THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale Sciences Chimiques et Biologiques pour la Santé (CBS2)

Centre de Biologie Structurale - CBS -
Unité mixte de recherche CNRS UMR 5048 - Université de Montpellier - Inserm U 1054

Impact de la phosphorylation des régions C-terminales, désordonnées et fonctionnelles, des RCPGs sur la signalisation cellulaire dépendante de l'arrestine

Présentée par Myriam GUILLIEN
Le 14 décembre 2021

Sous la direction du Dr. Nathalie SIBILLE

Devant le jury composé de

Pascale CRÉPIEUX, Directrice de recherche CNRS, PRC INRAE - CNRS - Univ. Tours, Nouzilly	Rapportrice
Olivier LEQUIN, Professeur Sorbonne Université, LBM, Paris	Rapporteur
Carine VAN HEIJENOORT, Directrice de recherche CNRS, ICSN, Gif-sur-Yvette	Examinatrice
Jean-Louis BANÈRES, Directeur de recherche CNRS, IBMM, Montpellier	Président du jury
Nathalie SIBILLE, Chargée de recherche CNRS, Centre de Biologie Structurale, Montpellier	Directrice de thèse

UNIVERSITÉ DE MONTPELLIER

REMERCIEMENTS

Ce travail de thèse a été effectué au sein du Centre de Biochimie Structurale (CBS) de Montpellier. Je remercie le directeur Pierre-Emmanuel Milhiet de m'avoir accueillie au sein du laboratoire. Je remercie le Labex EpiGenMed de m'avoir attribué une bourse de thèse m'ayant permis de vivre cette expérience unique, sous l'encadrement de ma directrice de thèse, Nathalie Sibille. Le projet a également été financé par l'Agence Nationale de la Recheche (ANR) et est en collaboration avec l'équipe de Jean-Louis Banères, que je remercie.

Je tiens à remercier les membres du jury d'avoir accepté de juger ce travail. Je remercie Pascale Crépieux et Olivier Lequin d'avoir accepté d'être rapporteurs de ce manuscrit. Je remercie Carine Van Heijenoort et Jean-Louis Banères d'avoir accepté d'examiner ce travail, et d'avoir participé tous les ans au développement de cette thèse lors des comités de suivi de thèse.

Je remercie plus particulièrement ma directrice de thèse, Nathalie Sibille, pour ces trois années passées ensembles. Merci pour l'apprentissage théorique et technique de la RMN. Merci d'avoir participé à mon développement de scientifique. Merci pour toutes les expériences professionnelles uniques auquelles j'ai pu participer (mentorat, encadrement de stagiaire, école d'été, congrès). Merci pour les moments non-scientifiques partagés ensembles. Merci pour ton écoute, ta bienveillance, ta patience (que j'ai parfois mise à rude épreuve ;)). Merci pour tes conseils. Je pars avec une valise complète de savoir-faire, savoir-être et de connaissances dans tous les domaines ! (de la biologie moléculaire à la physique quantique haha). Merci d'avoir fait de moi une « thésarde épanouie ».

Je remercie tous les membres de l'équipe 2 pour leur soutien et les bons moments passés ensembles, que ce soit au laser game (malgré une écrasante victoire de l'équipe possédant Carlos !), ou au minigolf. Je remercie plus particulièrement Aurélie Fournet pour son soutien et son travail sur le projet. Je remercie aussi Assia Mouhand, mon 'bînome arrestine', dont la bienveillance et le sourire quotidien « illumine » les journées de toute personne travaillant à ses côtés ;).

Merci à tous les étudiants, doctorants, post-doctorants, présents et passés, avec qui nous avons passé des soirées foot, des soirées bières, des soirées jeux de sociétés. Un merci plus particulier à Lucille Senicourt, Justine Girard et Anaïs Biquet. Je vous souhaite, tout comme à Assia et Aurélie, d'atteindre vos objectifs professionnels et personnels. Vous allez toutes me manquer (Nathalie comprise). Girl power !

Et enfin, aux personnes les plus importantes de ma vie, mes amis Laure Souillès, Noémie Portier, Laura Guée, Benjamin Vaissier ; ma belle-famille, Laetitia Lerat et sa famille ; ma mère, Françoise Coppenet ; mon beau-père, Valery Seillier ; mes frères Thibault, Grégoire et Aymeric Guillien ; et mon fiancé, Jordan Vasselin. Vous m'avez tous, tout au long de cette aventure, soutenue avec beaucoup de patience, d'amitié et d'amour. Merci.

RÉSUMÉ

Les récepteurs couplés aux protéines G (RCPGs) sont des acteurs importants de la signalisation cellulaire et représentent, à ce titre, la cible d'un tiers des médicaments mis sur le marché. Il est donc fondamental de mieux comprendre les mécanismes moléculaires associés à leurs signalisations dépendante et indépendante des protéines G. La signalisation indépendante des protéines G est médiée par l'interaction des régions C-terminales des RCPGs avec les protéines arrestines. Cependant, ces régions sont intrinsèquement désordonnées (IDR) et variables en longueur et en séquence, rendant leur étude difficile. Cette interaction fonctionnelle est modulée par la phosphorylation des domaines C-terminaux des RCPGs par des kinases spécifiques (GRKs). Cette thèse a pour objectifs de comprendre l'impact de la phosphorylation par les GRKs sur la conformation des domaines C-terminaux des RCPGs et leur interaction avec leur partenaire physiologique arrestine. La stratégie mise en place a consisté à produire des protéines recombinantes des domaines C-terminaux, sous forme isolées et solubles, de trois RCPGs (RCPG-Cter) de la classe A : les récepteurs vasopressine V2 (V2R-Cter), ghréline (GHSR-Cter) et β2-adrénergique (β2AR-Cter). La phosphorylation par GRK2 ainsi que l'utilisation de variants phosphomimétiques de ces domaines C-terminaux a permis d'étudier l'impact de la phosphorylation sur leur conformation et leur interaction avec l'arrestine-2. Les analyses structurale et fonctionnelle de ces régions désordonnées ont été effectuées par l'utilisation de méthodes biophysiques complémentaires comme la résonance magnétique nucléaire (RMN), la diffusion des rayons X aux petits angles (SAXS) et la spectroscopie de fluorescence (FRET). Mes travaux de thèse contribuent à une nouvelle image de la signalisation des RCPGs dépendante de l'arrestine. Ils montrent que la phosphorylation par les GRKs modifie la conformation des domaines C-terminaux des RCPGs et module la région d'interaction de ces domaines avec l'arrestine.

Mots clés : RMN, IDP, RCPG, GRK, arrestine, interaction

SUMMARY

Impact of phosphorylation of the functional and disordered C-terminal regions of GPCR on arrestin signaling pathways

G protein-coupled receptors (GPCRs) are central actors of cell signaling. They are the target of one third of the current cellular drugs. Thus, it is crucial to understand the molecular mechanism behind G-protein dependent and independent signaling pathways. G-protein independent cell signaling is mediated by the interaction of the cytoplasmic C-termini of GPCRs with arrestin. These C-termini are intrinsically disordered regions (IDRs) and are variable in length and sequence, making their study more elusive. This interaction is regulated by the phosphorylation of GPCR C-terminal domains by specific GPCR kinases (GRKs). My thesis objectives aimed at understanding the impact of GRK phosphorylation on the conformation and the interaction of GPCR C-termini with their arrestin partner. The strategy implemented was to produce excised recombinant C-terminal domain of three well-studied GPCRs: vasopressin V2 (V2R), growth hormone secretagogue 1a (GHSR) and β2-adrenergic (β2AR) receptors. Phosphorylation by GRK2, as well as, the use of phosphomimetic variants of these GPCR C-terminal domains allowed us to study the impact of phosphorylation on their conformation and interaction with arrestin-2. Structural and functional analyses of these disordered regions were performed using complementary biophysical methods such as solution-state nuclear magnetic resonance (NMR), small angle X-ray scattering (SAXS) and fluorescence spectroscopy (FRET). My thesis work postulates for a new picture of GPCR signaling dependent of arrestine. It shows that phosphorylation by GRKs changes the conformation of the C-terminal domain of GPCRs and modulates the binding region of these domains with arrestin.

Keywords: NMR, IDP, GPCR, GRK, arrestin, interaction

TABLE DES MATIÈRES

REMERCIEMENTS **1**

RÉSUMÉ **3**

SUMMARY **4**

TABLE DES MATIÈRES **5**

LISTE DES TABLEAUX **8**

LISTE DES ILLUSTRATIONS **9**

LISTE DES ABRÉVIATIONS **13**

PRÉAMBULE ET ORGANISATION DU MANUSCRIT **16**

1 CONTEXTE BIOLOGIQUE **18**

1.1 La signalisation des RCPGs : État de l'art **19**

1.2 La voie de signalisation dépendante de l'arrestine **39**

1.2.1 Les arrestines visuelles et les β-arrestines 39

1.2.2 Activation de l'arrestine 40

1.2.3 Fonction des arrestines 41

1.2.4 La phosphorylation des RCPGs par les GRKs 47

1.2.5 L'interface RCPG:arrestine 53

2 OBJECTIF DE LA THÈSE **64**

2.1 Objectif de la thèse **65**

2.2 Système modèle de RCPGs **67**

2.2.1 Récepteurs de la classe A des RCPGs 67

2.2.2 Des régions C-terminales aux récepteurs entiers 69

2.3 Stratégie d'étude **70**

2.3.1 Étude des régions C-terminales des RCPGs 70

2.3.2 Étude du réarrangement global de l'arrestine 71

3 MATÉRIEL ET MÉTHODES **73**

3.1 Production des protéines **74**

3.1.1 Production des domaines C-terminaux des RCPGs et de l'arrestine-2 74
3.1.2 Production des kinases GRK2 et GRK6 81
3.2 Caractérisation biophysique 94
3.2.1 Caractérisation globale des IDPs 94
3.2.2 Estimation de la constante de dissociation 100
3.3 La Résonance Magnétique Nucléaire (RMN) 106
3.3.1 Principe de la RMN 106
3.3.2 Expérimentation 107
3.3.3 Attribution 3D des régions C-terminales des RCPGs 107
3.3.4 Caractérisation structurale et dynamique des RCPG-Cters 110
3.3.5 Étude de l'interaction RCPG-Cter:arrestine-2 par CSP 122
3.3.6 Étude de la phosphorylation des RCPG-Cters 124
3.3.7 Analyse statistique 126
4 RÉSULTATS ET DISCUSSION 128
4.1 ARTICLE 1 : problématique et résumé 129
4.2 ARTICLE 2 : problématique et résumé 160
4.3 PARTIE 3 : problématique et résumé 200
4.3.1 Pourquoi mettre en place la production des GRKs au laboratoire ? 202
4.3.2 Suivi de la phosphorylation des IDPs par RMN 203
4.3.3 Optimisation du tampon de phosphorylation 206
4.3.4 Les sites de phosphorylation de GRK2 209
4.3.5 La phosphorylation module le repliement des RCPG-Cters et leur région d'interaction à l'arrestine-2 221
4.4 PARTIE 4 : problématique et résumé 225
4.4.1 Analyse de la conformation de l'arrestine wt et R169E seule 226
4.4.2 Analyse de la conformation de l'arrestine wt et R169E en complexe 228
5 CONCLUSION 233

5.1 Conclusion générale 234

5.2 Perspectives 237

ANNEXES 240

BIBLIOGRAPHIE 286

LISTE DES TABLEAUX

Table 1 : Les arrestines partagent ~ 50 % d'identité de séquence. 39
Table 2 : Les GRKs partagent au minimum 30 % d'identité de séquence. 47
Table 3 : Structures de l'arrestine en interaction avec un RCPG. 53
Table 4 : Les RCPGs modèles utilisés dans ce manuscrit partagent ~ 23 % d'identité de séquence. 67
Table 5 : Séquence des peptides C-terminaux des récepteurs β2AR, V2R et GHSR. 70
Table 6 : Tampons de purification des domaines C-terminaux des RCPGs et de l'arrestine-2. 76
Table 7 : Rendement de production des peptides des RCPGs marqués ^{15}N et de l'arrestine-2 en fin de purification. 79
Table 8 : Séquence des amorces utilisées pour vérifier par PCR l'intégration du gène GRK dans le bacmide. 84
Table 9 : Tampons de purification de GRK2. 90
Table 10 : Expériences utilisées pour l'attribution de la chaîne principale des peptides des RCPGs. 109
Table 11 : Concentration en peptides des RCPGs utilisées pour la mesure des différents paramètres RMN. 111
Table 12 : Corrélation entre les valeurs de couplage $^{3}J_{HNHA}$ et les structures secondaires transitoires des IDPs. 114
Table 13 : Recherche d'un motif consensus de phosphorylation pour GRK2. 220

LISTE DES ILLUSTRATIONS

Figure 1 : Conformation basale de l'arrestine-2 (Figure modifiée à partir de (Seyedabadi *et al*, 2021)). 40

Figure 2 : Changement conformationnel de l'arrestine en complexe avec les RCPGs (Figure modifiée à partir de (Gurevich & Gurevich, 2019b)). 41

Figure 3 : Représentation schématique des fonctions de l'arrestine dans la signalisation des RCPGs. Figure modifiée à partir de (Luttrell et al, 2001). 42

Figure 4 : Désensibilisation des récepteurs par compétition pour le site de liaison entre les protéines G et les arrestines. (Figure extraite de (Chaturvedi *et al*, 2020)). 43

Figure 5 : L'arrestine sert d'adaptateur à la séquestration des récepteurs dans les endosomes. (Figure modifiée à partir de (Moore *et al*, 2007). 44

Figure 6 : Implication de l'arrestine dans la croissance cellulaire et la survie via la voie ERK (Figure modifiée à partir de (Ahn *et al*, 2020)). 46

Figure 7 : Modèle d'activation des GRKs par la liaison au récepteur activé (Figure modifiée à partir de (Chen *et al*, 2021)). 48

Figure 8 : Expression différenciée des GRKs ubiquitaires 50

Figure 9 : Représentation schématique du phospho-barcode du récepteur β2-adrénergique (β2AR) (Figure modifiée à partir de (Wisler *et al*, 2014). 51

Figure 10 : Représentation schématique du modèle du phospho-barcode (Figure modifiée à partir de (Wisler *et al*, 2014). 52

Figure 11 : Les deux modes de liaison de l'arrestine au récepteur (Figure modifiée à partir de (Smith & Rajagopal, 2016)et (Shukla *et al*, 2013)) 54

Figure 12 : Interaction des boucles de l'arrestine avec le sillon cytoplasmique du récepteur. 56

Figure 13 : Comparaison des complexes RCPG:arrestine (Figure extraite de (Seyedabadi *et al*, 2021)) 57

Figure 14 : L'arrestine contient des poches permettant d'accueillir les phosphates des RCPGs. 59

Figure 15 : La localisation des phosphates des RCPGs affecte les interfaces d'interaction de l'arrestine avec ses partenaires (Figure modifiée à partir de (He *et al*, 2021)). 61

Figure 16 : Structure du phospho-peptide V2Rpp4 en complexe avec l'arrestine-2 (Figure extraite de (He *et al*, 2021). 62
Figure 17 : Sites de phosphorylation des GRKs dans les domaines C-terminaux des récepteurs β2AR, V2R et GHSR (Figure extraite de GPCRdb (https://gpcrdb.org/)). 68
Figure 18 : Utilisation d'un acide glutamique pour mimer la phosphorylation. 70
Figure 19 : La mutation R175E entraîne une destruction du cœur polaire de l'arrestine-1 72
Figure 20 : Représentation schématique des protéines recombinantes utilisées dans ce manuscrit. 75
Figure 21 : Purification et pureté de l'arrestine-2. 78
Figure 22 : En SEC, le domaine C-terminal de GHSR élue plus rapidement que ce qui est attendu pour une protéine de 3.6 kDa. 79
Figure 23 : Clivage du domaine C-terminal de β2AR par la 3C protéase. 80
Figure 24 : Production du bacmide recombinant exprimant les protéines GRKs chez *E. coli* DH10Bac (Figure modifiée à partir de (Ciccarone *et al*, 1998)). 82
Figure 25 : Séquence des protéines recombinantes GRK2 et GRK6 produites au laboratoire. 83
Figure 26 : Vérification par PCR de l'intégration du gène GRK dans le bacmide. ... 85
Figure 27 : Vérification par PCR de l'intégration du gène GRK6 dans le bacmide avec les amorces sens et anti-sens PUC/M13. 86
Figure 28 : Infection des cellules *Sf9* par les virus recombinants. 88
Figure 29 : Test d'expression des protéines GRK2 et GRK6 à différents ratios bacmide:*Sf9*. 89
Figure 30 : Chromatogrammes issus de la chromatographie d'affintié de GRK2 (A) et GRK6 (B). 91
Figure 31 : Chromatogrammes issus de la chromatographie échangeuse de cation de GRK2 (A) et GRK6 (B). 92
Figure 32 : Caractérisation spectrale des domaines désordonnés et globulaires de la protéine HPV E7 (Figure modifiée à partir de (Chemes *et al*, 2012)). 95
Figure 33 : Fonctionnement d'un SEC-MALS (Figure modifiée à partir de (Minton, 2016)). 96
Figure 34 : Représentation schématique de la technique de SAXS (Figure modifiée à partir de (Blanchet & Svergun, 2013)). 97

Figure 35 : Les différentes courbes pouvant être obtenues par SAXS (Figure modifiée à partir de (Blanchet & Svergun, 2013)). 98
Figure 36 : Représentation schématique du principe général de la polarisation de la fluorescence (Figure extraite de (Moerke, 2009)). 101
Figure 37 : Marquage du domaine C-terminal de GHSR phosphomimétique par l'alexa Fluor® 488 succinimidyl ester (Alexa488 SE). 102
Figure 38 : Principe de la MST (Figure modifiée à partir de (Jerabek-Willemsen *et al*, 2014)..... 104
Figure 39 : Marquage des cystéines de l'arrestine par l'alexa Fluor® 488 maleimide C5 (Alexa488 MC_5). 105
Figure 40 : Représentation schématique du principe de la RMN..... 106
Figure 41 : Stratégie d'attribution RMN. 108
Figure 42 : Relation entre le couplage scalaire à trois liaisons $^3J_{HNHA}$ et l'angle φ du squelette protéique. 113
Figure 43 : Relation entre les RDCs et l'orientation du vecteur $\boldsymbol{NH}$ dans les structures secondaires transitoires des IDPs. (Figure modifiée à partir de (Jensen *et al*, 2009)). 116
Figure 44 : Milieux orientants $C_{12}E_5$/hexanol et phage Pf1 par rapport au champ magnétique B_0. 116
Figure 45 : Observation du dédoublement quadrupolaire de la raie du deutérium (ΔD) sur un spectre RMN ^2H enregistré en milieu $C_{12}E_5$/hexanol (Figure extraite de (Rückert & Otting, 2000)). 117
Figure 46 : Phénomène de relaxation des spins..... 119
Figure 47 : Obtention des formes paramagnétiques et diamagnétiques. 121
Figure 48 : Suivi de la phosphorylation des IDPs par RMN. 204
Figure 49 : Disparition de pics par échange proton amide avec l'eau à pH 7.5. 205
Figure 50 : Mise au point du protocole de phosphorylation des régions C-terminales des RCPGs par GRK2. 208
Figure 51 : Cinétique des sites phosphorylés du domaine C-terminal de β2AR (β2AR-Cter) par GRK2..... 209
Figure 52 : Perturbation des sites voisins aux sites phosphorylés par GRK2 du domaine C-terminal de β2AR (β2AR-Cter). 211
Figure 53 : Attribution du domain C-terminal de GHSR phosphorylé par GRK2.... 213

Figure 54 : Visualisation de potentiels sites phosphorylés sur le domaine C-terminal de V2R (V2R-Cter) par GRK2..... 215

Figure 55 : Comparaison de nos sites phosphorylés déterminés par RMN (orange) aux sites connus et prédits pour être phosphorylés dans les domaines C-terminaux des récepteurs β2AR, GHSR et V2R. 216

Figure 56 : Effet de la phosphorylation de GRK2 sur la conformation de la région C-terminale de GHSR et sur son interaction avec l'arrestine-2, et comparaison avec les résultats du phosphomimétiques 222

Figure 57 : Expérience de digestion trypsique limitée de l'arrestine sauvage (wt, A) et R169E (B). 226

Figure 58 : Analyse par SEC-SAXS des arrestines sauvage (wt, violette) et R169E (orange) 227

Figure 59 : Analyse par SEC-SAXS de l'arrestine-2 en complexe avec le variant phosphomimétique de GHSR-Cter (pm-GHSR-Cter), sans saturation (A) ou avec saturation (B) de la colonne en peptide. 229

Figure 60 : Interaction entre l'arrestine sauvage (wt) ou R169E et le peptide de GHSR-Cter (sauvage, wt ; phosphomimétique, pm). 230

Figure 61 : Expériences de FRET avec ou sans les peptides C-terminaux des RCPGs..... 232

Figure 62 : La phosphorylation par les GRKs régule la conformation des domaines C-terminaux des RCPGs et leur interaction à l'arrestine..... 237

LISTE DES ABRÉVIATIONS

Abs	Absorbance
ADN	Acide DésoxyriboNucléique
AP-2	Protéine adaptatrice 2 du système endocytique
ARN	Acide RiboNucléique
ATP	Adénosine TriPhosphate
AVP	Arginine VasoPressine
BRET	Transfert d'énergie par résonance de bioluminescence
β2AR	Récepteur β2-adrénergique
3C	Protéase 3C PreScission dérivée du rhinovirus humain
C7pp	Phospho-peptide synthétique du domaine C-terminal de CXCR7
CD	Dichroïsme Circulaire
Cryo-EM	Microscopie électronique cryogénique
CSA	Anisotropie de déplacement chimique
CSI	Index de déplacement chimique
CSP	Perturbation du déplacement chimique
CXCR7	Récepteur de chimiokine s'apparentant aux RCPGs de classe A
2D	2 dimensions
3D	3 dimensions
Da	Dalton : poid moléculaire (1 Da ≈ 1 g/mol)
DO	Densité Optique
DSF	Fluorimétrie différentielle
DTT	DiThioThréitole
ERK	Kinase régulée par le signal extracellulaire
E. coli	Bactérie *Escherichia Coli*
ECL	Boucle extra-cellulaire
EDTA	Acide éthylène-diamine-tétra-acétique
eIF4e	Protéine 2 qui se lie au facteur de traduction eucaryote 4E
Fab	Anticorps composé uniquement du fragment de liaison à l'antigène
FM	Logiciel *Flexible Meccano*
FRET	Transfert d'énergie de fluorescence par résonance de type Förster
FT	Molécules non-retenues sur les colonnes de chromatographies

HEK	Cellules embryonnaires humaines issues du rein
HetNOE	Effet Oerhauser hétéronucléaire
HRP	Enzyme peroxydase de raifort
HSQC	Corrélation hétéronucléaire à simple Quantum
GFP	Protéine fluorescente verte
GHSR	Récepteur sécrétagogue de l'hormone de croissance de type 1
GRK	Kinase spécifique des RCPGs
GRP	Phosphatases des RCPGs
GST	Glutathione S Transferase
ICL	Boucle intra-cellulaire
IDP	Protéine Intrinsèquement Désordonnée
IDR	Région Intrinsèquement Désordonnée
INEPT	Renforcement de transfert de polarisation entre deux noyaux
IPTG	IsoPropyl β-D-1-ThioGalactopyranoside
ITC	Calorimétrie par titrage isotherme
K_d	Constante de dissociation
K_m	Constante de Michaelis
LB	Milieu riche de culture Luria Broth
M2R	Récepteur muscarinique de type 2
M9	Milieu minimum de culture à supplémenter en azote et en carbone
MALS	Diffusion de la lumière aux petits angles
MAPK	Protéine kinase activée par des mitogènes
MEK-1/2	MAPK/ERK kinase-1/2
MS	Spectrométrie de masse
MST	Thermophorèse de fluorescence à micro-échelle
N5052	Milieu de culture auto-inductible à supplémenter en azote
Nb	Nanocorps, fragment d'anticorps composé d'un seul domaine
NLS	Séquence de localisation nucléaire
NTSR1	Récepteur à la neurotensine de type 1
ON	Sur la nuit
PAGE	Electrophorèse sur gel de polyacrylamide
PCR	Réaction de polymérisation en chaîne
PEG	Alkyl-Poly(Ethylène Glycol)
pm	Variant phosphomimétique

PMSF	PhenylMethylSulfonyl Fluoride
PP-2A	Protéine Phosphatase-2A
PPI	Interaction protéine-protéine
PPII	Hélice polyproline de type 2
PTM	Modification post-traductionnelle
PRE	Augmentation de la relaxation paramagnétique
Raf	Protéine issue d'un gène proto-oncogène pouvant induire des fibrosarcome à accélération rapide
RCPG	Récepteur Couplé aux Protéines G
RCPG-Cter	Domaine C-terminal des récepteurs couplés aux protéines G
RDC	Couplage dipolaire résiduel
RF	Onde radiofréquence
RLuc	Protéine fluorescente Renilla Luciferase
RMN	Résonance Magnétique Nucléaire
Rho	Récepteur à la rhodopsine
RX	Cristallographie aux rayons X
SAXS	Diffusion des rayons X aux petits angles
SDS	Laurylsulfate de sodium
SEC	Chromatographie d'exclusion de taille
Sf9	Cellules d'insectes de *Spodoptera frugiperda*
siRNA	Petit ARN interférent
SLiM	Petit motif linéaire de séquence protéique
SOC	Milieu Luria Broth super optimal avec répression des catabolites
Src	Protéine kinase proto-oncogène pouvant induire un sarcome
T4L	Protéine lysozyme du phage T4
TBST	Solution saline tamponnée en Tris contenant du Tween
TCPE	Tris(2-CarboxyEthyl)Phosphine
TEV	Protéase issue du virus responsable de la maladie du tabac
TM	Hélice transmembranaire
V2R	Récepteur à la vasopressine V2
V2Rpp	Phospho-peptide synthétique du domaine C-terminal de V2R
wt	Sauvage
YFP	Protéine fluorescente jaune

PRÉAMBULE ET ORGANISATION DU MANUSCRIT

Dans les organismes pluricellulaires, les cellules communiquent entre elles pour régir leur métabolisme, leur développement, leur organisation et leur activité. On parle de communication cellulaire. Cette communication cellulaire passe par l'échange de signaux, qui sont portés par des molécules variées (hormones, neurotransmetteurs, peptides, lipides), d'une cellule à une autre. Le plus souvent, les cellules réceptrices reçoivent et transmettent le signal à l'intérieur de la cellule, en passant à travers la membrane plasmique grâce à des récepteurs transmembranaires. La plus grande famille de récepteurs membranaires humains est la famille des Récepteurs Couplés aux Protéines G (RCPGs). Leur signalisation implique deux voies intracellulaires principales qui sont la voie canonique dépendante de la protéine G et la voie dépendante des protéines arrestines. Mes travaux de thèse font partie d'un projet global (ANR-10-LABX-12-01 et ANR-17- CE11-0022-01) dont l'objectif est de décrire les principes généraux de l'interaction RCPG:arrestine.

Mes travaux de thèse vont vous être présentés à travers cinq chapitres.

Le **chapitre 1** présente le contexte biologique du sujet de thèse. Il comprend un état de l'art de la signalisation des RCPGs et de l'importance du désordre dans cette signalisation. Cette partie sera présentée à travers la revue « IDPs and their complexes in GPCR and nuclear receptor signaling » (Guillien *et al*, 2020). Suite à la revue, la signalisation des RCPGs dépendante de l'arrestine sera reprise et décrite plus en détail.

Le **chapitre 2** présente les objectifs du sujet de thèse et la stratégie mise en place pour y répondre.

Le **chapitre 3** correspond au matériel et méthode. Il contient les protocoles de production des protéines étudiées et les méthodes utilisées pour leur caractérisation structurale, dynamique et fonctionnelle : le dichroïsme circulaire (CD), la chromatographie d'exclusion couplée à la diffusion de la lumière à plusieurs angles (SEC-MALS), la diffusion des rayons X aux petits angles (SAXS) et la résonance magnétique nucléaire (RMN).

Le chapitre 4 est consacré aux résultats et à leur discussion. Les résultats sont séparés en quatre parties dont les deux premiers sont présentés sous forme d'article. L'*article 1* reprend la caractérisation structurale et dynamique des domaines C-terminaux des RCPGs dans leur forme basale, non-phosphorylée. Cette étude

n'avait jamais été réalisée auparavant et a permis de mettre en évidence la présence de structures secondaires transitoires dont la conformation diffère lorsqu'elles sont en interaction avec l'arrestine. L'*article 2* présente les profils de structuration de variants phosphomimétiques des domaines C-terminaux et leur comparaison à leur forme basale, non-phosphorylée. Cette comparaison a montré que les mimes de la phosphorylation ont pour effet de modifier la conformation des régions C-terminales des RCPGs et permettent de déclencher l'interaction avec le partenaire arrestine. Dans la *partie 3*, nous présentons la mise en place au laboratoire de la phosphorylation des régions C-terminales par les GRK2 et 6. L'analyse de la conformation de GHSR phosphorylé par GRK2, et de son interaction avec l'arrestine, a montré que la phosphorylation module le repliement du C-terminus de GHSR et sa région d'interaction à l'arrestine, tout comme les variants phosphomimétiques. La *partie 4* est consacrée à l'étude de la conformation globale de l'arrestine sauvage ou variant R169E, libre et/ou liée par SAXS et par FRET (Transfert d'énergie de fluorescence par résonance de type Förster). Ces analyses montrent que les différences conformationnelles entre les formes inactives, pré-actives et actives de l'arrestine sont subtiles et probablement associées à des changements de dynamiques.

Enfin, **le chapitre 5** est une conclusion générale qui reprend l'ensemble des résultats obtenus durant ma thèse et leur développement futur (ouverture).

1 CONTEXTE BIOLOGIQUE

1.1 LA SIGNALISATION DES RCPGS : ÉTAT DE L'ART

L'état de l'art des récepteurs couplés aux protéines G (RCPGs) et l'importance du désordre dans leur signalisation sont présentés dans ce paragraphe à travers la version pré-publiée de la revue « IDPs and their complexes in GPCR and nuclear receptor signaling » (Guillien *et al*, 2020). Cette revue a été publiée dans le journal Progress in Molecular Biology and Translational Science, 2020, volume 174, pages 105-155. L'article original est disponible sur pubmed (PMID : 32828464) et sur ScienceDirect. (DOI : 10.1016/bs.pmbts.2020.05.001).

En résumé, la revue est composée de trois parties principales. La partie '1 - introduction' présente ce qu'est une Région ou une Protéine Intrinsèquement Désordonnée (IDR ou IDP). La partie '2 - G protein-coupled receptors' décrit les grands principes de la signalisation des RCPGs et l'importance du désordre dans cette signalisation. Finalement, la partie '3 - Nuclear receptors and their transcriptional coregulators' concerne les récepteurs nucléaires et ne sera pas nécessaire à la compréhension de cette thèse. Cette troisième partie a donc été déplacée en Annexe 1 du manuscrit (Annexe 1, p. 241) où la version intégrale pré-publiée de la revue est disponible (références incluses).

CHAPTER FOUR

IDPs and their complexes in GPCR and nuclear receptor signaling

Myriam Guillien[a], Albane le Maire[a], Assia Mouhand[a], Pau Bernadó[a], William Bourguet[a], Jean-Louis Banères[b], Nathalie Sibille[a,*]

[a]Centre de Biochimie Structurale (CBS), Inserm, CNRS, Univ Montpellier, Montpellier, France
[b]Institut des Biomolécules Max Mousseron (IBMM), UMR 5247 CNRS, Université Montpellier, ENSCM, 15 av. Charles Flahault, 34093, Montpellier, France.

*Corresponding author: e-mail address: nathalie.sibille@cbs.cnrs.fr

Contents

1-Introduction
2-G protein-coupled receptors
  2.1 GPCRs : general trends and function
  2.2 Structural organization of GPCRs and activation process
  2.3 Extramembrane domains
    2.3.1 Intrinsic disorder
    2.3.2 Post-translational modifications
    2.3.3 Ligand binding and coupling to signaling partners
  2.4 Therapeutic implications
  2.5 Conclusions
3-Nuclear receptors and their transcriptional coregulators
  3.1 Biological context
  3.2 The NTD/AF-1 of NRs and its interactions with transcriptional coregulators
  3.3 Transcriptional coregulators and their interaction with the CTD/AF-2 of NRs
  3.4 Importance of NR-NTDs in diseases and as pharmaceutical targets
  3.5 Liquid-Liquid Phase Separation (LLPS) in NRs
4-Conclusions & Perspectives
Acknowledgments
References

Abstract

G protein-coupled receptors (GPCRs) and Nuclear Receptors (NRs) are two signaling machineries that are involved in major physiological processes and, as a consequence, in a substantial number of diseases. Therefore, they actually represent two major targets for drugs with potential applications in almost all public health issues. Full exploitation of these targets for therapeutic purposes nevertheless requires opening original avenues in drug design, and this in turn implies a better understanding of the molecular mechanisms underlying their functioning. However, full comprehension of how these complex systems function and how they are deregulated in a physiopathological context is obscured by the fact that these proteins include a substantial number of disordered regions that are central to their mechanism of action but whose structural and functional properties are still largely unexplored. In this chapter, we describe how these intrinsically disordered regions (IDR) or proteins (IDP) intervene, control and finely modulate the thermodynamics of complexes involved in GPCR and NR regulation, which in turn triggers a multitude of cascade of events that are exquisitely orchestrated to ultimately control the biological output.

Keywords

Intrinsically disordered proteins (IDPs), G protein-coupled receptors (GPCR), nuclear receptors (NRs), macromolecular complexes, signaling, G protein-coupled receptor kinase (GRK), arrestins, transcriptional coregulators, post-translational modifications (PTMs), ligands

1. Introduction

Why are Intrinsically Disordered Proteins (IDPs) so fascinating? They were discovered in the mid-90s and it is now well established that IDPs are very common in eukaryotes (40%)[1] where they confer complexity to these organisms.[2,3] Moreover, their inherent flexibility renders IDPs accessible to post-translational modifications (PTMs)[4] (phosphorylation,[5] ubiquitination,[6] palmitoylation, acetylation, sumoylation,[7] methylation). More than 300 PTMs can occur physiologically,[8] leading to the availability of more than 140 amino acids produced by these various modifications.[9] IDPs lack stable tertiary and secondary structure.[10] This is due to their specific amino acid sequences depleted in hydrophobic residues and enriched in polar and charged residues that are uncompensated; and that often have low complexity.[11] Their transient secondary structures are often conserved and linked to function.[12] Indeed, IDPs can adapt to many partners by changing their structure (promiscuity) and their ability to interact with high specificity and low affinity allow them to finely tune the thermodynamic properties of their complexes.[13]

With all these general biophysical features, it is not surprising that IDPs are linked to very specific functions such as signaling, transcription, etc.[14-16] As a consequence, their deregulation is at the origin of a multitude of diseases,[17] such as cancer,[18,19] neurodegeneration,[20,21] cardiovascular diseases[22] and diabetes[23]. In order to understand the role of disorder in function and disease, we have to characterize their structure and partner-recognition mechanisms. IDPs and their complexes are very difficult to characterize at the structural level due to their inherent flexibility.[24] The most suited way to address their study is by combining data from complementary techniques.[25] Indeed, Nuclear Magnetic Resonance (NMR) is the best-adapted technique to highlight the hot spots of IDPs in their free state.[26,27] Most of the times, these identified Short Linear Motifs (SLiMs)[28] or Molecular Recognition Features (MoRFs)[29] are evolutionary conserved[30] and are the functional part of IDPs.[31] These interacting regions identified in the complex, which usually fold upon binding,[32,33] can then be delimitated by NMR, synthesized and submitted to co-crystallization trials. In parallel, Small Angle X-ray Scattering (SAXS) and Small Angle Neutron Scattering (SANS) can provide the overall envelope of the entire system.[34,35] Finally, all these data are generally integrated into computational tools to deliver accurate structural models.[36,37] These data can be complemented or validated

by other biophysical methods such as FRET, interactions studies *in vitro* (fluorescence anisotropy, ITC, MST, etc), mutational analyses and cell-based assays allowing connecting detailed structural models with biological function. Such an integrative strategy where experimental information obtained from complementary techniques is integrated and simultaneously interpreted to derive a unified model of the biological system has already allowed the understanding of the structural bases of crucial biological processes involving IDPs.[14,38,39]

In this chapter, we will present two biological machineries involved in cell signaling for which structural disorder plays key roles: G Protein-Coupled Receptors (GPCRs) and Nuclear Receptors (NRs). GPCRs are the target of almost 40% of drugs on the market,[40-42] and NR ligands constitute 10-20% of the worldwide pharmaceutical market[43] and, as a consequence, it is essential to better understand their functioning to design more effective drugs without side effects. GPCRs and NRs are representative examples of the complementarity of ordered and disordered regions associated with distinct and complementary functions.[15] Both GPCRs and NRs (Figs. 2 and 5) share many structural features. They have a well-folded core domain whose high-resolution tridimensional structures have been solved by X-ray crystallography and cryo-EM, and unfolded domains that are not well characterized because they are missing from the electron densities (see below §2.1 and 3.1). For class A or (rhodopsin-class) GPCRs, the disordered domains are the N- and C-terminal domains plus extra- and intra-cellular loops. For NRs, the main disordered domains are the N-terminal Activation Domain (NTD), a linker between the DNA and ligand binding domains (DBD and LBD, respectively) and the H12 helices at the C-terminus of LBD. Both N-terminal domain of GPCRs and H12 of NRs undergo folding upon ligand binding processes that trigger different functional outputs. For GPCRs, this disorder to order transition triggers the transduction of signal via a transmembrane core reorganization, and for NRs it triggers the switch of the repressor complex by the activator one to initiate transcription of the target genes. NR coregulators are among the first proteins reported to be functional despite being mainly disordered.[44,45] Transcriptional coregulators are very large IDPs that include corepressorss and coactivators, containing around 2,400 and 1,400 amino acids, respectively. Concerning the cytoplasmic C-terminal domain of GPCRs, they were discovered with the tridimensional structure of the beta-2 adrenergic receptor

(β2AR)[46]. This structure came out seven years after the first membrane protein solved by X-ray crystallography, the Rhodopsin.[47] Similarly as NTD of NRs, C-terminal domains of class A GPCRs are very variable in length and amino acid composition (from 30 to 325 residues), and are recognized as fully disordered.[48] Intrinsically disordered coregulators, NTDs of NRs and C-terminal domains of GPCRs are submitted to a multitude of PTMs. Together with these PTM-sites, these highly flexible IDPs and IDRs contain SLiMs and/or MoRFs regions (for a review, see [49] for GPCRs and [50] for NRs). This combination increases the degree of diversity of binding sites, thus making the modulation of their interactions with their physiological partners more multifaceted, which ultimately allows these complexes to play their role of super regulators. These regions were assigned in the literature (described in [51] and illustrated in Fig. 1) as functional semi-foldons (i.e. partially folded protein segments with transient residual structure) and inducible foldons (i.e. segments that are able to fold upon interacting with specific binding partners). Moreover, the functions of non-foldons (i.e. protein segment with intrinsically disordered nature), which are found in this kind of regulating systems, are very important because of their obviously entropic chain activities. Indeed, this activity is remarkably well exemplified by coregulators that encompass multiple binding sites that increase the cooperativity in their interaction with NRs. This conformational variability complements the dynamical assembly of foldons (i.e. ordered segments or independently foldable regions) represented by the central core of GPCR and NRs (Figs. 2 and 5), which hold complementary function (transduction of signal for GPCRs and DNA and ligand recognition for NRs). The last piece of the mosaic of the protein structure-function space (Fig. 1) is illustrated by the functional unfoldons (a part of the ordered protein that has to undergo order-to-disorder transition to make protein active). This activity is exemplified in GPCRs interaction with arrestin, where the flexible cytosolic domain of GPCRs releases the C-terminal tail of arrestin that in turn will expose some specific interaction regions that will trigger different downstream effectors in the cell for specific functional responses.

The two complex machineries presented in this review, GPCRs and NRs, are illustrative of the structure-function continuum that undergoes IDPs and IDRs in these signaling and regulation systems to elegantly control cell signaling by orchestrating a combination of modular assembly of different type of foldons[51].

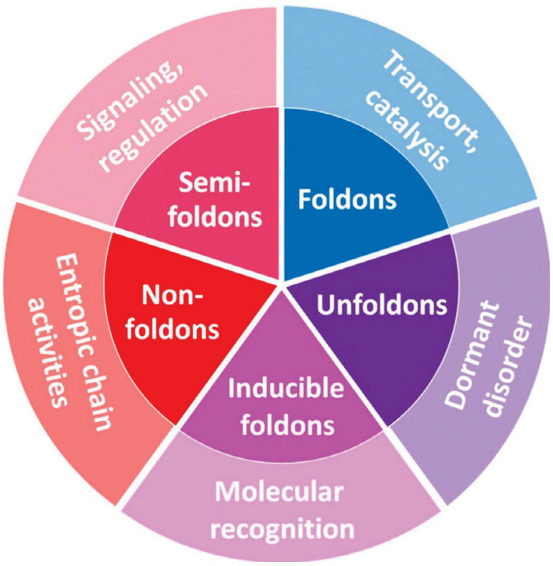


Fig. 1 Schematic representation of the mosaic nature of the protein structure–function space, where differently folded segments of a protein (foldons, semi-foldons, non-foldons, inducible foldons, and unfoldons) might possess different biological functions. Reproduced with permission from (Uversky, 2015).

2. G protein-coupled receptors

2.1 GPCRs: general trends and function

G protein-coupled receptors (GPCRs) are one of largest class of membrane receptors. They play a central role in most intercellular communication processes, including vision, metabolism and neurological communication.[52,53] As such, they represent a major target family for drugs, with potential applications in many clinical spheres including cancer, metabolic, inflammatory and neurological disorders or hypertension, among many others.[41]

GPCRs transmit information through a mechanism that primarily involves activation by an external agonist compound and subsequent coupling to intracellular signaling partners. Historically, the first of these partners to be identified were G proteins, from which GPCRs got their name. G proteins are αβγ-heterotrimers composed of diverse combinations of the twenty-one α, five β and twelve γ subunit isoforms.[54] Coupling to GPCRs triggers GDP-to-GTP exchange in the α subunit, resulting in the activation of the G protein. Activated G proteins in turn activate their downstream effectors to trigger signaling.[54] Besides G proteins, GPCRs can trigger signaling through an extensive set of cytoplasmic partners. Among them are arrestins. There are four isoforms of arrestins, including the two visual ones (arrestin

1 and 4), which are involved in both signaling and receptor desensitization.[55] In the classical paradigm, activated GPCRs are first recognized by G protein-coupled Receptor Kinases (GRKs), leading to receptor phosphorylation at specific sites.[56] The phosphorylated receptor becomes competent for interacting with arrestins. Arrestin binding then triggers recruitment of the internalization machinery to the activated GPCR, leading either to signal termination or to a second wave of signaling through internalized receptors.[57] In addition, arrestin recruitment to the phosphorylated receptor can initiate an alternate intracellular signal.[55,58] Whether this additional signal is correlated or not to G protein activation is still under debate, however.[59] Finally, in addition to G proteins, GRKs and arrestins, GPCRs can interact with a full network of cytoplasmic proteins, including scaffolding ones, that further regulate the transmission of the signal. This interaction network has been extensively described in particular for the serotoninergic system in the brain.[60]

2.2 Structural organization of GPCRs and activation process

The GPCR family is composed of almost 800 different members separated into five different classes (Glutamate, Rhodopsin, Adhesion, Frizzled/Taste2 and Secretin).[61] Although these classes differ in their structure and function, they all share common structural elements. Indeed, GPCRs are primarily composed of two separated domains, the intra- and extramembrane segments. For all GPCRs, the membrane domain is composed of seven transmembrane helices labeled TM1 to TM7.[61] The extramembrane region differs, depending on the class of GPCR considered. In some classes, *i.e.* glutamate and secretin, these regions include large extracellular domains that participate to ligand binding and receptor activation. In others, in particular the rhodopsin-like, the extramembrane regions essentially consist in relatively short N- and C-extensions, including a cytoplasmic H8 helix, and extra- and intracellular loops connecting the transmembrane helices. These segments in the rhodopsin class of GPCRs will be the main focus of the present chapter, as the large N-terminal parts of the glutamate and secretin receptors essentially display the characteristics of folded protein domains. In contrast to the transmembrane sequences that include highly conserved motifs, the extracellular and cytoplasmic regions of GPCRs are far more divergent in sequence.[62]

The three-dimensional organization of the membrane domain of GPCRs will be described here briefly, as it is not the focus of the present chapter. Starting from the

first crystal structures of rhodopsin[47] and the β_2-adrenergic receptor,[46,63] the structural arrangement of the membrane domain of GPCRs has been progressively illuminated during the last decade through a series of crystal structures of receptors loaded with either agonists or antagonists (for a review, see [64]). These structures provided a detailed description of the general architecture of the helical domains, of the orthosteric and/or allosteric ligand binding sites, as well as the conformational changes associated with agonist binding. Structural analyses subsequently provided the arrangement of the complexes of receptors with their signaling partners. The first structure was obtained through X-ray crystallography with the β_2-adrenergic receptor associated to Gs.[65] The subsequent explosion in cryo-electron microscopy (cryoEM) approaches then provided us with a plethora of structures of receptors from different classes, *i.e.* rhodopsin (for a review, see [66]), secretin (for a review, see [67]), glutamate[68] and Smoothened[69], in complex with different G protein partners (G11, Gs, Gi/o). More recently, cryoEM structures were obtained for receptors in complex with arrestins[70-72] or even with both G proteins and arrestins[73]. All these structures provide a detailed description of the three-dimensional arrangement of the complexes as well as the changes in receptor and signaling protein conformation associated to the activation process.

Besides the transmembrane domains, GPCRs include a significant proportion of residues that lie out of the lipid bilayer. These include the N- and C-terminal domains as well as in the intra- and extracellular loops. In contrast to our initial view, these regions are not to be considered as ancillary domains. Depending on the class of GPCR considered, they range from ca. 40% (rhodopsin-like receptors) to ca. 70% (glutamate) of the total receptor residues.[74] However, so far, their structural arrangement remains obscure. Indeed, both their conformational dynamics and their modification for crystallization purposes (truncation of the N- and C-terminal regions, fusion of protein partners in the intracellular loops or N-terminus) precluded their observation in most of the GPCR crystal structures. This dynamic character also makes the use of modeling methods uncertain. In a few crystal- and cryoEM structures of the receptor:effector complexes, some of these regions are visible, however, in particular when their dynamics is restrained because of the direct interaction with the signaling partner (see below).

GPCRs transmit an information through concerted changes in their conformation. The three-dimensional structure of receptors in complex with antagonists, agonists and signaling proteins provided a detailed description of these changes. These include large amplitude motions in the TM helices as well as specific conserved microswitches. However, the diversity in signaling properties of GPCRs strongly suggests that this process can be hardly accounted for with the limited number of conformational states described in the crystal and cryoEM structures. Combination of computational and experimental approaches, including fluorescence and NMR, rather indicates that control and modulation of signaling involves a strong conformational plasticity GPCRs. The current model proposes that receptors explore complex conformational landscapes populated with multiple states, and that they adapt their shape in response to their environment through a series of conformational rearrangements that are central to the selective activation of downstream signaling partners.[75,76]

2.3 Extramembrane regions

As state above, a significant part of GPCRs lies outside the membrane. These include the N- and C-terminus as well as the intra- and extracellular loops connecting the TM segments. In the case of rhodopsin-like receptors, they represent about 40% of the total receptor residues. Initially, these loop segments were essentially considered as necessary to maintain the architecture of the receptor central core. This is the case, as they likely constrain the respective arrangement of the transmembrane helices,[77] but not only. Indeed, in addition to their structural role, these regions are deeply involved in the functional properties of GPCRs and are subject to specific post-translational modifications that modulate GPCR function. As such these regions display some of the characteristics of intrinsically disordered regions (IDRs).

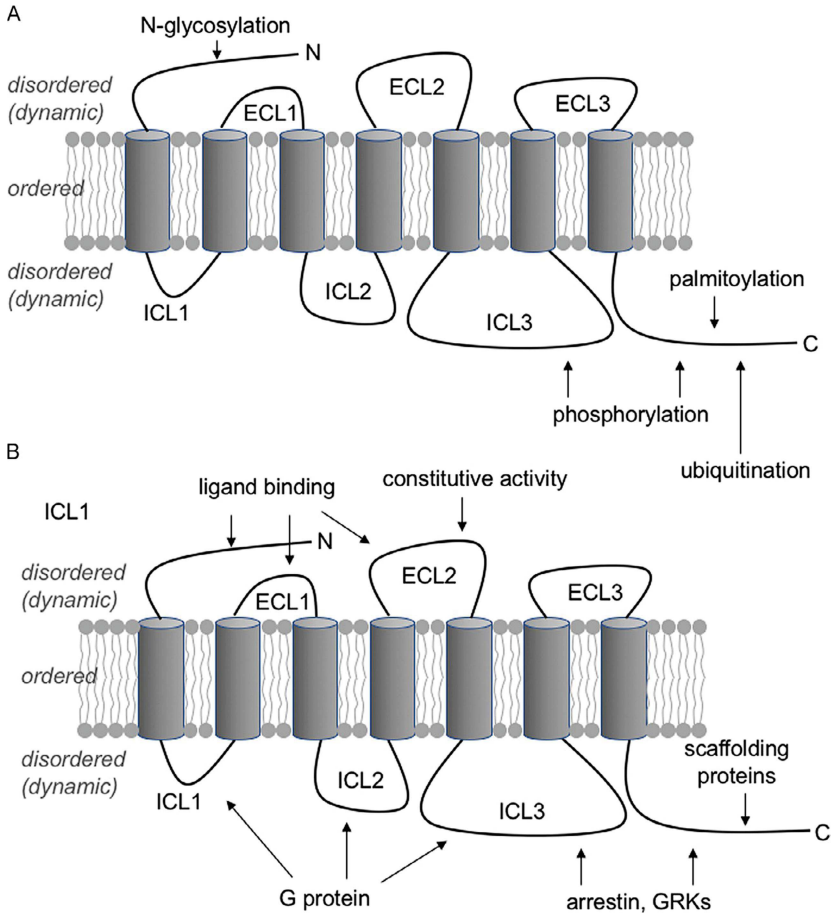


Fig. 2 Schematic representation of class A GPCR structural and functional organization. GPCRs are primarily organized in two kinds of domains, the ordered transmembrane helices and the dynamic extra- and intracellular loops and N- and C-terminal extensions. The general (but not restrictive) nature of PTMs (A) and binding properties (B) in the extra-membrane regions are indicated.

2.3.1 Intrinsic disorder

Using the FoldIndex program, Jaakola and co-workers[62] concluded that these regions show the characteristic patterns of IDRs. Accordingly, in the case of the dopamine receptor for instance, the occurrence of repeats in the region encoding the third intracellular loop cause this region to have a high disorder index.[78] Of importance, the amino-acid content in these unstructured regions is significantly different from that found in disordered proteins, with a particularly high content in basic (R, K) and histidine residues.[62] This specific composition is probably directly related to the functional properties of these regions (*e.g.* interaction with signaling partners and scaffolding proteins). Interestingly, independently of their intrinsic properties, the structural and dynamics features of these regions also appear to be intimately linked to the dynamics of the transmembrane domains. For instance, the structural arrangement of the C-terminal domain of the β2-adrenergic[48] or

vasopressin[79] receptors is dependent on the activation state triggered by pharmacologically distinct ligands. Specifically, biased ligands triggered a different arrangement of the C-terminal end of the receptor relative to the transmembrane domain. It is then tempting to speculate that the arrangement of the C-terminal domain occurs as in concerted movement with the dynamics of the TM helices, ultimately modulating the interaction with cognate cytoplasmic proteins.

2.3.2 Post-translational modifications

GPCRs undergo a full series of post-translational modifications (PTM) that primarily involve the unstructured extra-membrane domains, in particular the intracellular loops and C-terminal domain. Initially considered as an ancillary process, PTM now appear as central to GPCR functioning and regulation. Accordingly, dysregulation of the PTM machinery has been linked to pathological conditions, *e.g.* nephrogenic diabetes insipidus[80] or obesity[81]. The major GPCR modifications include glycosylation, ubiquitination, lipidation or phosphorylation, in addition to other far less described, more limited modifications (*e.g.* tyrosine sulfation). However, the impact of these modifications on the structure of GPCRs is largely unexplored, as most modification sites were removed in many structural studies, in particular in crystallization assays, to increase the homogeneity of receptor preparations.

GPCR glycosylation has been recently reviewed in details.[82] Briefly, glycosylation, and in particular N-glycosylation, essentially affects the N-terminal region of GPCRs and is primarily involved in their folding and transport to the cell surface.[83-87] As such, it may indirectly affect signaling by modulating the number of receptors present at the cell surface. In addition, glycosylation can have a direct impact on receptor functioning. For instance, activation of the β2-adrenergic receptor by pathogens has been shown to involve traction forces on the receptor N-terminus *via* N-glycan chains.[88] GPCRs can also be ubiquitinated, and ubiquitination has been proposed to regulate GPCR signaling through degradation of cell-surface receptors.[89] In addition, this modification could also directly impact on signaling, as demonstrated for instance in the case of CXCR4-dependent MAPK activation.[90] Lipidation of GPCRs essentially involves addition of a palmitoyl chain on a conserved cysteine of the receptor C-terminus. This results in the anchoring of this region to the membrane, creating an additional intracellular loop. As such, it modulates the geometrical

features of this region and, as a consequence, the receptor signaling and internalization properties.[91]

Phosphorylation actually appears as one of the most important PTM in the context of GPCR-dependent signal regulation. This has pushed forward, during the last decade, the development of novel analytical techniques and strategies, based in particular on mass spectrometry and specific antibodies, to identify residues subjected to phosphorylation under a variety of conditions.[92] Phosphorylation occurs in response to agonist-mediated receptor activation and is carried out by different classes of kinases. These essentially include second-messenger dependent protein kinases (PKA, PKC), and GRKs. Both phosphorylate specific Ser and Thr residues on the extra-membrane, intracellular regions of GPCRs. We will put the focus here on GRKs, as they appear as a class of kinases that phosphorylate GPCRs specifically.

GRK is a family composed of seven members (1 to 7) that are subdivided into three subfamilies (GRK1/7, GRK2/3 and GRK4/5/6).[93] They share a common structural organization (Fig. 3A) composed of a specific and conserved N-Terminal domain (αN), a discontinuous Regulator of G-protein signaling Homology domain (RH), a kinase domain (KD) belonging to the family of AGC kinases, and a C-terminal domain (CT) containing specific structural elements to each subfamily that is responsible for their membrane targeting.[56] GRKs require induced conformational rearrangements to become active (Fig. 3C). In the inactive conformation of GRKs the two N- and C-lobes of the kinase domain are splayed apart ('open conformation') compared to active AGC kinase structures ('closed conformation'),[94,95] and is stabilized by the interaction between the RH and the kinase domains.[96,97] The activation of GRKs is promoted by docking to activated GPCRs.[98,99] Indeed, in the model of GRK5/β2AR complex,[97] the binding of β2AR induce GRK5 conformational changes caused by the disruption of a transient 'ionic lock' between the RH and the kinase domains, and by the closure of the catalytic domain ('closed conformation'). The stabilization of this active state of GRKs involves two conserved elements (Fig. 3B): the C-tail extension of the kinase domain (AGC C-tail) and the αN domain. During the activation and thus the closure of the kinase domain, the central part of the AGC C-tail (called the active site tether (AST)) and the αN domain become more ordered.[100-102] Indeed, the αN domain that folds in an αN-helix and the AGC C-tail

bring together the N- and C-lobes of the KD, stabilizing it in its 'closed' and active conformation (Fig. 3C).[96,101,102] We see here again, the functional importance of the disorder to order transition for the kinase activation and receptor phosphorylation.[103-107]

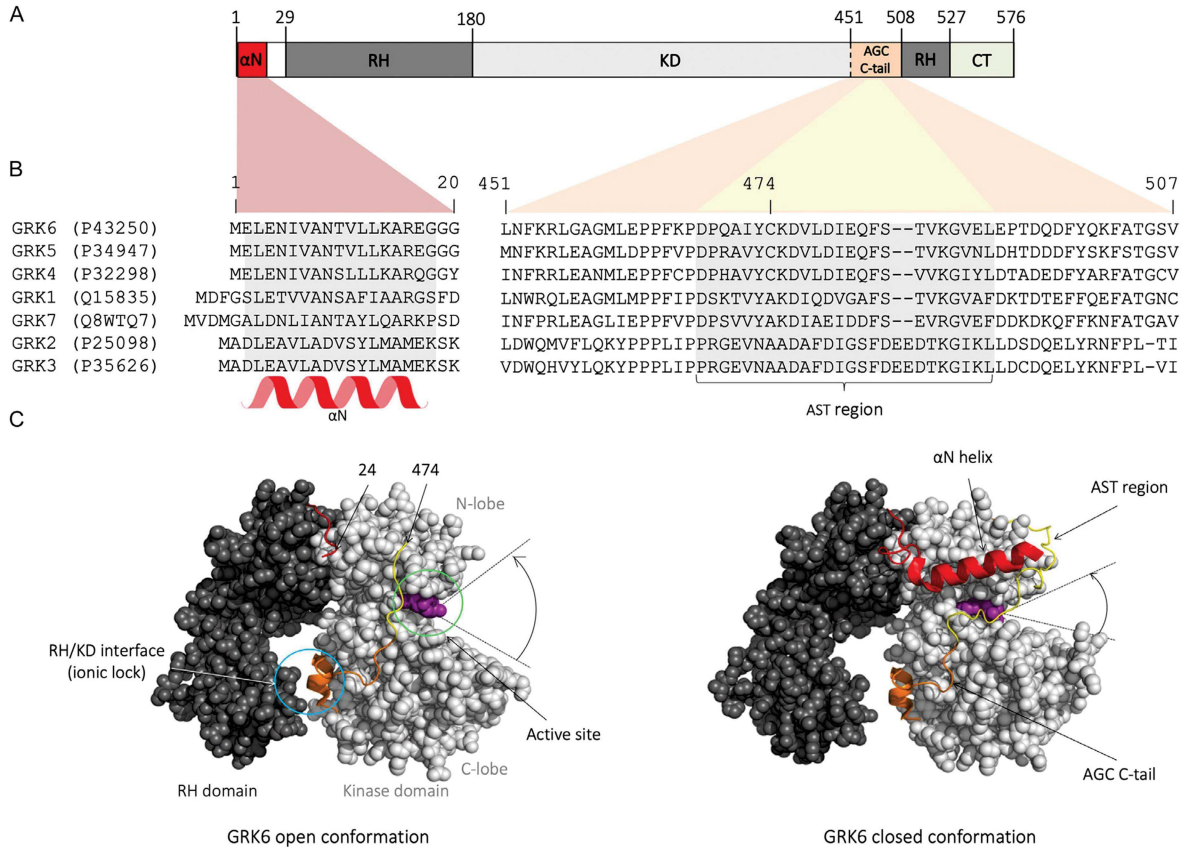


Fig. 3 Importance of the conserved αN domain and AGC C-tail in GRK activation. (A) GRKs share a common structural organization. Residue numbers correspond to those of human GRK6.[108] The N-terminal domain (αN, red) is followed by a regulator of G-protein signaling homology domain (RH, dark grey) surrounding a kinase domain (KD, light gray) and its AGC C-tail (orange). The C-terminal domain (CT, green) contains elements for membrane targeting specific to each subfamily.[56] (B) BLASTp **alignment and conservation of** the αN and the AGC C-tail domains of human [GRK100,105] (sequence numbering based on GRK6 and Uniprot codes of human GRKs are in brackets). The αN helix and the AST region are shown below the alignment. (C**)** Structures of GRK6 in open (PDB 2ACX, on the left)[108] and closed (PDB 3NYN, on the right)[102] conformations. The αN domain and the AGC C-tail are represented in cartoon, the numbering on the left figure correspond to the end and the beginning of segments 1-23 and 475-491 which are absent of the electron density. C-terminal domains (CT) are not represented. The active site (green circle) contains adenylyl-imidodiphosphate (AMP-PNP, ATP analogue) and sangivamycin (nucleoside analogue, kinase inhibitor) (in purple) in open and closed conformation, respectively and its closure is represented by the angle between the N- and C-lobes of the kinase domain. The RH/KD interface containing the ionic lock (blue circle).

Phosphorylation of GPCRs essentially occurs in their third intracellular loop and C-terminal region, two regions that are central to the coupling to signaling partners. It is required for recruitment of arrestins, which ultimately leads to either receptor internalization or activation of alternate signaling pathways. Importantly, besides modulating the interaction with signaling partners, phosphorylation of the GPCR C-terminal tail could also affect the way it interacts with the membrane, and therefore its structural and dynamics features.[109] Even if it was initially considered that the most important feature for arrestin recruitment was the number of the phosphate residues and their spacing in the C-terminal domain,[110] GPCR phosphorylation appears to be a more complex process. First, the different GRK isoforms may regulate distinct, specific physiological functions of GPCRs.[111] In addition, these proteins appear to be able to specifically interact with different conformational states of the receptors. For instance, GRK2 can phosphorylate the D2 dopamine receptor without G protein activation.[112] In this context, an emerging model is that of the "phospho-barcode" that states that distinct receptor conformations trigger interaction with different GRKs that in turn phosphorylate the receptor at distinct sites, establishing a "barcode" that ultimately dictates the functional outcome. This model was established by R. Lefkowitz and co-workers for the β2-adrenergic receptor.[111] It has since then been extended to a full series of GPCRs including GPR120[113] and the ghrelin receptor[114].

2.3.3 Ligand binding and coupling to signaling partners

As stated above, besides being the site for PTMs, in particular phosphorylation, the extramembrane regions of GPCRs display one of the main characteristics of intrinsically disordered regions, i.e. interaction with cognate protein partners. This is the case, at least to some extent, for the extracellular regions. First, the extracellular loops can regulate the constitutive, ligand-independent, activity of some GPCRs such as the ghrelin receptor[115] and the orphan receptor GPR52[116]. In the latter case, the second extracellular loop directly plays the role of an intrinsic agonist for this otherwise orphan receptor. In addition, these regions can participate to ligand recognition and binding. This can occur through a direct interaction with the ligand, as demonstrated in particular for the chemokine[117-119] or adenosine[120] receptors. They can also play the role of a funnel delineating the binding pathway of ligands, as shown for the leukotriene[121] and ghrelin[122] receptors. Finally, the extracellular loops, in particular ECL2, can also affect indirectly the binding parameters for ligands by

playing the role of a lid that precludes dissociation of compounds from their binding site within the transmembrane domains.[123]

Such a role as a "binding platform" for associated partners is even more relevant for the intracellular parts of GPCRs. First, the C-terminus appears to serve as a platform for assembling a full series of scaffolding proteins, in particular the so-called GPCR Interacting Proteins (GIPs), through the recognition of small motifs on the C-terminal regions of GPCRs. This leads to the formation of large functional protein:protein complexes involved in signaling. The role of these assemblies is particularly well-documented in the physiology and pathology of the central nervous system.[124] The intracellular loops of GPCRs are also involved in the interaction with G proteins. Initially demonstrated through mutagenesis and chimeric exchange experiments,[125] this has been further directly visualized in the cryoEM structures of the receptor:G protein complexes. Although the ICL3 loop is poorly resolved in some structures, direct contacts between the three intracellular loops and the G protein α and β subunits have been observed in essentially all these structures, as reviewed recently.[126] In the same way, hydrogen/deuterium exchange mass spectrometry experiments with the isolated β_2-adrenergic receptor and Gs indicated a specific involvement of the receptor intracellular loops in the interaction with the G protein, with specific residues in the second loop playing a central role in the stabilization of the complex between the β2-adrenergic receptor and nucleotide free Gs.[127] The intracellular loops and C-terminal end are also major players in the interaction with arrestins. As stated above, the GPCR C-terminus is selectively phosphorylated by GRKs, and this PTM is the starting point for recruitment of arrestin and further activation of this protein.

Arrestins are briefly described here as, in contrast to G proteins, their structural features and conformational changes upon coupling to GPCRs may share some characteristics with IDPs. Arrestins are a family of increasing importance in GPCR signaling. This family is composed of four members (arrestin1[128,129], arrestin2[130], arrestin3[131] and arrestin4[132]) that share a similar 3D structure (Fig. 4). Arrestins encompass a N- and C-terminal lobe around a polar core composed of charged amino acids: Asp26, Arg169, Lys170, Asp290, Asp297, and Arg393 (numbering based on arrestin2). Arrestin inactive conformation is maintained by a salt bridge between Arg169 and Asp290,[133] and a 'three-element' (TE) interaction constituted of

an α helix (H1), a β strand (BI) in the N-terminal lobe and a β strand (BXX) in the C-tail of arrestin (Fig. 4).[134] The formation of a β-sheet between the two strands of the TE lays this inactive state in a 'closed' conformation. Upon binding of the phosphorylated GPCR C-terminal domain, arrestin undergoes a conformational change to its active state[135-139]: (1) a twisting of the N and C-lobes (~20°); (2) a movement of the finger loop (that can be inserted in the core of the receptor, see below) and of the middle loop; (3) a release of arrestin C-tail; (4) a disruption of the polar core and (**5**) a conformational change of HI, BI and BII. The release of the flexible arrestin C-tail during its transition from an inactive to an active form, allows an exposition of binding sites for its partner interactions that trigger downstream signaling pathways.[140]

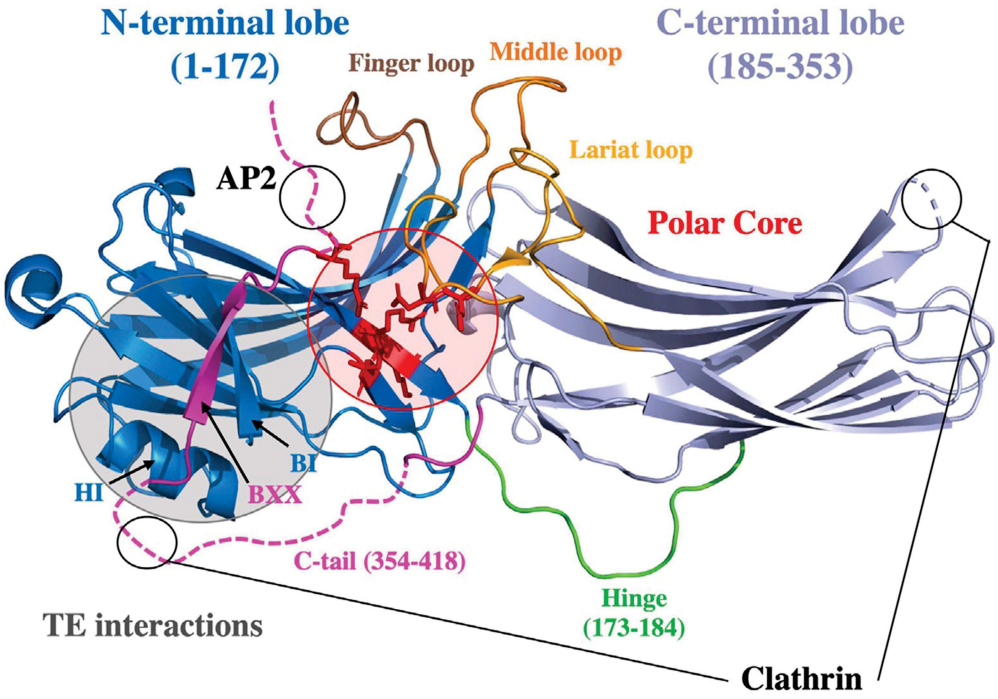


Fig. 4 Common structural organization of arrestin family (PDB:1G4M/X-ray structure of arrestin2 inactive state of [130]). Arrestin family shares similar elongated shape and present two concave lobes mostly composed of antiparallel ß-sheets: a N-terminal lobe (blue; residues 1-172) and a C-terminal lobe (light blue, residues 185-353), linked by a hinge in green (residues 173-184); flexible loops: finger loop (brown), middle loop (orange) and lariat loop (light orange). The polar core is highlighted by a red circle, mostly composed of charged amino-acid (in red, main and side chain of polar core residues) and the three-elements (TE) interaction highlighted by a grey circle involving β-strand 1 and 20 (BI and BXX) and α-helix 1 (HI). Arrestin C-tail (pink, residues 354-418) encompasses β-strand 20 (BXX) (which interact with the N-terminal lobe in the 'closed' state). Concerning non-visual arrestin isoform, its C-tail contains AP2 binding site (^{391}FARQRLK^{397}; not present in the structure) and one clathrin binding site (^{376}LIELD^{380}; not present in the structure), the second clathrin binding site is localized in a flexible loop of C-terminal lobe (black circles).[142]

More precisely, arrestin C-tail regions, which are present in the electron density of the inactive state X-ray structure (residues 354-357 and 383-393) (Fig. 4), are conversely absent in the active form X-ray structures (complex of arrestin2 cross-linked with a synthetic phosphorylated peptides mimicking the C-terminal domain of V2[137] or GPCR-arrestin complexes[71-73,141]). Indeed, in this active 'open' state, the disruption of TE interactions and the release of the C-tail bring some disorder and flexibility in the complex. The aforementioned exposed arrestin C-tail's binding sites allow the interaction with downstream effectors such as clathrin and AP2 (Fig. 4), which are important for the internalization of GPCR.[140,142] These binding sites are conserved motifs, where the AP2 binding site is ^{391}FARQRLK^{397},[140,143] and the clathrin binding site motif is LøXø(D/E),[142] (where ø corresponds to a hydrophobic residue and X to any polar residue). A second clathrin binding motif $(L/I)_2$GXL has been found in a flexible loop of the C-terminal lobe (Fig. 4).[142,144] It was suggested that this second clathrin binding site, which is overlapped with a membrane binding site, would enhance the stability of the interaction between the receptor and its partner arrestin.[144]

With regard to regions of GPCRs involved in the interaction with arrestins, in the recent cryoEM structures of the neurotensin receptor in complex with arrestin, the receptor C-terminus binds the N-terminal lobe of arrestin, a region that contains all the basic amino-acids that interact with phosphorylated residues of the receptor.[71] In the same way, in the complex between the muscarinic receptor and arrestin, the arrestin N-lobe engages the phosphorylated V2R C-terminal domain of the chimeric receptor, with critical phosphorylated residues of the receptor in interaction with basic amino-acids of arrestin.[72] These features are fully consistent with the previous crystal structure of arrestin associated to a phosphorylated peptide from the vasopressin V2 receptor C-terminal domain.[137] In this structure also, the peptide phosphates directly interact with arrestin through charge-charge interactions. Of importance, in these cryo-EM structures, not only the C-terminus of the receptors is in interaction with arrestin, but also the intracellular loops participate to complex formation. In the case of the neurotensin receptor, arrestin engages the central core and third intracellular loop of the receptor in addition to its C-terminus.[71] In the same way, arrestin interdomain loops engage the receptor core and its second intracellular loop in the complex with the muscarinic receptor.[72] Such an interaction pattern is fully consistent

with the model where arrestins were proposed to interact with GPCRs in a "two-site" manner, with a first binding site corresponding to phosphorylated C-terminus of the receptors and the other to the GPCR central core, including the intracellular loops.[73,145] Importantly, each of these interactions could be associated to a distinct functional outcome, thus expanding the repertoire for modulating GPCR-mediated signaling.

2.4 Therapeutic implications

So far, most if not all therapeutic interventions targeting GPCR involved the ligand binding sites, either orthosteric or allosteric. However, the unstructured parts of GPCRs can also be responsible for disfunction of the signaling systems, as evidenced by the fact that several diseases are associated to mutations in the extra- or intracellular, unstructured parts of the receptors.[146] It cannot be excluded that this effect is due to the impact of such mutations on the three-dimensional arrangement of the receptor, *e.g.* by affecting the relative arrangement of thee TM helices. However, these mutations could also alter the interaction pattern with specific signaling partners, as these are all regions that are responsible for protein:protein contacts in the signaling complexes. The extra-membrane regions of GPCRs are far less conserved than the transmembrane domains. Hence, this could pave the way for more specific interventions affecting protein:protein contacts in the signaling complexes. However, so far, if one excepts the pepducin approach that is based on peptides mimicking to intracellular loops of GPCRs and has been proposed for treating some physiological dysfunctions,[147,148] the regions lying outside the membrane have been largely unused in terms of therapeutical applications. Hence, protein:protein interaction (PPI) inhibitor strategies aimed at exploiting the structural and sequence diversity of these regions to regulate the composition and functioning of the signaling complexes could constitute an original, alternate, avenue in drug discovery.

2.5 Conclusion

In addition to many GPCR ligands where conformational disorder is an intrinsic feature of the binding process,[122,149,150] receptors themselves include a substantial number of their residues outside the membrane that present many of the characteristic of intrinsically disordered regions (IDRs). These include lack of defined

structure and high conformational dynamics, interaction with multiple protein components of the signaling complexes, and post-translational modifications that specifically regulate these interactions. As such, they represent major players in the signaling processes. However, so far, their structural and functional properties are largely unexplored, in particular because of these conformational dynamics and labile structural and functional elements. Hence, future research in molecular pharmacology is also to be directed to illuminate our present understanding of the structural and functional relevance of GPCR loops and N- and C-terminal extensions in the context of signaling.

1.2 LA VOIE DE SIGNALISATION DEPENDANTE DE L'ARRESTINE

La revue précédente a permis de présenter toutes les informations majeures concernant la signalisation des récepteurs couplés aux protéines G (RCPGs). Je vais maintenant présenter plus en détail la signalisation des RCPGs dépendante de l'arrestine.

1.2.1 Les arrestines visuelles et les β-arrestines

Chez les mammifères, la famille des arrestines est composée de quatres isoformes : l'arrestine-1 (anciennement nommée antigène S ou 'protéine de 48 kDa ou arrestine des bâtonnets) (Wilden *et al*, 1986), l'arrestine-2 et l'arrestine-3 (ou β-arrestine-1 et -2) (Attramadal *et al*, 1992) et l'arrestine-4 (ou X-arrestine ou arrestine des cônes) (Murakami *et al*, 1993; Craft *et al*, 1994). Les arrestines partagent ~ 50 % d'identité de séquence (Table 1) et une organisation structurale commune (Chapitre 1, revue p. 19, Figure 4). Selon leur séquence et leur fonction, les arrestines sont divisées en deux sous familles : les arrestines-1 et -4 visuelles, qui sont exprimées spécifiquement dans la rétine (cellules bâtonnets et cônes, respectivement), et les arrestines-2 et -3 non-visuelles qui sont exprimées ubiquitairement et partagent 76 % d'identité de séquence (Table 1). De ce fait, les β-arrestines régulent la transduction du signal des RCPGs non-visuels.

	Arrestine-1	Arrestine-2	Arrestine-3	Arrestine-4
Arrestine-1	X	57.4	53.4	49.0
Arrestine-2		X	76.1	58.7
Arrestine-3			X	53.8
Arrestine-4				X

Table 1 : Les arrestines partagent ~ 50 % d'identité de séquence. Les valeurs représentent le pourcentage d'identité entre protéines. Le pourcentage d'identité des arrestines visuelles est coloré en bleu et celui des arrestines non-visuelles est coloré en vert. Les alignements ont été effectués avec BlastP (blast.ncbi.nlm.nih.gov) à partir des codes UniProt des arrestines humaines : arrestine-1 : code P10523 ; arrestine-2 : code P49407 ; arrestine-3 : code P32121 ; arrestine-4 : code P36575.

1.2.2 Activation de l'arrestine

Dans son état basal, la conformation de l'arrestine est stabilisée par deux sets d'interactions impliquant la queue C-terminale de l'arrestine : l'interaction des trois éléments (TE) et l'interaction du cœur polaire (Figure 1) (Hirsch *et al*, 1999; Seyedabadi *et al*, 2021). L'interaction des trois éléments implique des résidus hydrophobes de l'hélice alpha I (HI) et du brin bêta I (BI) du domaine N-terminal, ainsi que le résidu I386 de la queue C-terminale de l'arrestine (BXX). Le cœur polaire est composé de résidus chargés qui forment des interactions électrostatiques entre le domaine N-terminal, la boucle gate (partie de la boucle lariat) et la queue C-terminal de l'arrestine.

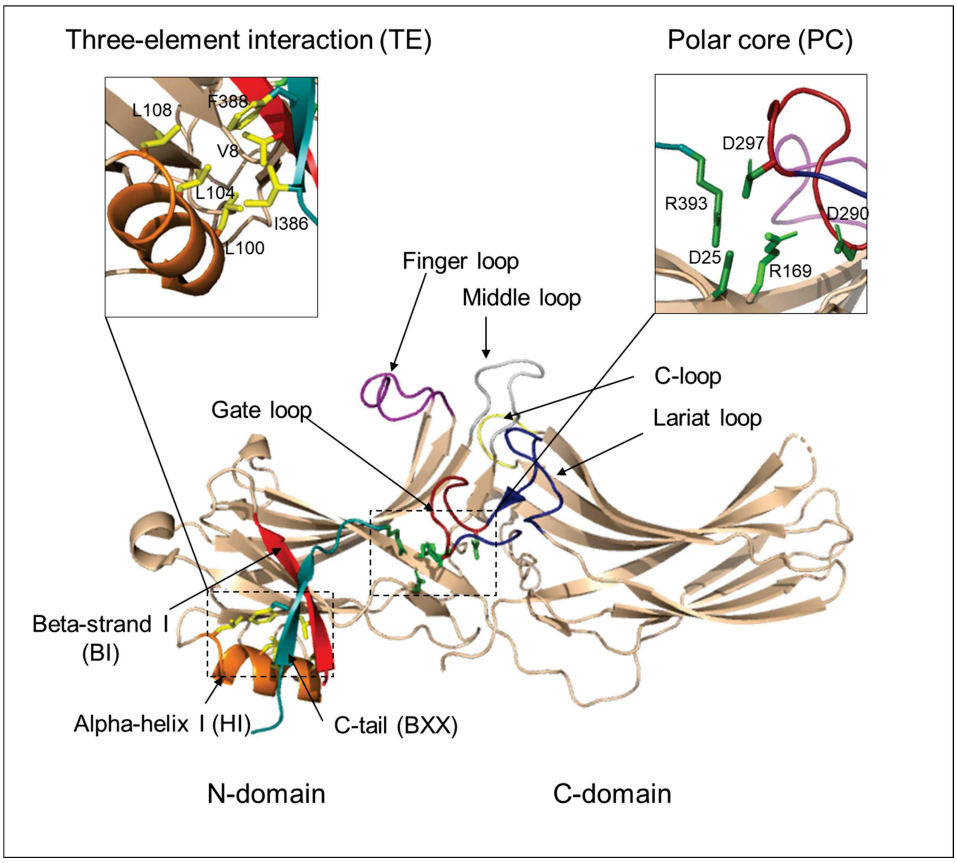


Figure 1 : Conformation basale de l'arrestine-2 (Figure modifiée à partir de (Seyedabadi *et al*, 2021)). La conformation de l'arrestin-2 basale (PDB : 1G4M (Han *et al*)) est stabilisée par les interactions hydrophobes des trois éléments (TE, encart de gauche) et les interactions électrostatiques du cœur polaire (PC, encart de droite). Les boucles fonctionnellement importantes et les résidus critiques sont indiqués et mis en évidence comme suit : queue C-terminale (dont le brin BXX), bleu claire ; hélice alpha I (HI), orange ; brin bêta I (BI), rouge ; boucle gate, rouge foncé ; boucle finger, violette ; boucle middle, grise ; boucle C, jaune et boucle lariat, bleu foncée ; chaîne latérale des résidus formant les interactions des trois éléments, jaune ; chaîne latérale des résidus formant le cœur polaire, verte.

L'interaction de l'arrestine avec le domaine C-terminal phosphorylé des RCPGs entraîne un déplacement de la queue C-terminale de l'arrestine et, par conséquent, une destruction des interactions des trois éléments et du cœur polaire (Shukla *et al*, 2013). Ainsi, les interactions qui stabilisent la conformation basale sont détruites, ce qui permet notamment une rotation du domaine C-terminal par rapport au domaine N-terminal (Figure 2).

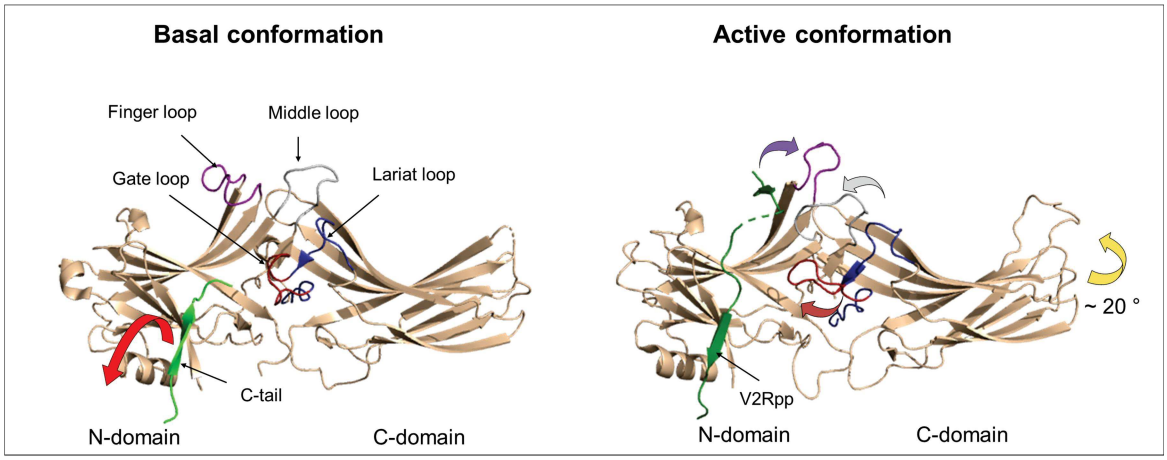


Figure 2 : Changement conformationnel de l'arrestine en complexe avec les RCPGs (Figure modifiée à partir de (Gurevich & Gurevich, 2019b)). Les structures proviennent des PDB 1G4M (Han *et al*) pour la conformation basale et 4JQI (Shukla *et al*, 2013) pour la conformation active. Lors de l'interaction, la queue C-terminale de l'arrestine (vert clair) est déplacée (la flèche rouge indique la direction du mouvement) par le domaine C-terminale du récepteur V2R (vert foncé). Les boucles de l'arrestine subissent aussi des déplacements qui sont indiqués par le même code couleur que celui des boucles : finger, violette ; middle, grise ; lariat, bleu et gate, rouge foncé. La rotation du domaine C-terminal par rapport au domaine N-terminal est représentée par la flèche jaune.

Parmi les autres changements, de nombreuses boucles de l'arrestine, dont les boucles finger, middle et lariat, sont déplacées (Figure 2). Ces changements structuraux, décrits à partir de structures cristallographiques, sont certainement dynamiques en solution et permettraient d'exposer différents sites de liaison de l'arrestine à ses autres partenaires cytosoliques, déclenchant différentes fonctions des arrestines.

1.2.3 Fonction des arrestines

Les arrestines ont de nombreuses fonctions (Ahn *et al*, 2020) qui peuvent être simplifiées en trois fonctions principales : la désensibilisation des récepteurs, le trafic

intracellulaire ainsi que l'internalisation des récepteurs, et la transduction de signaux par interaction avec d'autres partenaires cytosoliques (Figure 3).

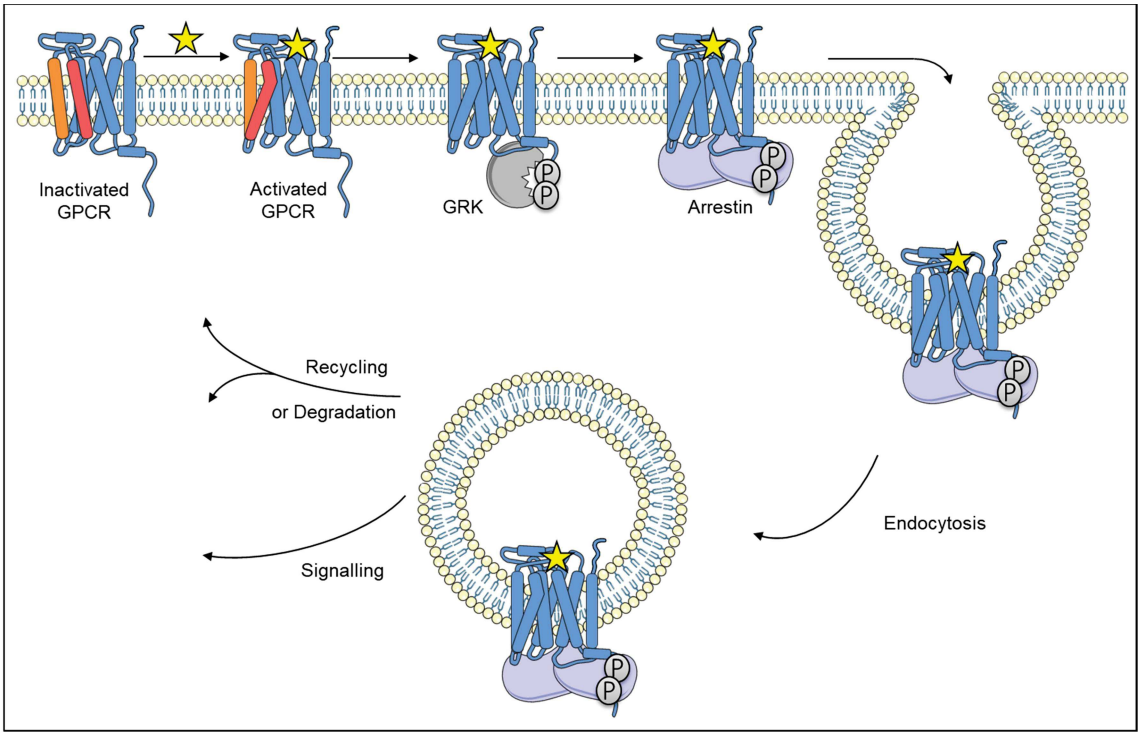


Figure 3 : Représentation schématique des fonctions de l'arrestine dans la signalisation des RCPGs. Figure modifiée à partir de (Luttrell et al, 2001). Dans le paradigme classique, la liaison de molécules extracellulaires (agoniste, étoile jaune) aux RCPGs mène à des changements conformationnels du récepteur. Parmi ces changements, les hélices transmembranaires 5 (orange) et 6 (rouge) sont déplacées et conduisent à l'ouverture d'une cavité inter-hélicoïdale (ou sillon) sur le côté cytoplasmique du récepteur. Ainsi activé, le RCPG est phosphorylé par les kinases GRKs, puis il interagit avec la protéine arrestine. La liaison de l'arrestine déclenche alors la désensibilisation et l'internalisation du récepteur ainsi que l'activation de voie de signalisation, comme la voie ERK (kinase régulée par le signal extracellulaire).

1.2.3.1 La désensibilisation des récepteurs

Historiquement, la première fonction attribuée aux protéines arrestines est la désensibilisation des RCPGs. La désensibilisation des récepteurs consiste à stopper la signalisation dépendante des protéines G après une première activation grâce à une action combinée des GRKs et des arrestines (Smith & Rajagopal, 2016; Caron & Barak, 2019). Elle a été décrite pour la première fois en 1986, où la protéine 'arrestine' a 'arrêté' la signalisation de la voie G de la rhodopsine phosphorylée par GRK1 (Wilden *et al*, 1986). Peu de temps après, il a été montré que le récepteur β2-adrénergique (β2AR) voyait sa capacité à activer la voie G être diminuée de 80 %

après phosphorylation par un extrait partiellement purifié de GRK2 contenant un analogue de l'arrestine-1 qui sera par la suite appelé β-arrestine (Benovic *et al*, 1987). Depuis, il est accepté que la désensibilisation des récepteurs est universelle et est dépendante des kinases GRKs et des protéines arrestines (Lohse *et al*, 1992; Kohout & Lefkowitz, 2003).

Suite à la découverte du rôle des arrestines dans la désensibilisation des récepteurs, le mécanisme par lequel cette fonction opère a été étudié. Benovic et son équipe (Krupnick *et al*, 1997b) ont montré que la liaison de la protéine G au RCPG décroît au fur et à mesure que l'arrestine se lie au récepteur lors d'expériences de liaison compétitive. Cette compétition entre arrestine et protéine G a été confirmée lorsque des structures atomiques obtenues par cristallographie aux rayons X (RX) et par cryo-microscopie électronique (cryo-EM) ont montré que ces deux effecteurs possèdent un site de liaison commun au récepteur (Szczepek *et al*, 2014; Chaturvedi *et al*, 2020). Par exemple, l'hélice α5 de la protéine G et la boucle 'finger' de l'arrestine occupent une position similaire dans le sillon formé par le basculement des hélices transmembranaires 5 et 6 (TM5 et TM6) du récepteur à la neurotensine (NTSR1) sous forme activée (Figure 4).

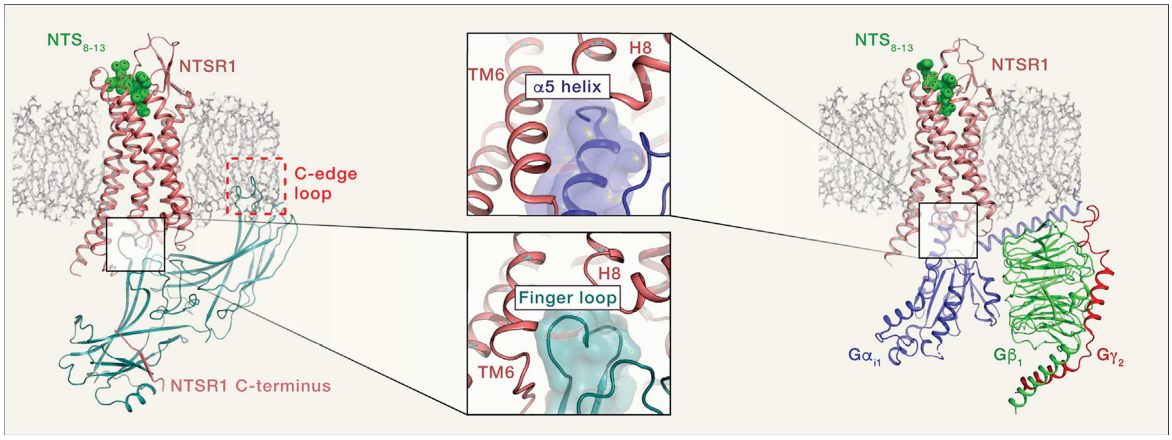


Figure 4 : Désensibilisation des récepteurs par compétition pour le site de liaison entre les protéines G et les arrestines. (Figure extraite de (Chaturvedi *et al*, 2020)). Le panel de gauche montre la structure cryo-EM du complexe NTSR1:arrestine-2 (PDB : 6UP7). Le panel de droite montre la structure cryo-EM du complexe NTSR1:$G_{\alpha i1}$ (PDB : 6OS9). Le panel du milieu indique que l'hélice α5 de la protéine G et la boucle finger de l'arrestine partage une interface d'interaction similaire au domaine centrale du récepteur.

Ainsi, la désensibilisation des récepteurs serait due à une exclusion stérique de la protéine G par la liaison de l'arrestine au récepteur phosphorylé.

1.2.3.2 Le trafic intracellulaire et l'internalisation des récepteurs

Lorsque les récepteurs sont désensibilisés, ils sont moins réactifs à une nouvelle simulation. Par conséquent, il existe un mécanisme permettant leur réactivation, nommé resensibilisation, qui implique la déphosphorylation des récepteurs. La séquestration des récepteurs dans les endosomes (von Zastrow & Kobilka, 1992) est nécessaire à leur déphosphorylation par la phosphatase GRP (G protein-coupled receptor phosphatase), membre de la famille des Protéines Phosphatases-2A (PP-2A) (Pitcher *et al*, 1995). En effet, le pH acide des endosomes permet d'altérer la conformation du récepteur et sa déphosphorylation par la GRP (Krueger *et al*, 1997).

Les RCPGs peuvent être internalisés pas différents mécanismes de séquestration (pour revue (Moore *et al*, 2007)). Cette section présente uniquement le mécanisme d'endocytose des récepteurs associé à l'arrestine.

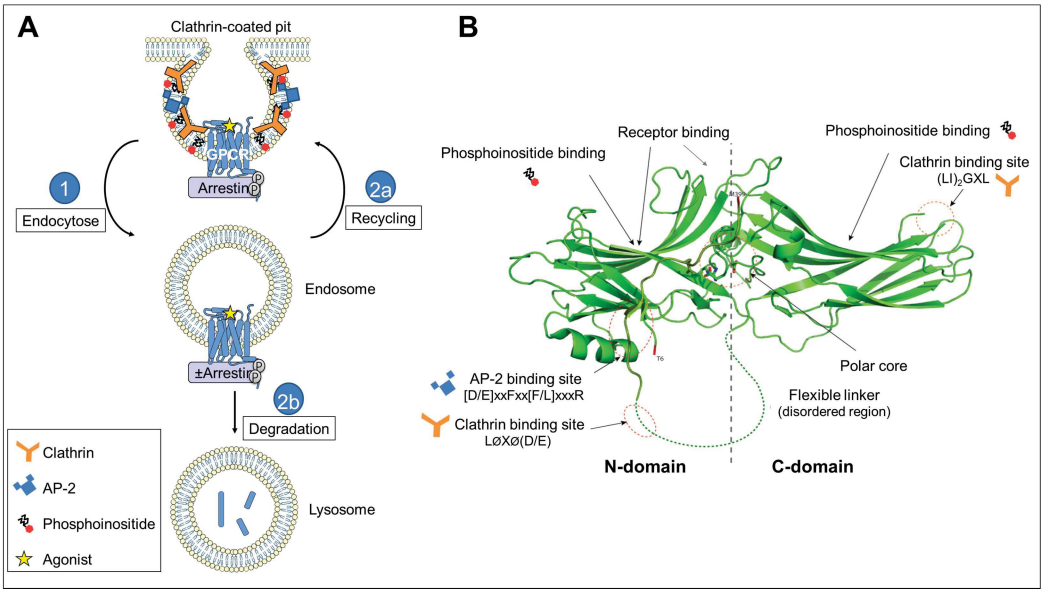


Figure 5 : L'arrestine sert d'adaptateur à la séquestration des récepteurs dans les endosomes. (Figure modifiée à partir de (Moore *et al*, 2007). **A)** Les récepteurs sont dirigés vers les vésicules recouvertes de clathrines grâce à l'interaction de l'arrestine avec les trois éléments du système endocytique : les clathrines (orange), l'adaptateur AP-2 (bleu) et les phosphoinositides (rouge et noir). Cela permet l'endocytose du récepteur en complexe (+) ou non (-) avec l'arrestine (1). Le récepteur internalisé peut ensuite être recyclé à la membrane (2a) ou dégradé dans les lysosomes (2b). **B)** Sites d'interaction de l'arrestine-2 avec les trois éléments du système endocytique (Figure extraite de (Kang *et al*, 2014)). La liaison aux clathrines se fait grâce au motif conservé LØXØ(D/E) (où Ø est un acide aminé aliphatique) (Krupnick *et al*, 1997a) et au motif $(LI)_2GXL$ (Kang *et al*, 2009). La liaison à l'adaptateur AP-2 se fait via le motif conservé [D/E]xxFxx[F/L]xxxR (Edeling *et al*, 2006). La liaison aux phosphoinositides se fait grâce à des résidus basiques dont K233, R237 et K251 (Gaidarov *et al*, 1999). Il est intéressant de noter que l'arrestine-1 ne contient pas les motifs nécessaires à ces interactions, à l'exception du motif $(LI)_2GXL$, ce qui est en accord avec l'absence apparente d'internalisation de la rhodopsine.

Il a été montré que les récepteurs sont généralement dirigés vers les vésicules recouvertes de clathrine grâce à l'interaction des β-arrestines avec plusieurs éléments composant le système endocytique : les protéines clathrines, l'adaptateur AP-2 ainsi que des molécules de phosphoinositides (Figure 5) (Kang *et al*, 2009; Edeling *et al*, 2006; Gaidarov *et al*, 1999).

Curieusement, le suivi de l'arrestine marquée par la protéine fluorescente verte (GFP) en cellule HEK (cellules embryonnaires humaines issues du rein) a permis de mettre en évidence que l'arrestine n'internalise pas toujours avec le récepteur (Zhang *et al*, 1999). Cette observation a conduit à la formation de deux classes de RCPGs, classe A et B, basées sur leur caractéristique de liaison à l'arrestine (Oakley *et al*, 2000). Les récepteurs de classe A, dont le récepteur β2-adrénergique (β2AR) et le récepteur à la ghréline (GHSR), se lie plus fortement à l'arrestine-3 qu'à l'arrestine-2. Ces récepteurs forment des complexes transitoires avec l'arrestine qui se dissocient à proximité de la membrane plasmique. Ainsi l'arrestine n'internalise pas avec le récepteur. À l'inverse, les récepteurs de classe B, dont le récepteur à la vasopressine V2 (V2R), interagissent avec la même affinité pour les deux β-arrestines et forment des complexes plus stables permettant l'internalisation du complexe entier.

Après internalisation, les récepteurs sont soit recyclés à la membrane, soit dégradés dans les lysosomes (Figure 5). Ainsi, l'arrestine participe à la régulation du trafic intracellulaire des RCPGs et à leur internalisation (Moore *et al*, 2007).

1.2.3.3 La transduction de signaux par assemblage multi-protéique

Avec plus de 400 partenaires identifiés à ce jour, les β-arrestines sont des protéines dites « hub » (Crépieux *et al*, 2017) qui servent d'échafaudage protéique en aval des RCPGs. Ainsi, elles sont impliquées dans de très nombreuses fonctions comme la croissance et la migration cellulaire, la transcription ou encore le réarrangement du cytosquelette (Ahn *et al*, 2020). Parmi ces interactants certains sont capables d'interagir avec l'arrestine sous forme libre, étendant le rôle des arrestines à des signalisations indépendantes des RCPGs. Cette section présente l'implication de l'arrestine dans la régulation de la voie de signalisation ERK (kinase régulée par le signal extracellulaire).

La voie de signalisation ERK est une des trois voies de signalisation MAPK (protéine kinase activée par des mitogènes) impliquées dans la croissance cellulaire et la survie. Elle est composée d'une cascade de phosphorylation partant de la kinase Raf (protéine proto-oncogène pouvant induire des fibrosarcomes à accélération rapide), qui phosphoryle la kinase MEK-1/2 (MAPK/ERK kinase-1/2) qui phosphoryle à son tour la kinase ERK-1/2 (ou MAPK). Ainsi activée, ERK peut phosphoryler des facteurs de transcription impliqués dans la progression du cycle cellulaire, ou des protéines cytosoliques (Robinson & Cobb, 1997).

La β-arrestine, en complexe avec le RCPG, active la voie ERK de différentes manières (Figure 6). Elle peut interagir avec la protéine Src (protéine kinase proto-oncogène pouvant induire un sarcome), ce qui permet l'activation indirecte de la kinase Raf, et l'activation subséquente des voies de signalisation nucléaire de ERK (Luttrell *et al*, 1999; Ahn *et al*, 2020).

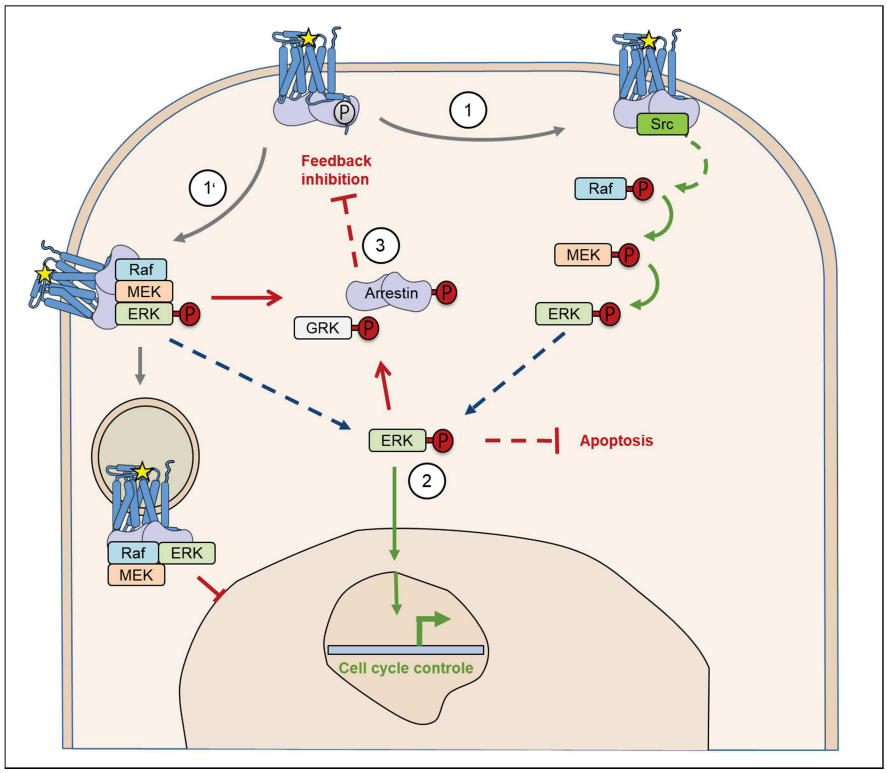


Figure 6 : Implication de l'arrestine dans la croissance cellulaire et la survie via la voie ERK (Figure modifiée à partir de (Ahn *et al*, 2020)). L'arrestine participe à l'activation de la voie ERK par interaction avec la protéine Src (1 sur le schéma) ou par échafaudage protéique des protéines Raf, MEK et ERK (1'). Suite à son activation, ERK va promouvoir la division cellulaire (2) et la phosphorylation de protéines cytosoliques, comme les GRKs et les arrestines (3, mécanisme d'auto-régulation).

À l'inverse, la formation d'un complexe multi-protéique formé des kinases Raf, MEK et ERK en interaction avec l'arrestine bloque la translocation de ERK dans le noyau (Luttrell *et al*, 2001; Ahn *et al*, 2020). Cela peut favoriser la phosphorylation de protéines cytosoliques comme les GRKs (Pitcher *et al*, 1999) et les arrestines (Lin *et al*, 1999), ce qui atténuerait l'activation de ERK par des mécanismes de rétro-inhibition.

1.2.4 La phosphorylation des RCPGs par les GRKs

1.2.4.1 Caractéristiques principales des GRKs

1.2.4.1.1 La famille des GRKs

La phosphorylation du domaine C-terminal et/ou des boucles intracellulaires des récepteurs par les kinases GRKs (kinase spécifique des RCPGs) permet la liaison de l'arrestine au RCPG (Seyedabadi *et al*, 2021). Il existe 7 GRKs différentes, numérotées de 1 à 7 en fonction de leur ordre de clonage et classées en trois sous-familles selon leur séquence et leur fonction : GRK1/7, GRK2/3 et GRK4/5/6 (Premont *et al*, 1999). Les GRKs partagent au minimum ~ 30 % d'identité de séquence entre elles mais ce chiffre augmente au sein des sous-familles avec 46.7 % d'identité entre GRK1 et 7, ~ 68 % d'identité entre GRK4, 5 et 6 et 84 % d'identité entre GRK2 et 3 (Table 2).

	GRK1	GRK2	GRK3	GRK4	GRK5	GRK6	GRK7
GRK1	X	33.4	33.5	47.7	47.8	48.3	46.9
GRK2		X	84.0	36.4	39.1	38.7	31.2
GRK3			X	36.5	39.2	38.9	32.2
GRK4				X	68.0	66.6	47.9
GRK5					X	72.3	45.9
GRK6						X	46.5
GRK7							X

Table 2 : Les GRKs partagent au minimum 30 % d'identité de séquence. Les valeurs représentent le pourcentage d'identité entre protéines. Les pourcentages d'identité au sein des sous-familles sont colorés en bleu pour GRK1/7, en vert pour GRK2/3 et en rose pour GRK4/5/6. Les alignements ont été effectués avec BlastP (blast.ncbi.nlm.nih.gov) à partir des codes UniProt des GRKs humaines : GRK1 : Q15835 ; GRK2 : P25098 ; GRK3 : P35626 ; GRK4 : P32298 ; GRK5 : P34947 ; GRK6 : P43250 ; GRK7 : Q8WTQ7.

1.2.4.1.2 Organisation structurale et mécanisme d'activation

Les GRKs partagent une organisation structurale commune qui subit des changements conformationnels lors de leur activation (Chapitre1, revue, p. 19). Dans la revue nous avons expliqué que la conformation active des GRKs est caractérisée par un rapprochement des deux lobes du domaine catalytique (N-/petit lobe et C-/large lobe) (Figure 7) (Komolov *et al*, 2015, 2017). Ce rapprochement est stabilisé par l'interaction du domaine catalytique avec deux autres régions des GRKs : le domaine N-terminal (domaine αN) et la région AST (ancrage du site actif ; élément régulateur situé à proximité du site d'ancrage du substrat dans le site actif). De facon interessante, ces deux régions sont des IDRs qui sont repliées uniquement dans la conformation active de la kinase (Singh *et al*, 2008; Boguth *et al*, 2010).

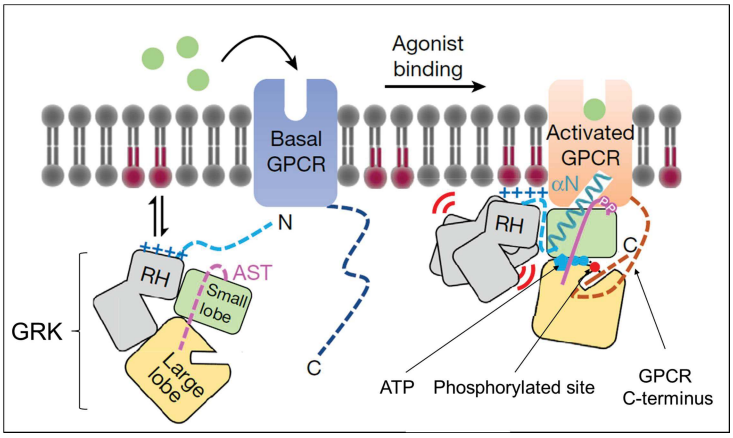


Figure 7 : Modèle d'activation des GRKs par la liaison au récepteur activé (Figure modifiée à partir de (Chen *et al*, 2021)). A l'état initial, les lobes du domaine catalytique (rectangles vert et jaune) sont dans une conformation ouverte, éloignés l'un de l'autre. Le domaine αN (bleu) et la région AST (rose) sont désordonnés. Après activation du RCPG, le domaine αN forme une hélice (αN) qui interagit avec le sillon cytoplasmique du récepteur. Cette interaction consolide l'hélice αN et ses interactions avec l'AST et les lobes du domaine catalytique, ce qui stabilise la GRK dans une conformation fermée, catalytiquement compétente. Le complexe RCPG:GRK serait aussi stabilisé par des interactions ioniques entre les lipides chargés négativement (rouge foncé) et des acides aminés chargés positivement du domaine RH (gris, domaine d'homologie des protéines régulant la signalisation des protéines G).

Même s'il est connu que la liaison des GRKs aux RCPGs favorise leur activation (Palczewski *et al*, 1991; Chen *et al*, 1993), le mécanisme par lequel les récepteurs contribuent à cette activation restait imprécis jusqu'à très récemment. Cette année, la toute première structure d'un récepteur en complexe avec une GRK a été obtenue grâce à des réticulations chimiques ('cross-linking') et à l'ajout d'anticorps stabilisant

le complexe (Chen *et al*, 2021). Ce complexe rhodopsine:GRK1 valide l'implication du domaine αN et de la région AST dans la fermeture du domaine kinase. De plus, la structure de cryo-microscopie électronique montre une interaction entre le domaine αN et le sillon cytoplasmique du récepteur activé qui pourrait stabiliser le repliement du domaine αN en hélice. La conservation de l'interface rhodopsine:GRK1 chez d'autres récepteurs et d'autres kinases a permis aux auteurs de proposer un modèle d'activation générale des GRKs (Figure 7). Dans ce modèle, le récepteur contribue à l'activation des GRKs par stabilisation du domaine αN en hélice.

1.2.4.1.3 Spécificité des GRKs

Il est connu qu'une GRK peut phosphoryler plusieurs RCPGs et qu'un même RCPG peut être phosphorylé par différentes GRKs (Inglese *et al*, 1993; Carman & Benovic, 1998). Cependant, encore à ce jour, les sites phosphorylés par les différentes GRKs sont peu connus et le mécanisme par lequel les GRKs phosphorylent spécifiquement tel ou tel site reste un mystère. La spécificité des protéines kinases est généralement dominée par les résidus acides, basiques ou hydrophobes adjacents au résidu phosphorylé (Blom *et al*, 2004). Partant de ce postulat de nombreuses études ont cherché à identifier les préférences des GRKs pour différents substrats peptidiques, selon le taux d'intégration de phosphates radioactifs. Ces études ont montré que GRK1 et GRK2 préfèrent les substrats présentant des acides aminés acides aux alentours du site phosphorylé (Onorato *et al*, 1991; Asai *et al*, 2014). À l'inverse, GRK5 et GRK6 préfèrent les substrats ayant des résidus basiques voisinant le site de phosphorylation (Kunapuli *et al*, 1994; Loudon & Benovic, 1994).

Il a aussi été proposé que les GRKs phosphorylent de manière spécifique les récepteurs en fonction du type de cellules et de tissus (Butcher *et al*, 2011). Cela pourrait être associé aux localisations cellulaires de chaque GRKs. Les GRKs 1 et 7 sont spécifiques du système visuel (Kühn & Dreyer, 1972; Weiss *et al*), alors que GRK4 est exprimée plus largement dans le cervelet, les testicules et les reins (Premont *et al*, 1996). Les quatre autres GRKs (GRK2, 3, 5 et 6) sont exprimées dans de nombreux tissus avec différents niveaux d'expression (Figure 8) (Benovic *et al*, 1991; Benovic & Gomez, 1993; Kunapuli *et al*, 1994).

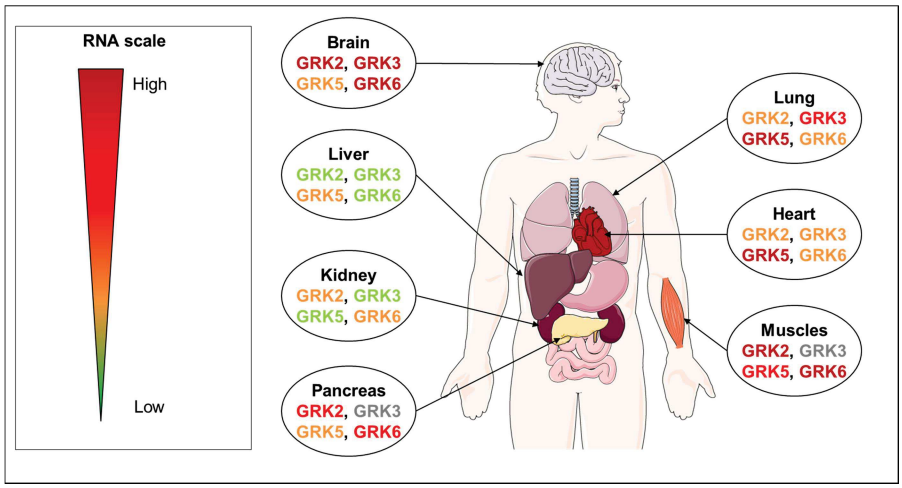


Figure 8 : Expression différenciée des GRKs ubiquitaires (selon (Benovic *et al*, 1991; Benovic & Gomez, 1993; Kunapuli *et al*, 1994)). Les ARNs (Acide RiboNucléiques) des GRKs ont été isolés à partir de différents organes et quantifiés pour représenter la distribution tissulaire des GRKs. Pour chaque GRKs, la quantité d'ARN est représentée par un gradient de couleur : du moins présent (vert) au plus présent (rouge foncé). L'ARN de GRK3 n'a pas été quantifié dans le pancréas et les muscles (indiqué en gris).

En plus d'une expression dépendante du type cellulaire, les GRKs peuvent aussi présenter des localisations sub-cellulaires différentes. Par exemple, dans les neurones humains, GRK2 et GRK3 n'ont pas de localisation préférentielle, alors que GRK5 et GRK6 sont préférentiellement localisées aux membranes synaptiques (Bychkov *et al*, 2011). La raison de cette localisation particulière des GRKs n'est pas connue. Cependant, une distribution similaire a été identifiée dans le cerveau du rat (Ahmed *et al*, 2008), suggérant que cette localisation subcellulaire spécifique est importante sur le plan fonctionnel.

1.2.4.2 Lien entre phospho-barcode et code conformationnel de l'arrestine

Etant donné que la phosphorylation par les GRKs est un prérequis à la liaison de l'arrestine aux RCPGs, Lefkowitz et son équipe se sont penchés sur la contribution des différentes GRKs dans la signalisation dépendante de l'arrestine pour différents récepteurs dont le récepteur β2-adrénergique (β2AR) (Kim *et al*, 2005; Ren *et al*, 2005; Nobles *et al*, 2011). Pour le récepteur β2AR, ils ont utilisé une stratégie de 'perte-de-fonction' basée sur l'utilisation de petit ARN interférents (siRNA) pour supprimer les fonctions de GRK2 et/ou GRK6 en cellules HEK stimulée par l'isoprotérénol (agoniste qui active la voie G et la voie arrestine) (Nobles *et al*, 2011). Les auteurs ont choisi les cellules HEK car dans ces cellules seules GRK2 et GRK6

régulent le recrutement de l'arrestine-3 au récepteur β2AR (Violin *et al*, 2006). Ils ont montré que l'activation du récepteur β2AR entraîne la phosphorylation des sites T360, S364, S396, S401, S407, et S411 par GRK2 et des sites S355 et S356 par GRK6 (Figure 9). Ces motifs sont respectivement impliqués dans l'internalisation du récepteur (sites GRK2), dans l'activation de la voie ERK (sites GRK6) et dans la désensibilisation du récepteur (sites GRK2 et 6) (Figure 9).

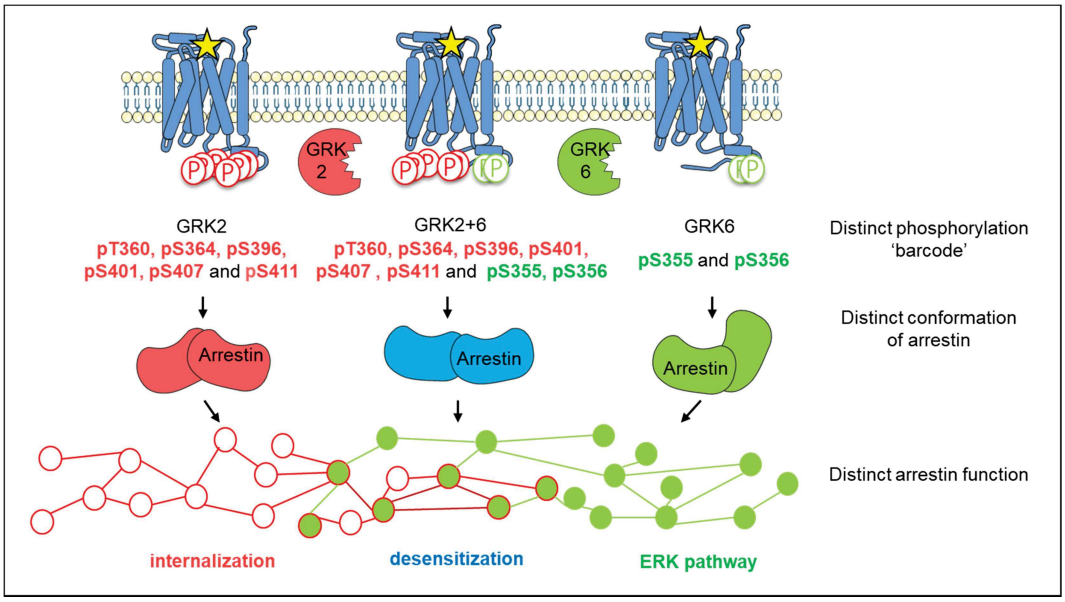


Figure 9 : Représentation schématique du phospho-barcode du récepteur β2-adrénergique (β2AR) (Figure modifiée à partir de (Wisler *et al*, 2014)**.** Les kinases GRK2 (rouge) et GRK6 (verte) phosphorylent différents sites du récepteur β2AR après activation par l'isoprotérénol (étoile jaune). Ces motifs de phosphorylation (barcode) dirige la conformation de l'arrestine et sont impliqués dans l'internalisation du récepteur (sites GRK2, rouge), dans l'activation de la voie ERK (sites GRK6, vert) et dans la désensibilisation du récepteur (sites GRK2 et 6). Ce modèle a été obtenu grâce à l'utilisation d'ARNs interférents contre GRK2 et/ou GRK6 en cellules HEK.

L'effet de ces motifs de phosphorylation sur la conformation de l'arrestine a aussi été étudié par transfert d'énergie par résonance de bioluminescence (BRET). Le signal BRET a été mesuré à partir de la protéine fluorescente Renilla Luciferase (RLuc) et de la protéine fluorescente jaune (YFP) fusionnées aux extrémités N- et C-terminales de l'arrestine-3 (RLuc-arr3-YFP). Après traitement avec les différents siRNAs, la mesure du signal BRET a permis de mettre en évidence que l'arrestine adopte des conformations différentes selon la suppression des phosphorylations de GRK2 ou de GRK6. Ces résultats indiquent que les motifs de phosphorylation dirigent les conformations et les fonctions de l'arrestine en complexe avec le récepteur. De plus,

ils ont montré que l'activation du récepteur par le carvedilol, agoniste biaisé qui active uniquement la voie arrestine, entraîne seulement la phosphorylation des sérines S355 et S356, sites phosphorylés par GRK6.

En conclusion de cette étude, Lefkowitz et son équipe proposèrent un modèle, le modèle du phospho-barcode, pouvant expliquer la signalisation différenciée de l'arrestine en fonction de l'activation des RCPGs (Figure 10). Dans ce modèle, l'activation du récepteur par divers ligands engagerait des GRKs spécifiques qui pourraient phosphoryler différents sites des domaines C-terminaux des RCPGs. Ces motifs de phosphorylation, ou 'phospho-barcode', formeraient ainsi un 'code-barres' qui indiquerait ou déterminerait la conformation à adopter par l'arrestine et ses capacités fonctionnelles (Figure 10, (Nobles *et al*, 2011; Liggett, 2011)).

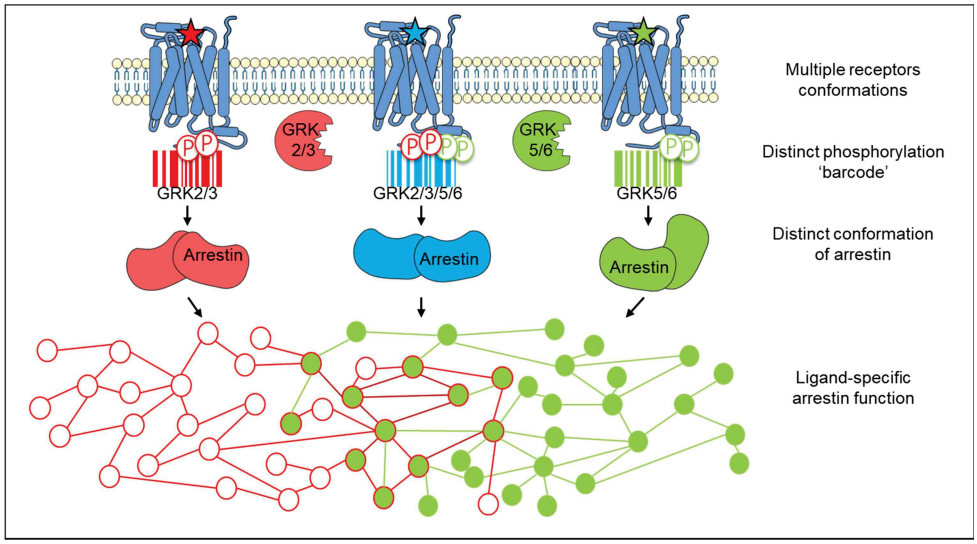


Figure 10 : **Représentation schématique du modèle du phospho-barcode** (Figure modifiée à partir de (Wisler *et al*, 2014). L'activation des RCPGs par différents ligands stabiliserait des conformations distinctes du récepteur. Ces conformations uniques recruteraient des GRKs spécifiques qui phosphoryleraient des sites particuliers. La localisation de ces sites formerait un motif de phosphorylation, ou 'barcode' qui stabiliserait une conformation spécifique de l'arrestine liée. Cela déclencherait une fonction particulière de l'arrestine, propre à cette conformation (exposition d'une surface d'interaction spécifique à un effecteur particulier de la cellule).

Depuis lors, le phospho-barcode modèle a été étendu à de nombreux RCPGs, comme le récepteur à la ghréline (présenté Chapitre 4, p. 217). Aujourd'hui, le modèle du phospho-barcode est devenu le modèle d'interaction des RCPGs avec l'arrestine.

1.2.5 L'interface RCPG:arrestine

Complexe	Phosphorylation du C-terminus du RCPG	références
V2Rpp:arrestine-2:Fab30 V2R=vasopressine 2 (343-371)	Synthèse chimique sur les sites : pT347, pS350, pS357, pT359, pT360, pS362, pS363, pS364	PDB : 4JQI (RX) 2.6 Å (Shukla *et al*, 2013)
T4L-β2V2R:arrestine-2:Fab30 T4L-β2V2R=fusion T4 lysozyme –β2AR(1-341)-V2R(343-371)	Co-expression en cellules d'insectes avec GRK2 Pas d'indication concernant la phosphorylation	(CN-EM) (Shukla *et al*, 2014)
rhodopsine:ArrFL-1 ArrFL-1= boucle 'finger' de l'arrestine 1 (67-77)	Phosphorylation *In vitro* par GRK1 Pas d'indication concernant les sites phosphorylés	PDB : 4PXF (RX) 2.7 Å (Szczepek *et al*, 2014)
T4L-rhodopsine-arrestine-1 Protéine de fusion	Expression de la protéine de fusion en cellules HEK pT336, pS338	PDB : 5W0P (RX) 3.0 Å (Zhou *et al*, 2017)
T4L- β2V2R:arrestine-2:Fab30:Nb32	Expression du récepteur en cellules d'insectes Domaine C-terminale non visible dans la structure	(CN-EM) (Cahill *et al*, 2017)
T4L-β2V2R:arrestine-2:Fab30:Gs:Nb35	Co-expression en cellules d'insectes avec GRK2 pS357, pT359, pT360, pS362, pS363 et pS364	PDB : 6NI2 (EM) 4.0 Å (Nguyen *et al*, 2019)
NTSR1:arrestine-2 NTSR1=neurotensine-1	Phosphorylation *In vitro* par GRK5 pS396, pS401, pS403, pS404, pT407, pT413	PDB : 6UP7 (EM) 4.2 Å (Huang *et al*, 2020)
M2Rpp:arrestine-2:Fab30 M2Rpp: fusion muscarinique M2(2-466)- V2R(343-371)	Synthèse chimique sur les sites : pT347, pS350, pS357, pT359, pT360, pS362, pS363, pS364	PDB : 6U1N (EM) 4.0 Å (Staus *et al*, 2020)
C7pp:arrestine-3 C7pp=CXCR7 (331-345)	Synthèse chimique sur les sites : pS335, pT338, pT341	PDB : 6K3F (RX) 2.3 Å (Min *et al*, 2020)
β1AR6P:arrestine-2:Fab30 β1AR6P=β1-adrénergique dinde (32-329)-V2R(355-371)	Synthèse chimique sur les sites : pS357, pT359, pT360, pS362, pS363, pS364	PDB : 6TKO (EM) 3.3 Å (Lee *et al*, 2020)
V2Rpp-1:arrestine-2:Fab30	Synthèse chimique sur les sites : pS350, pS357, pT359, pT360, pS362, pS363, pS364	PDB : 7DF9 (RX) 3.2 Å (He *et al*, 2021)

Table 3 : **Structures de l'arrestine en interaction avec un RCPG.** Les complexes formés sont indiqués dans la colonne de gauche. La colonne centrale montre les sites identifiés comme phosphorylés dans la région C-terminale des récepteurs. La colonne de droite présente : les références (code PDB et auteurs), la résolution de la structure (Å), ainsi que la méthode utilisée pour l'obtention de la structure correspondante : RMN (résonance magnétique nucléaire), RX (cristallographie aux rayons X), CN-EM (microscopie électronique en coloration négative) ou EM (cryo-microscopie électronique). Sauf indication, les numérotations des récepteurs correspondent aux protéines humaines. Les protéines Fab30 (Fab, fragment de liaison à l'antigène), Nb32 et Nb35 (Nb, nanocorps) sont des anticorps ajoutés pour stabiliser les complexes. Les récepteurs sont aussi souvent produits avec le tag T4L (lysozyme du phage T4) pour améliorer leur stabilité.

Le modèle du phospho-barcode postule que les agonistes et les motifs de phosphorylation dirigent la conformation liée de l'arrestine au RCPG. Depuis, plusieurs structures, à résolution atomique, d'arrestines seules, ou en interaction avec des récepteurs, ont été obtenues et mettent en évidence les nombreuses conformations que les arrestines peuvent adopter (Table 3). Ces structures ont permis de comprendre comment les arrestines se lient aux récepteurs phosphorylés et adoptent une conformation active (Seyedabadi *et al*, 2021).

1.2.5.1 Les modes de liaisons de l'arrestine aux RCPGs

1.2.5.1.1 Les deux types d'interaction RCPG:arrestine

L'architecture du complexe RCPG:arrestine a été visualisée pour la première fois en 2014 par microscopie électronique en coloration négative (Shukla *et al*, 2014). Pour cela, les auteurs ont utilisé un récepteur β2-adrénergique chimérique dont le domaine C-terminal a été remplacé par le domaine C-terminal du récepteur vasopressine V2 (β2V2R, Table 3). La commutation du domaine C-terminal de β2AR par celui de V2R permet au récepteur chimérique d'acquérir les caractéristiques d'un récepteur de classe B vis-à-vis de l'interaction à l'arrestine (Chapitre 1, p. 44). Ce récepteur chimérique a donc une meilleur affinité pour l'arrestine et forme des complexes plus stable que β2AR sauvage (wt) (Oakley *et al*, 2000). Avec l'ajout de l'anticorps Fab30 (fragment de liaison à l'antigène), pour stabiliser le complexe β2V2R:arrestine-2, les images en particule unique ont révélé deux conformations.

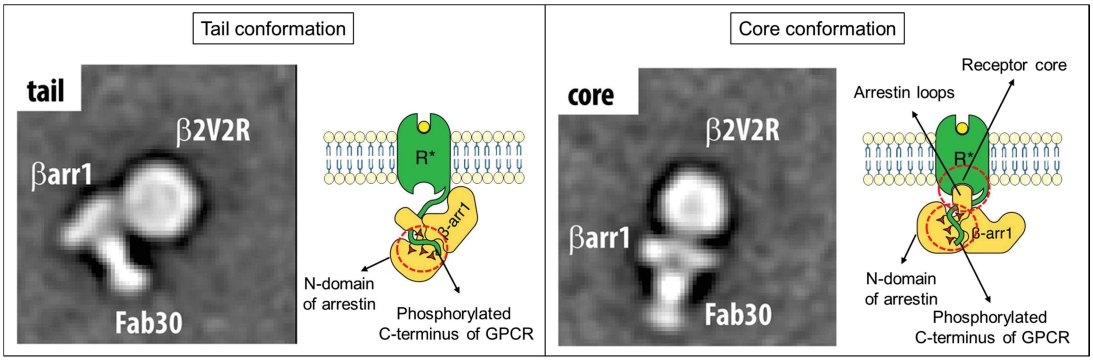


Figure 11 : **Les deux modes de liaison de l'arrestine au récepteur** (Figure modifiée à partir de (Smith & Rajagopal, 2016)et (Shukla *et al*, 2013)). Les images de microscopie électronique en particule unique montrent que le complexe β2V2R:arrestine-2 adopte deux conformations : une conformation 'tail' où le 'senseur de phosphorylation' de l'arrestine (domaine N-terminal) se lie à la région C-terminale du récepteur ; et une conformation 'core' où les 'senseurs d'activation' (boucles de l'arrestine) et le 'senseur de phosphorylation' de l'arrestine interagissent respectivement avec le cœur du récepteur (sillon cytoplasmique) et le domaine C-terminal du récepteur.

Dans la première conformation, l'arrestine s'accroche au récepteur par une interaction ponctuelle impliquant uniquement le domaine C-terminal de V2R fusionné à β2AR (Figure 11). Cette conformation, nommée conformation 'tail', représente 63 % des particules visualisées. Les 37 % restant de particules montrent que l'arrestine se lie à la face cytoplasmique du récepteur activé, formant la conformation 'core' (Figure 11).

En combinant ces observations à des études biochimiques, les auteurs ont obtenu un modèle à basse résolution de la conformation globale du complexe. Ils ont montré que la liaison des arrestines aux RCPGs se fait par deux types d'interaction (Figure 11). La première interaction a lieu entre le domaine N-terminal de l'arrestine et le domaine C-terminal des RCPGs (conformation 'tail') (Figure 11). La deuxième interaction engage les boucles de l'arrestine (boucles finger, middle et lariat, Figure 12) à l'intérieur de la cavité cytoplasmique du récepteur, s'insérant entre les hélices TM3, 5 et 6. Ainsi, le domaine N-terminal de l'arrestine est aussi appelé 'senseur de phosphorylation' et les boucles de l'arrestine sont nommées 'senseurs d'activation'. Dans la conformation 'core' les deux interactions prennent place en même temps (Figure 11).

1.2.5.1.2 Cas des super-complexes (ou mégaplexes)

L'existence de la conformation 'tail', qui implique uniquement la région C-terminale des RCPGs pour l'interaction avec l'arrestine, laisse l'opportunité à une protéine G de se lier simultanément au récepteur via le domaine central de ce dernier, formant un 'super-complexe' ou 'mégaplexe'. Récemment, une structure de cryo-microscopie électronique du mégaplexe Gs:β2V2R:arrestine-2 stabilisé par les anticorps Fab30 et Nb35 a été obtenue et a permis de voir que les trois composants adoptent leur conformation active dans ce complexe (Nguyen *et al*, 2019). Les trois protéines étant actives, il est possible qu'un mégaplexe permette une signalisation simultanée des voies G et arrestines. Cela serait en accord avec les travaux de Lefkowitz et son équipe, qui ont montré l'existence des mégaplexes dans les endosomes des cellules HEK et leur fonction dans la signalisation des protéines G médiée par des RCPGs internalisés (Thomsen *et al*, 2016).

1.2.5.2 Rôle de la cavité cytoplasmique des RCPGs dans la liaison de l'arrestine

L'activation des RCPGs est accompagnée par un mouvement important des hélices transmembranaires (TM) 5 et 6, qui ouvre une cavité côté cytoplasmique du récepteur (Chapitre 1, p. 41, Figure 3). Les boucles de l'arrestine (finger, C-, middle et lariat) sont engagées dans cette cavité, entre les hélices transmembranaires (Figure 12) (Gurevich & Gurevich, 2019b).

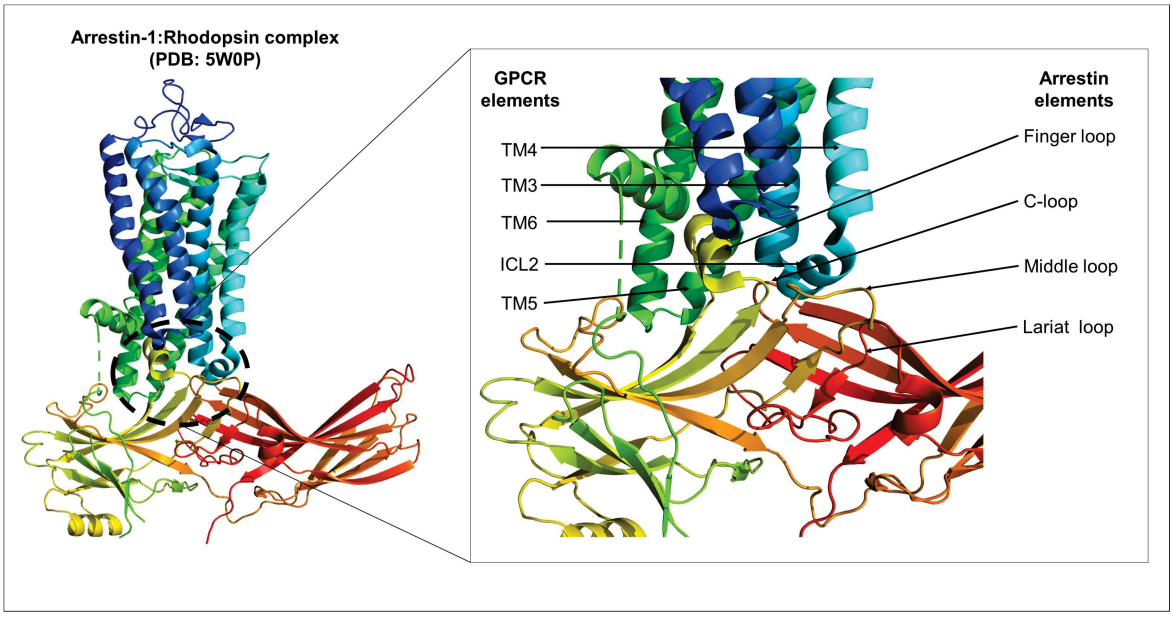


Figure 12 : Interaction des boucles de l'arrestine avec le sillon cytoplasmique du récepteur. La structure de cryo-microscopie électronique de l'arrestine-1 en complexe avec la rhodospine **(Zhou *et al*, 2017)** montre que les boucles 'finger', 'C-' et 'middle' interagissent avec les éléments du domaine central du récepteur, comme la boucle ICL2 et les hélices transmembranaires TM3, 4, 5 et 6.

Ces boucles font toutes parties de l'interface RCPG:arrestine pour les complexes rhodopsine:arrestine-1 (Zhou *et al*, 2017), muscarinique:arrestine-2 (Staus *et al*, 2020) et neurotensine:arrestine-2 (Huang *et al*, 2020) (Table 3). Cependant, ces éléments interagissent avec différentes parties du récepteur (Figure 13). Par exemple, la boucle intracellulaire ICL2 des récepteurs rhodopsine (Rho) et muscarinique (M2R) est prise en sandwich entre les boucles finger et middle de l'arrestine, alors que dans le complexe neurotensine:arrestine-2, ICL2 est localisée sur le côté extérieur de ces boucles (Figure 13).

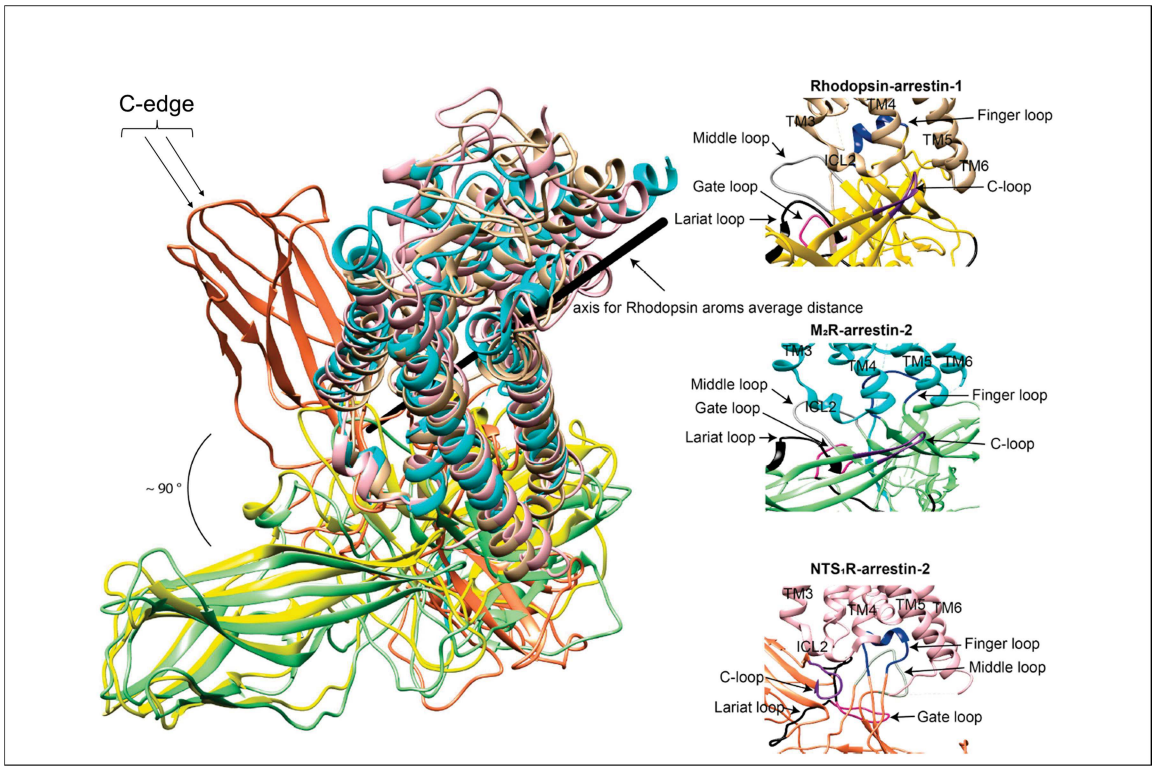


Figure 13 : **Comparaison des complexes RCPG:arrestine** (Figure extraite de (Seyedabadi *et al*, 2021)). A gauche : superposition des complexes rhodopsine:arrestine-1 (jaune:brun ; PDB 5W0P), M2R:arrestine-2 (vert:bleu ; PDB 6U1N) et NTS1R:arrestine-2 (orange:rose ; PDB 6UP7). La superposition a été faite en se basant sur l'alignement de séquences des récepteurs. A droite : zoom sur les boucles de l'arrestine en contact avec le récepteur pour chaque complexe. Les boucles finger (bleue foncé), middle (grise), lariat (noire), gate (rose) et C (violette) sont mises en évidence pour mieux voir les différences de conformations.

La boucle finger de l'arrestine est celle qui forme le plus de contact avec les éléments de la cavité cytoplasmique, dont les hélices TM3, TM4, TM5, TM6 et la boucle ICL2. En effet, 37 % des contacts avec le récepteur β1-adrénergique (β1AR) proviennent de la boucle finger de l'arrestine-2 (Lee *et al*, 2020). Curieusement, cette boucle adopte différentes conformations en complexe avec les récepteurs rhodopsine (Rho), muscarinique (M2R), et neurotensine (NTSR1). Dans les complexes avec Rho et NTSR1, la boucle finger est repliée en hélice alors que dans le complexe muscarinique:arrestine-2, elle forme une boucle étendue (Figure 13). Ces structures ont mis en évidence que la boucle finger est très flexible. Cette flexibilité pourrait lui permettre de s'adapter à différents récepteurs (Seyedabadi *et al*, 2021). En effet, la boucle finger forme des interactions électrostatiques avec des résidus chargés des RCPGs mais la position de ces résidus n'est pas conservée entre les récepteurs. Cela donnerait lieu à des espacements distincts de charges qui

pourraient forcer le positionnement de la boucle finger et affecter également la conformation globale de l'arrestine.

La dernière différence majeure entre ces complexes concerne l'orientation de l'arrestine par rapport au récepteur. Dans le complexe neurotensine:arrestine-2 (Huang *et al*, 2020), les boucles du domaine C-terminal de l'arrestine (C-edge, Figure 13) interagissent avec la micelle dans laquelle est solubilisée le récepteur entrainant un basculement de ~ 40° de l'arrestine par rapport au récepteur. À l'inverse, dans les complexes rhodopsine:arrestine-1 (Zhou *et al*, 2017) et muscarinique:arrestine-2 (Staus *et al*, 2020), l'arrestine est basculée de ~ 15° par rapport au récepteur, soit une différence de 90° entre les domaines C-terminaux des arrestines dans ces différents complexes (Figure 13). Ainsi, l'arrestine pourrait prendre de nombreuses conformations et orientations différentes selon les interactions qu'elle effectue avec le RCPG et la membrane plasmique (Huang *et al*, 2020).

1.2.5.3 Rôle du domaine C-terminal des RCPGs dans la liaison de l'arrestine

Dans ce paragraphe, sauf indication, les numérotations utilisées correspondent à celle de l'arrestine-1 humaine.

1.2.5.3.1 Rôle des motifs de phosphorylation dans l'interaction à l'arrestine

Le modèle phospho-barecode présenté précédement (Chapitre 1, p. 50) montre l'importance de la phosphorylation des domaines C-terminaux des RCPGs pour la liaison à l'arrestine. Ainsi, il doit exister un mécanisme moléculaire expliquant comment les phosphorylations déclenchent une liaison de plus ou moins haute affinité aux récepteurs. Dans cet objectif, l'équipe de Xu (Zhou *et al*, 2017) a comparé les structures cristallographiques de l'arrestine liée au domaine C-terminal phosphorylé des RCPGs, existantes en 2017 : le complexe rhodopsine:arrestine-1 (Zhou *et al*, 2017) et le complexe V2pp:arrestine-2, où V2pp est un phospho-peptide synthétique du domaine C-terminal du récepteur V2R (Shukla *et al*, 2013). Dans ces deux complexes, les domaines C-terminaux phosphorylés sont repliés sous la forme d'un brin bêta qui se lie au feuillet bêta du domaine N-terminale de l'arrestine. Pour cela, les résidus phosphorylés et chargés négativement des récepteurs interagissent avec les résidus chargés positivement des arrestines. Ces résidus lysines et

arginines de l'arrestine sont situés dans des poches très conservées : poche A (K19, R22, R175), poche B (K19, R33 et K304) et poche C (K18 et K114) (Figure 14).

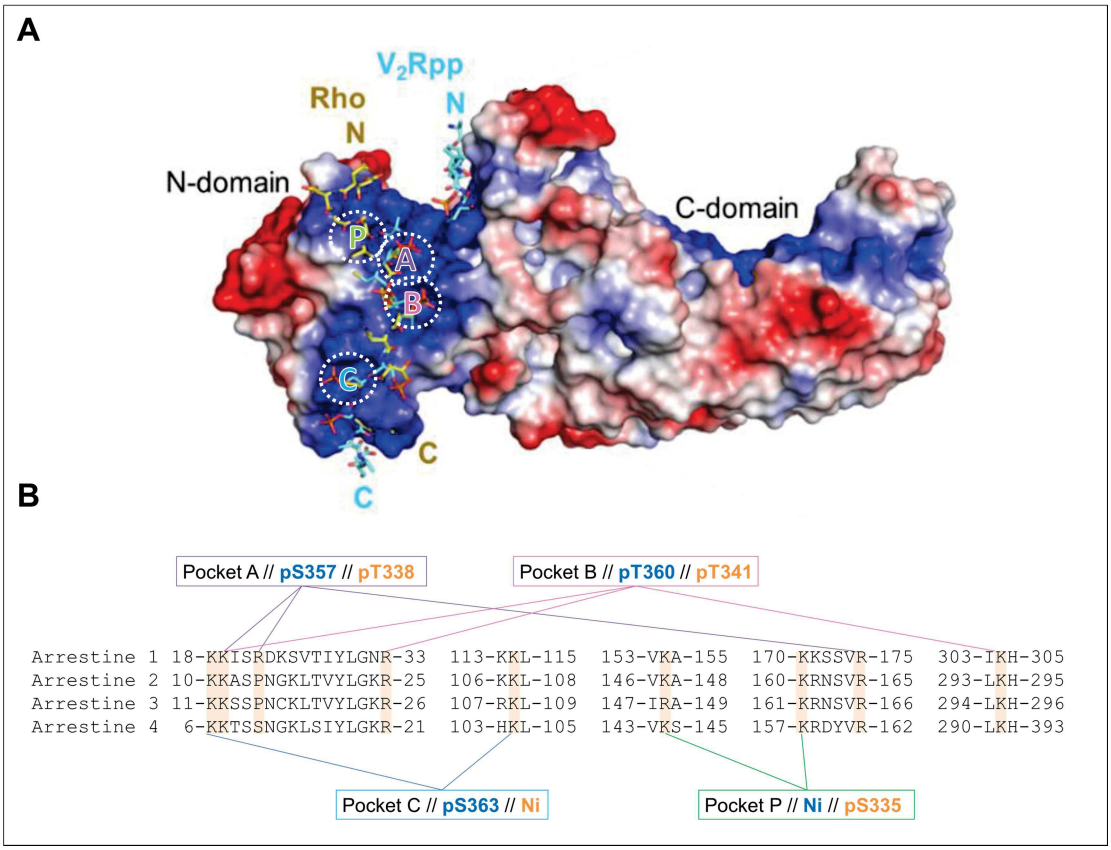


Figure 14 : L'arrestine contient des poches permettant d'accueillir les phosphates des RCPGs. A) Superposition des phospho-peptides de la rhodopsine (PDB 5W0P, jaune) et du V2R (PDB 4JQI, bleu clair) sur le potentiel électrostatique de l'arrestine-2 (Figure modifiée à partir de (Min *et al*, 2020)). Les charges positives sont colorées en bleu et les charges négatives sont représentées en rouge. Les poches de liaison A, B, C et P à la surface de l'arrestine sont entourées en pointillés. **B)** Alignement des séquences des poches de liaison des différentes arrestines humaines. Les sérines et thréonines du récepteur V2R (en bleu) et CXCR7 (en orange) interagissant avec chaque poche sont indiquées dans les encadrés respectifs. 'Ni' signifie 'pas d'interaction'.

La distribution des résidus chargés négativement sur les régions C-terminales des RCPGs a conduit à un modèle général où des motifs de phosphorylation (ou phospho-codes) serviraient de région d'interaction de haute affinité avec l'arrestine (Zhou *et al*, 2017). La liaison à l'arrestine se ferait grâce au motif p[S/T]Xp[S/T]XXp[S/T]/E/D (motif court) ou p[S/T]XXp[S/T]XXp[S/T]/E/D (motif long), où p[S/T] représente une sérine ou une thréonine phosphorylée et X peut être n'importe quel résidu excepté une proline dans la deuxième occurrence (XX). L'absence d'un des trois résidus chargés du motif (p[S/T] ou p[S/T]/E/D), donnerait

lieu à des motifs incomplets, ou partiels, qui auraient une affinité plus faible pour l'arrestine. Plusieurs motifs de phosphorylation sont localisés dans les régions C-terminales des RCPGs et la phosphorylation d'un motif particulier permettrait l'interaction avec l'arrestine dans une conformation spécifique associée à une fonction de signalisation. Ceci permet de donner une description structurale au modèle du phospho-barcode initialement proposé par Lefkowski et son équipe (Nobles *et al*, 2011).

Basée sur ces travaux, l'équipe de Lee (Min *et al*, 2020) a créé un phospho-peptide du récepteur CXCR7 (nommé C7pp) dont les résidus S335, T338 et T341 sont phosphorylés et forment le motif long p[S/T]XXp[S/T]XXp[S/T]/E/D. Ce peptide est capable d'interagir avec l'arrestine-3, et a permis l'obtention d'une structure cristallographique. Dans cette structure, le peptide de CXCR7 forme une boucle allongée et le premier, deuxième et troisième phosphate du motif forment des ponts ioniques avec les arginines R148 (nouvelle poche P), R166 (poche A) et R26 (poche B) de l'arrestine-3 (numérotation de l'arrestine-3 humaine) (Figure 14). Cette nouvelle poche P est aussi conservée parmi les arrestines (Figure 14) et elle serait capable d'accommoder les motifs longs p[S/T]XXp[S/T]XXp[S/T]/E/D via R148 mais aussi les motifs courts p[S/T]Xp[S/T]XXp[S/T]/E/D via K161 (numérotation de l'arrestine-3 humaine).

Ainsi, les travaux de (Zhou *et al*, 2017) et (Min *et al*, 2020) suggèrent que les motifs de phosphorylation permettent l'interaction aux arrestines.

1.2.5.3.2 Du motif de phosphorylation des RCPGs à la conformation de l'arrestine

Les deux études précédentes ont permis de fournir un motif de phosphorylation qui serait nécessaire à l'interaction avec l'arrestine. Cependant, elles n'expliquent pas comment les motifs de phosphorylation régulent la conformation de l'arrestine. Pour cela, il est nécessaire de comprendre comment la contribution individuelle, et les différentes combinaisons, de chaque phosphorylation modifie le comportement de l'arrestine. C'est l'objectif que s'est fixé Yu et son équipe (He *et al*, 2021). Pour cela, ils ont identifié sept poches conservées de l'arrestine-2 (V1 à V7, Figure 15A), qui présentent une interaction avec chacun des phosphates du phospho-peptide synthétique du récepteur V2R (V2Rpp) selon (Shukla *et al*, 2013) (Figure 15A). Ils ont ensuite créé des phospho-peptides synthétiques qui diffèrent de V2Rpp par la position d'un ou deux phosphates (V2Rpp1, V2Rpp3, V2Rpp4 et V2Rpp6-7, Figure

15B), et ils ont obtenu les structures de ces peptides en interaction avec l'arrestine-2, par cristallographie aux rayons X.

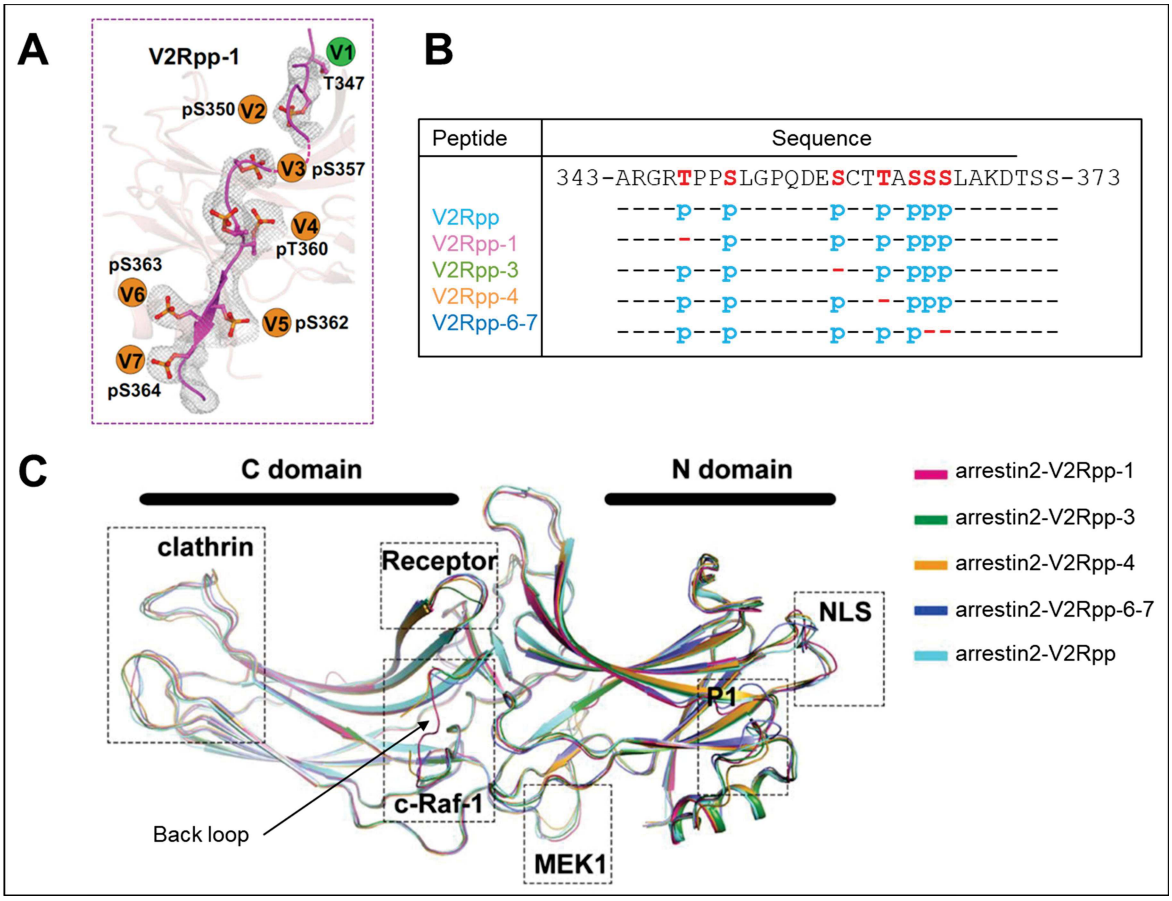


Figure 15 : La localisation des phosphates des RCPGs affecte les interfaces d'interaction de l'arrestine avec ses partenaires (Figure modifiée à partir de (He *et al*, 2021)). **A)** Carte de densité électronique du peptide V2Rpp-1 en complexe avec l'arrestine-2. Les poches de liaison des phosphates sur l'arrestine ont été définies par les auteurs selon la structure du complexe V2Rpp:arrestine2 (PDB 4KQI) : poche V1 (Y71, Q73, R85), poche V2 (R70, K145, K170, R175), poche V3 (K170, R175, K19), poche V4 (K19, K304, R33), poche V5 (V15 et F17), poche V6 (K18, Y29, K114) et poche V7 (K114) (numérotation selon l'arrestine-1 humaine) (Yang *et al*, 2015). **B)** Séquence des phospho-peptides dérivés de V2Rpp. Les sites phosphorylés dans chacun de ces peptides sont présentés par un 'p' bleu en dessous de la séquence. Les tirets rouges représentent les sites qui diffèrent de V2Rpp et qui ne portent pas de phosphate **C)** Superposition des structures de l'arrestine dans les différents complexes V2pp:arrestine-2. Les régions d'interaction de l'arrestine avec ses partenaires sont encadrées : le sillon cytoplasmique des RCPGs activés, les clathrines, les protéines Raf (c-Raf-1), MEK1 et Src (P1) et la séquence de localisation nucléaire (NLS). Ces régions présentent des changements structuraux entre les différents complexes, comme la boucle back dans le complexe V2Rpp-1:arrestine-2.

A l'exception du phosphopeptide V2Rpp-4, comme nous le verrons ci-dessous, les interactions V2pp:arrestine-2 ont toutes été conservées comme décrites dans (Shukla *et al*, 2013). Cependant, même si les interactions phosphates:arrestines sont

conservées entre les différents complexes, des régions de l'arrestine présentent des conformations différentes qui modifient ses interfaces d'interaction avec ses partenaires cytosoliques, comme MEK1 ou Raf (Figure 15C). Par exemple, l'absence de phosphorylation à la position T347 (V2Rpp-1) supprime les interactions du phospho-peptide-1 avec la poche V1, ce qui modifie la conformation de la boucle back de l'arrestine, qui est impliquée dans l'interaction avec la protéine Raf (Figure 15).

Pour le complexe arrestine-2:V2Rpp-4, la perte du phosphate à la position T360 supprime son interaction avec la poche V4, comme prévu, mais interrompt aussi l'interaction entre pS357 et la poche V3. Une nouvelle interaction prend place entre pT359 de V2Rpp-4 et K11, R25 et K294 de l'arrestine-2 (Figure 16). Ainsi la fixation des phosphates à l'arrestine est interdépendante, permettant un mode de régulation de l'interaction avec l'arrestine de style rhéostat (Lee *et al*, 2010). En effet, la phosphorylation du RCPG à de multiples sites améliore sa fixation au partenaire arrestine.

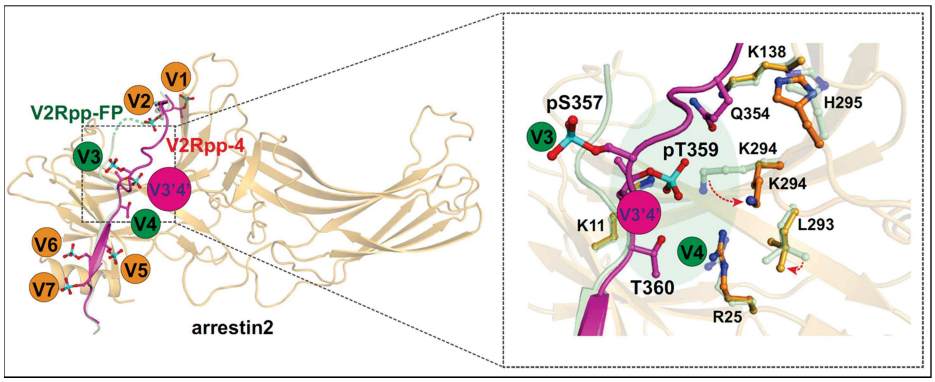


Figure 16 : Structure du phospho-peptide V2Rpp4 en complexe avec l'arrestine-2 (Figure extraite de (He *et al*, 2021). La suppression du phosphate en position T360 induit un réarrangement de la région C-terminale du récepteur vasopressine 2 allant de P353 à T360. Ce changement de conformation est accompagné de la perte de l'interaction entre le résidu phosphorylé pT359 et la poche V3, et de la transposition de pT359 qui interagit finalement avec K11, R25 et K294 (nouvelle poche V3'4').

En conclusion, Yu et son équipe (He *et al*, 2021), propose que la liaison des phosphates dans des poches spécifiques de l'arrestine, entraîne des changements conformationnels subtils de l'arrestine, qui modifie ses interfaces d'interaction avec ses autres partenaires cytosoliques, régulant ainsi son activité.

En regardant ces mécanismes d'un point de vue un peu plus dynamique, nous pourrions aussi envisager que les phosphates n'induisent pas de réorganisation de l'arrestine mais plutot que cette dernière présente plusieurs conformations en solution et qu'une seule de ces conformations reconnait un phospho-code spécifique des domaines C-terminaux des RCPGs. Ainsi, cette reconnaissance de phospho-code pourrait stabiliser l'arrestine dans une conformation particulière favorisant l'exposition d'une surface d'interaction caractéristique à un partenaire cytosolique, déclenchant ainsi une voie de la signalisation spécifique à cette conformation de l'arrestine.

2 OBJECTIF DE LA THÈSE

2.1 OBJECTIF DE LA THESE

Les Récepteurs couplés aux protéines G (RCPGs) sont des protéines transmembranaires impliquées dans la transmission de signaux extracellulaires. Leur activation donne lieu à une grande diversité de réponses cellulaires affectant de nombreuses fonctions biologiques. Ainsi, les RCPGs sont la cible de plus d'un tiers des médicaments présent sur le marché (Hauser *et al*, 2017). La signalisation des RCPGs est complexe car un seul récepteur peut activer à la fois des voies de signalisation dépendante et indépendante des protéines G. La signalisation indépendante des protéines G est médiée par l'interaction des récepteurs avec les protéines arrestines. Bien que la signalisation dépendante de l'arrestine soit une composante majeure du fonctionnement des RCPGs, nous commençons tout juste à en saisir les bases mécanistiques, même pour les RCPGs les mieux étudiés.

Le modèle actuel d'interaction RCPG:arrestine est le phospho-barcode modèle. Dans ce modèle, l'activation du récepteur par divers ligands mènerait à l'engagement de GRKs spécifiques qui pourraient phosphoryler différents sites des domaines C-terminaux des RCPGs. Ces motifs de phosphorylation, ou 'phospho-barcode', formeraient ainsi un 'code-barres' qui indiquerait ou déterminerait la conformation adoptée par l'arrestine et déterminerait ses capacités fonctionnelles (Chapitre 1, p. 50, Figure 10, (Nobles *et al*, 2011)). Il a été proposé que les motifs de phosphorylation permettant l'interaction avec l'arrestine soit composés des résidus p[S/T]Xp[S/T]XXp[S/T]/E/D (code court) ou p[S/T]XXp[S/T]XXp[S/T]/E/D (code long) (Zhou *et al*, 2017). Sachant que ces motifs sont présents en plusieurs exemplaires dans les domaines C-terminaux des RCPGs, leur phosphorylation différentielle serait associée à une fonction spécifique de l'arrestine. La liaison des phosphates ou des résidus chargés négativement des RCPGs dans des poches conservées et spécifiques de l'arrestine, permettrait de moduler la conformation de l'arrestine liée et ses fonctions (He *et al*, 2021). Cependant, les démonstrations expérimentales de ces modes de fonctionnement sont encore rares.

La région fonctionnelle des RCPGs qui régit l'interaction avec l'arrestine se situe dans le domaine C-terminal de ces premiers. Malheureusement cette région est souvent tronquée ou absente des structures atomiques (RX ou EM) à cause de sa grande flexibilité. En effet, ces régions sont prédites pour être intrinsèquement

désordonnées (IDRs) (Jaakola *et al*, 2005). Les IDRs sont des régions particulières qui ne contiennent pas ou peu de structure secondaire et tertiaire (Chapitre 1, Revue, p. 19). Cette caractéristique structurale leur confère une adaptabilité conformationnelle qui rend les IDRs très adaptés aux fonctions cellulaires versatiles, comme la signalisation (Wright & Dyson, 2015).

Dans ce projet, nous avons pour objectif de déchiffrer les mécanismes moléculaires impliqués dans l'interaction RCPG:arrestine, et donc à la signalisation dépendente de l'arrestine. Pour cela nous avons utilisé des méthodes biophysiques complémentaires comme la résonance magnétique nucléaire (RMN), la diffusion des rayons X aux petits angles (SAXS) et la spectroscopie de fluorescence (FRET). Ces méthodes ont été appliquées à des systèmes modèles de complexité croissante composés de peptides C-terminaux isolés de RCPGs, d'arrestines ainsi que de kinases purifiées et de récepteurs entiers assemblés dans des systèmes imitant la membrane (nanodisques). Pour être représentatif, nous avons travaillé sur trois RCPGs différents de classe A : le récepteurs vasopressine V2 (V2R), le récepteur sécrétagogue de l'hormone de croissance de type 1a, ou récepteur à la ghréline, (GHSR) et le récepteur β2-adrénergique (β2AR). Outre le fait que ces récepteurs appartiennent à deux classes d'interactant avec l'arrestine (Oakley *et al*, 2000), ils sont également des cibles pharmaceutiques importantes en raison de leur rôle majeur dans des processus physiologiques et physiopathologiques.

L'objectif principal de ma thèse a été d'étudier cette interaction du point de vu des récepteurs. Nous avons analysé l'effet de la phosphorylation par les GRKs sur la conformation et l'interaction du domaine C-terminal des RCPGs avec l'arrestine. Pour étudier l'effet de la phosphorylation, nous avons utilisé la kinase GRK2, produite au laboratoire, ainsi que des variants phosphomimétiques des domaines C-terminaux. En parallèle, nous avons aussi analysé le réarrangement global que subit l'arrestine en complexe avec les RCPGs.

Mes travaux de thèse fournissent une nouvelle image de la signalisation des RCPGs dépendante de l'arrestine. Nous avons mis en évidence quelques clés structurales et dynamiques requises pour l'interaction des RCPGs avec les arrestines, qui nous aideront par la suite à comprendre comment se déroule la signalisation indépendante de la protéine G.

2.2 SYSTEME MODELE DE RCPGS

2.2.1 Récepteurs de la classe A des RCPGs

De manière à obtenir une vision globale du rôle de la phosphorylation sur la conformation et l'interaction des RCPGs avec l'arrestine, nous avons travaillé sur trois récepteurs, non-visuels, de classe A (ou classe Rhodospine) : β2AR, GHSR et V2R. La classe A regroupe plus de 80 % des RCPGs humains et ses récepteurs sont très hétérogènes en séquence. Par exemple, β2AR, GHSR et V2R partagent ~ 23 % d'identité de séquence (Table 4). Cette hétérogénéité provient en partie de leur région extra-membranaire, qui représente, en moyenne, 40 % de leur résidus (Venkatakrishnan *et al*, 2014). En effet, ces régions transmembranaires sont très variables en longeur et en séquence.

	β2AR	GHSR	V2R
β2AR	X	23	21
GHSR		X	25
V2R			X

Table 4 : Les RCPGs modèles utilisés dans ce manuscrit partagent ~ 23 % d'identité de séquence. Les valeurs représentent le pourcentage d'identité entre protéines. Les alignements ont été effectués avec BlastP (blast.ncbi.nlm.nih.gov) à partir des codes UniProt des RCPGs humains : β2AR : P07550 ; GHSR : Q92847 ; V2R : P30518.

Ces récepteurs sont des cibles pharmaceutiques importantes en raison de leur rôle majeur dans certains processus physiologiques et physiopathologiques. Brièvement, le récepteur β2AR est impliqué dans la relaxation des muscles lisses et la broncho-dilatation. Les molécules activant ce récepteur, comme l'isoproterenol (agoniste) sont souvent utilisées pour traiter les troubles respiratoires, comme l'asthme (Altosaar *et al*, 2019). Concernant le récepteur GHSR, son activation par la ghréline, peptide de 28 acides aminés, entraine la libération de l'hormone de croissance qui stimule l'appétit. Ainsi, son effet orexigène (substance capable d'augmenter l'appétit) est souvent utilisé dans le traitement des désordres alimentaires comme le manque sévère d'appétit (Davenport *et al*, 2019). Le récepteur V2R lie la vasopressine (ou arginine vasopressine), peptide de 9 acides aminés, qui permet la réabsorption de l'eau au niveau des reins. Des molécules bloquant l'activité de ce récepteur, comme

le minirin (antagoniste) sont souvent utilisées dans le traitement de l'hyponatrémie (diminution de la concentration plasmatique de sodium causée par un excès d'eau par rapport au sodium) (Manning *et al*, 2008).

En plus d'être des cibles thérapeutiques, chacun des trois récepteurs présente un intérêt scientifique différent.

Le premier intérêt porte sur la connaissance ou non des motifs de phosphorylations spécifiques des régions C-terminales de ces récepteurs (cette partie sera détaillée au Chapitre 4, p. 215 pour les trois récepteurs). Les sites phosphorylés par GRK2 et GRK6 sont connus pour le récepteur β2AR (Figure 17) (Nobles *et al*, 2011), mais pas spécifiquement pour les deux autres récepteurs. Ainsi, β2AR qui est le récepteur le plus étudié jusqu'à présent, tant d'un point de vue structural que du code de phosphorylation, sera notre récepteur de référence. Les récepteurs GHSR et V2R présentent des sites phosphorylés uniquement après activation par leur agoniste respectif (vasopressine ou ghréline) (Figure 17) (Bouzo-Lorenzo *et al*, 2016; Innamorati *et al*, 1997). La phosphorylation dépendante de l'activation des RCPGs suggère fortement une activité des GRKs dont l'activation se fait par liaison aux RCPGs actifs (Chapitre 1, p. 48, Figure 7).

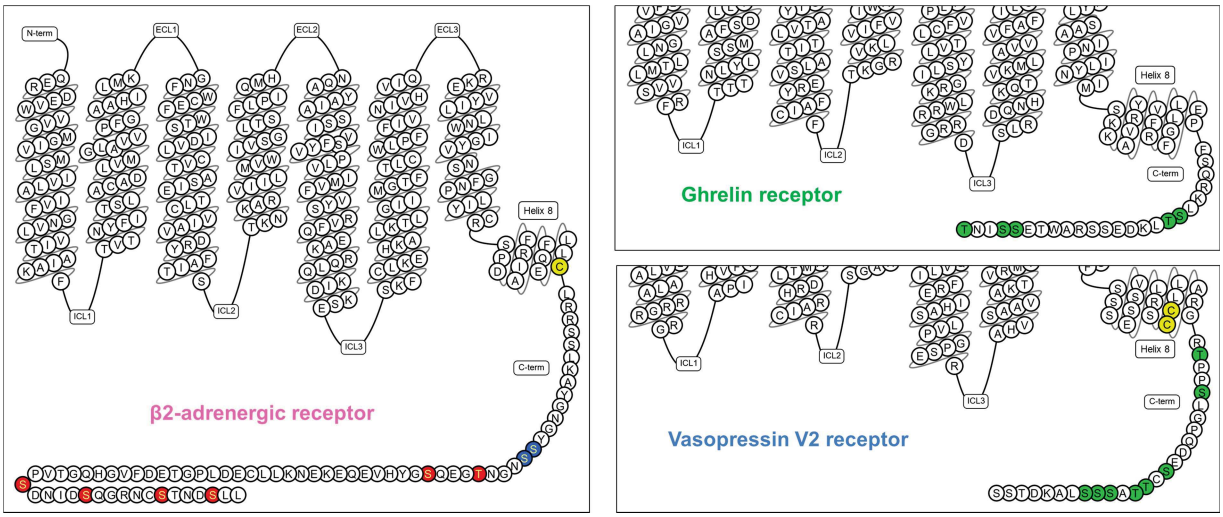


Figure 17 : Sites de phosphorylation des GRKs dans les domaines C-terminaux des récepteurs β2AR, V2R et GHSR (Figure extraite de GPCRdb (https://gpcrdb.org/)). Les sites colorés correspondent aux résidus phosphorylés par GRK6 (bleu), GRK2 (rouge) ou par des GRKs non-spécifiées (vert), selon (Nobles *et al*, 2011) pour β2AR ; (Bouzo-Lorenzo et al, 2016) pour GHSR et (Innamorati et al, 1998, 2001) pour V2R. Les cystéines palmitoylées sont colorées en jaune selon pour (O'Dowd *et al*, 1989) β2AR et selon (Sadeghi *et al*, 1997) pour V2R. Les régions du récepteur sont légendées comme suit : N-term, domaine N-terminal ; C-term, domaine C-terminal ; ECL, boucle extra-cellulaire ; ICL, boucle intra-cellulaire.

Pour GHSR, les sites phosphorylés ont été localisés par spectrométrie de masse. Pour V2R, les sites phosphorylés ont été localisés par mesure du taux d'intégration de phosphates radioactifs dans le domaine C-terminal, progressivement tronqué. Pour ces deux récepteurs, l'identité des GRKs impliquées dans la phosphorylation spécifique des sites n'est pas connue.

Le deuxième intérêt est que ces récepteurs appartiennent à des classes d'interactants différents vis-à-vis de l'arrestine, classe A ou B (décrites au Chapitre 1, p. 44). Les récepteurs β2AR et GHSR sont de la classe A, alors que le récepteur V2R est de la classe B. Ces récepteurs permettront d'observer s'il existe des mécanismes moléculaires différents entre les deux classes d'interactant de l'arrestine.

Le troisième intérêt est que ces trois RCPGs présentent un nombre de sites de palmitoylation différents (Figure 17). Le récepteur à la ghréline ne présentant aucune cystéine de palmitoylation, il sera une référence pour identifier s'il y a une différence dans le mécanisme d'interaction avec l'arrestine, lors de l'étude des récepteurs entiers reconstitués dans les nanodisques.

2.2.2 Des régions C-terminales aux récepteurs entiers

Nous souhaitons obtenir des informations structurales au niveau atomique sur les régions C-terminales des RCPGs. Pour se faire, dans un premier temps, nous avons travaillé uniquement sur des peptides cytosoliques de ces domaines (Table 5).

Il a été montré dans la littérature que des cystéines conservées dans les extrémités C-terminales des RCPGs présentent des palmitoylations qui permettent l'ancrage dans la membrane plasmique (Qanbar & Bouvier, 2003). C'est le cas de la cystéine 341 du récepteur β2-adrénergique (O'Dowd *et al*, 1989) et des cystéines 341 et 342 du récepteur vasopressine 2 (Sadeghi *et al*, 1997) (Figure 17). Dans l'objectif de travailler sur des extrémités C-terminales cytosoliques et solubles, les constructions ADN des C-termini ont été dessinées pour débuter après les cystéines palmitoylées. Pour GHSR, la séquence débute après le dernier acide aminé de l'hélice 8 (Figure 17). D'autre part, nous étudions le domaine C-terminal de GHSR dans son récepteur entier reconstitué en nanodisque et le récepteur entier en lui-même (travail en collaboration avec Jean-Louis Banères et non présenté dans ce manuscrit).

Récepteur	Nombre de résidus	Numérotation	pI théorique
β2AR (P07550)	72	342-413	4.71
GAMALRRSSLKAYGNGY**S**$_{355}$**S**$_{356}$NGN**T**$_{360}$GEQ**S**$_{364}$GYHVEQEKENKLLCEDLPGTEDFVGHQGTVP**S**$_{396}$DNID**S**$_{401}$ QGRNC**S**$_{407}$TND**S**$_{411}$LL Sites phosphorylé selon (Nobles *et al*, 2011)			
V2R (P30518)	29	343-371	6.16
GPMAARGR**T**$_{347}$PP**S**$_{350}$LGPQDE**S**$_{357}$C**T**$_{359}$**T**$_{360}$A**S**$_{362}$**S**$_{363}$**S**$_{364}$LAKDTSS Sites phosphorylés selon (Innamorati *et al*, 1998, 2001)			
GHSR (Q92847)	28	339-366	6.26
GPMAGFEPFSQRKL**S**$_{349}$**T**$_{350}$LKDESSRAWTE**S**$_{362}$**S**$_{363}$IN**T**$_{366}$ Sites phosphorylés selon (Bouzo-Lorenzo *et al*, 2016)			

Table 5 : Séquence des peptides C-terminaux des récepteurs β2AR, V2R et GHSR. Les codes UniProt (https://www.uniprot.org) des protéines sont indiqués entre parenthèse à côté des noms des récepteurs. Le tableau affiche aussi le nombre de résidu, la numérotation et le point isoélectrique (pI) des peptides des RCPGs utilisés dans cette étude. Les sites colorés correspondent aux résidus phosphorylés par GRK6 (bleu), GRK2 (rouge) ou par des GRKs non-spécifiées (vert).

2.3 STRATEGIE D'ETUDE

2.3.1 Étude des régions C-terminales des RCPGs

Dans l'objectif de décrire comment la phosphorylation par les GRKs affecte la conformation et l'interaction des RCPGs avec l'arrestine, nous avons mis en place, au laboratoire, la production et la purification des kinases GRK2 et GRK6 en cellules d'insectes, ainsi que la phosphorylation des peptides par GRK2. Le temps de mettre en place la phosphorylation par les GRKs, nous avons travaillé sur des peptides phosphomimétiques (notés 'variants' ou 'pm') des domaines C-terminaux.

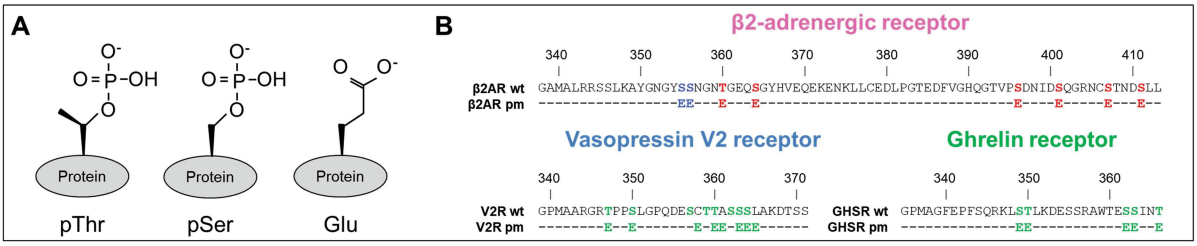


Figure 18 : Utilisation d'un acide glutamique pour mimer la phosphorylation. A) Comparaison des charges (à pH 6.7) et de la longueur entre une phospho-thréonine (pThr), une phospho-sérine (pSer) et un acide glutamique (Glu). **B)** Séquence des peptides C-terminaux non-phosphorylés (wt) et phosphomimétiques (pm) des récepteurs β2AR, V2R et GHSR. Les sites phosphorylés par GRK6 (bleu), GRK2 (rouge) et par des GRKs non-spécifiées (vert) ont été remplacés par des acides glutamiques (E), selon la Figure 17.

Dans la séquence de ces variants, les acides aminés phosphorylés par les GRKs ont été remplacés par des acides glutamiques de manière à mimer la charge et la longueur d'une phospho-sérine ou d'une phospho-thréonine (Figure 18).

Techniquement, la caractérisation biophysique et structurale des domaines C-terminaux des RCPGs dans leur état non-phosphorylé, phosphorylé et phosphomimétique a été faite en utilisant principalement la Résonance Magnétique Nucléaire (RMN) couplée à d'autres analyses biophysiques comme : le dichroïsme circulaire (CD), la diffusion de la lumière aux petites angles (MALS) et la diffusion des rayons X aux petits angles (SAXS). La fonctionnalité des peptides a été déterminée selon leur habilité à se lier à l'arrestine par RMN, ainsi que par anisotropie de fluorescence et par thermophorèse de fluorescence (MST).

2.3.2 Étude du réarrangement global de l'arrestine

Lors de l'interaction avec un RCPG, l'arrestine passe d'une conformation inactive à active (Chapitre 1, p. 40). Dans l'objectif d'étudier ce réarrangement structural, nous avons analysé l'arrestine dans différentes conformations par SAXS : inactive (arrestine seule), pré-active (variant R169E) ou active (arrestine en complexe avec le domaine C-terminal des RCPGs). Le variant R169E mime la conformation active de l'arrestine grâce à l'inversion de la charge du résidu R169 (numérotation de l'arrestine-2 humaine) qui compose le cœur polaire. Cette inversion de charge conduit à la destruction du cœur polaire et permet l'obtention d'une conformation pré-active de l'arrestine, comme décrit pour le variant R175 de l'arrestine-1 (équivalent du varient R169 de l'arrestine-2) (Granzin *et al*, 2015) (Figure 19). Par ailleurs, ce variant est capable d'interagir avec les RCPGs non-phosphorylés, il est donc souvent appelé arrestine 'phospho-indépendante' (Kovoor *et al*, 1999).

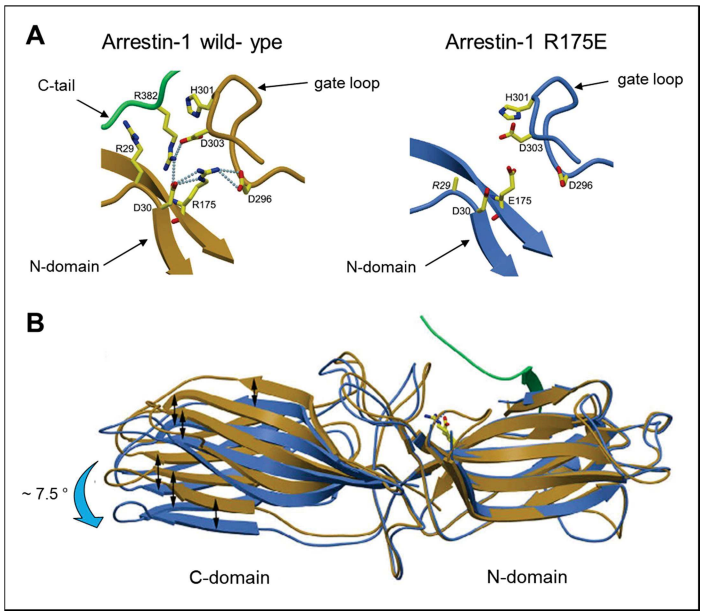


Figure 19 : La mutation R175E entraîne une destruction du cœur polaire de l'arrestine-1 permettant l'obtention d'une forme pré-active de l'arrestine (Figure modifiée de (Granzin *et al*, 2015)). **A)** Les liaisons hydrogènes (bleu) entre la boucle gate, le domaine N-terminal et la queue C-terminale qui stabilise l'arrestine sauvage (wt, à gauche, PDB : 3UGX) en conformation inactive sont complètement détruites chez le variant R175E (à droite, PDB : 4ZRG). **B)** Cette destruction du cœur polaire permet à l'arrestine de passer dans une conformation pré-active (bleu) avec une rotation de 7.5° du domaine C-terminal par rapport au domaine N-terminal. Dans sa conformation active, une rotation de ~20° est visualisé entre ces deux domaines (PDB : 5W0P). La structure basale de l'arrestine sauvage est en doré et le résidu R175 est représenté en bâton.

3 MATÉRIEL ET MÉTHODES

3.1 PRODUCTION DES PROTEINES

Nous avons effectué un travail important de biochimie afin de produire les peptides des RCPGs (RCPG-Cters) et l'arrestine-2 sauvage (wt) et son variant R169E, à partir de protocoles affinés au cours des dernières années. Nous avons aussi mis en place la production des kinases GRK2 et GRK6 en système d'expression baculovirus.

3.1.1 Production des domaines C-terminaux des RCPGs et de l'arrestine-2

Les étapes d'expression et de purification des RCPG-Cters sauvages (wt) et phosphomimétiques (pm), et des arrestines sauvages (wt) et R169E, sont identiques. Elles sont présentées ci-dessous.

3.1.1.1 Clonage en système d'expression bactérien Escherichia coli

Les ADNs codants pour les séquences protéiques des domaines C-terminaux des RCPGs (Chapitre 2, p. 70, Figure 18) et de l'arrestine-2 (arrestine-2 rat, code UniProt P29066) ont été commandés chez GeneArt® gene synthesis (life technologies). Ils ont été clonés dans des plasmides bactériens *Escherichia Coli* (*E. coli*) pETM11 pour l'arrestine et pETM33 pour les peptides des RCPGs, à l'exception du β2AR-Cter qui a été cloné en vecteur pET1a (tous ces plasmides étaient disponibles au laboratoire). Dans les vecteurs, l'expression génique est sous contrôle d'un promoteur T7 qui est induit en présence de lactose ou de son analogue l'IPTG (IsoPropyl β-D-1-ThioGalactopyranoside). Dans les plasmides pETM33 et pET1a, les protéines sont fusionnées en N-terminal à un tag Histidine-GST (six histidines suivies de la protéine Glutathione S Transferase) qui peut être clivé par utilisation des protéases 3C (protéase 3C PreScission dérivée du rhinovirus humain) ou TEV (protéase issue du virus responsable de la maladie du tabac), respectivement (Figure 20). Dans le plasmide pETM11, l'arrestine-2 est liée à un tag 6-histidines clivable par utilisation de la protéase TEV (Figure 20).

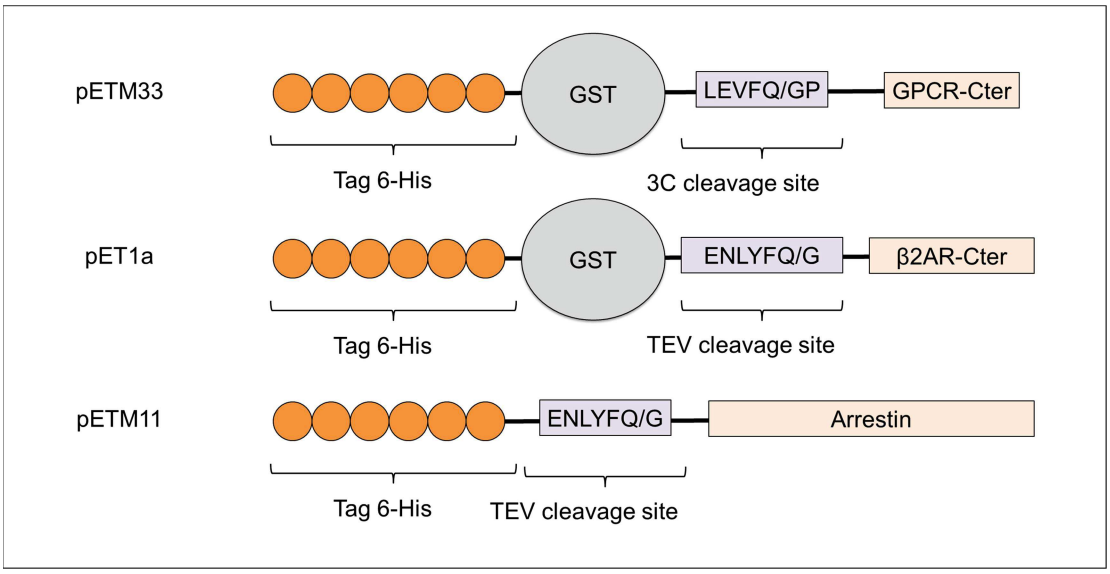


Figure 20 : Représentation schématique des protéines recombinantes utilisées dans ce manuscrit. Selon le vecteur de production utilisé (pETM33, pET1a ou pETM11), les constructions varient. Les RCPG-Cters ont un tag histidine-GST en N-terminal clivable par les protéases 3C PreScission (3C) ou TEV. Les arrestines ont uniquement un tag histidine en N-terminal clivable par la protéase TEV. Le tag histidine est composé de 6 histidines qui se suivent (rond orange). Les séquences de clivage de chaque protéase sont indiquées dans les encadrés violet clair.

Les bactéries *E. coli* (souche BL21 DE3, Fisher scientific) ont été transformées par choc thermique à 42 °C par un des plasmides précédents. Ensuite, 1 mL de milieu riche de culture, Luria Broth (LB), est ajouté et les cellules sont incubées 1 heure (h) à 37 °C. La solution bactérienne est ensuite étalée sur gélose LB-agar contenant 50 µg/mL de kanamycine, permettant de sélectionner uniquement les bactéries transformées par les plasmides pET qui contiennent un gène de résistance à l'antibiotique. Puis les cellules sont incubées sur une nuit (Over-Night, ON) à 37 °C. Le clonage a été vérifié par séquençage (Genewiz).

3.1.1.2 Cultures bactériennes

Un clone bactérien issu de la transformation est utilisé pour inoculer des pré-cultures de cellules bactériennes *E. coli* en milieu LB contenant 50 µg/mL de kanamycine et les cellules sont incubées ON à 37 °C. Ces pré-cultures servent à ensemencer 500 mL de culture contenant toujours 50 µg/mL de kanamycine. Pour l'arrestine-2, les cellules sont cultivées en milieu riche LB, à 37 °C avant induction, puis à 20 °C après induction durant une nuit. Pour les peptides des RCPGs, les cellules sont mises en cultures dans trois milieux différents selon le marquage souhaité : en milieu riche LB

pour les protéines non-marquées isotopiquement, en milieu auto-inductible N5052 enrichi en ^{15}N-NH_4Cl pour les protéines marquées ^{15}N ou en milieu minimum M9 enrichi en ^{15}N-NH_4Cl et en ^{13}C-glucose pour les protéines marquées ^{15}N et ^{13}C. Les cultures LB durent 16 h à 37 °C alors que les cultures N5052 durent 24 h à 25 °C. Pour le milieu M9, la culture est exceptionnellement ensemencée avec une pré-culture M9 non-marquée contenant 50 µg/mL de kanamycine et précédemment incubée ON à 37 °C. La culture M9 est ensuite incubée à 37 °C avant induction puis à 20 °C après induction pour une durée de 3 h. L'induction des milieux LB et M9 est effectuée lorsque la densité optique (DO) des cultures atteint 0.6 par ajout de 0.25 mM d'IPTG (IsoPropyl β-D-1-ThioGalactopyranoside) pour l'arrestine-2 ou par ajout de 2 mM d'IPTG pour les RCPG-Cters. Après culture, les cellules sont centrifugées à 6 000 g pendant 20 minutes (min) et le culot peut être conservé à -80 °C.

3.1.1.3 Purification

Les purifications des domaines C-terminaux des RCPGs et de l'arrestine-2 sont composées de trois et quatre étapes, respectivement : la lyse bactérienne, une chromatographie d'affinité histidine, une chromatographie échangeuse d'anions (pour l'arrestine uniquement) et une chromatographie d'exclusion. Les tampons utilisés à chaque étape de purification sont présentés Table 6.

Etape de purification	protéine	Tampon
Lyse cellulaire (Tampon de lyse)	RCPG-Cter	20 mM Tris 20 pH 7.5, 300 mM NaCl, 2 mM Dithiothréitole (DTT) et une tablette d'inhibiteur de protéases (Roche)
	arrestine	50 mM Tris pH 8.0, 200 mM NaCl, 2 mM DTT, 1 mg/ml lysosyme et des anti-protéases (Roche et 0.5 mM PMSF)
Affinité (tampon A)	RCPG-Cter	20 mM Tris 20 pH 7.5, 300 mM NaCl et 2 mM DTT
	arrestine	50 mM Tris pH 8.0, 200 mM NaCl et 2 mM DTT
Echangeuse d'anion (tampon B)	RCPG-Cter	/
	arrestine	20 mM Tris pH 8.0, 50 mM NaCl et 2 mM DTT
Exclusion (tampon C)	RCPG-Cter	50 mM Bis-Tris pH 6.7, 50 mM NaCl, et 2 mM DTT
	arrestine	50 mM Bis-Tris pH 6.7, 50 mM NaCl and 2 mM DTT

Table 6 : Tampons de purification des domaines C-terminaux des RCPGs et de l'arrestine-2.

3.1.1.3.1 Lyse cellulaire

Les culots bactériens sont repris dans 30 mL de tampon de lyse. Les cellules de culture des RCPG-Cters sont lysées par sonication pendant 15 min à 20 watts (cycles de 1 min entrecoupés de 1 min de pause). Pour les cultures d'arrestine-2, les cellules sont lysées par pression ('French press') à ~ 700 bar. Les débris cellulaires sont enlevés par centrifugation à 47 000 g pendant 30 min à 10 °C. Le surnagent est récupéré et filtré sur une membrane à 0.22 µM.

Les étapes suivantes de purification sont effectuées à température ambiante pour les RCPG-Cters et en chambre froide à 4 °C pour l'arrestine-2. A chaque étape, des échantillons sont conservés et analysés par SDS-PAGE (électrophorèse sur gel de polyacrylamide contenant du laurylsulfate de sodium) et la concentration des protéines est suivie par mesure de l'absorbance (Abs) à 280 nm.

3.1.1.3.2 Chromatographie d'affinité histidine

Après la lyse cellulaire, la solution protéique filtrée à 0.45 µm est injectée sur une colonne d'affinité histidine (5 mL HisTrap HP column, GE Healthcare) équilibrée en tampon A. Après lavage de la colonne, les protéines sont éluées par un gradient (linéaire ou en escalier) de tampon A contenant 500 mM d'imidazole. La solution d'élution est ensuite dialysée ON à 4 °C en tampon A contenant la protéase 3C (pour les protéines V2R-Cter et GHSR-Cter) ou TEV (pour l'arrestine-2 et β2AR-Cter) à ratios (g/g) 3C/protéine de 1/100 ou TEV/protéine de 1/50. Le lendemain, les protéines sont réinjectées sur la colonne d'affinité pour enlever le tag clivé.

3.1.1.3.3 Chromatographie échangeuse d'anion

Cette étape n'était pas présente dans le protocole initial de purification de l'arrestine. Elle a été ajoutée de manière à augmenter la pureté de l'échantillon, ce qui a permis par la suite une bien meilleure conservation de l'arrestine à -80 °C (Figure 21).

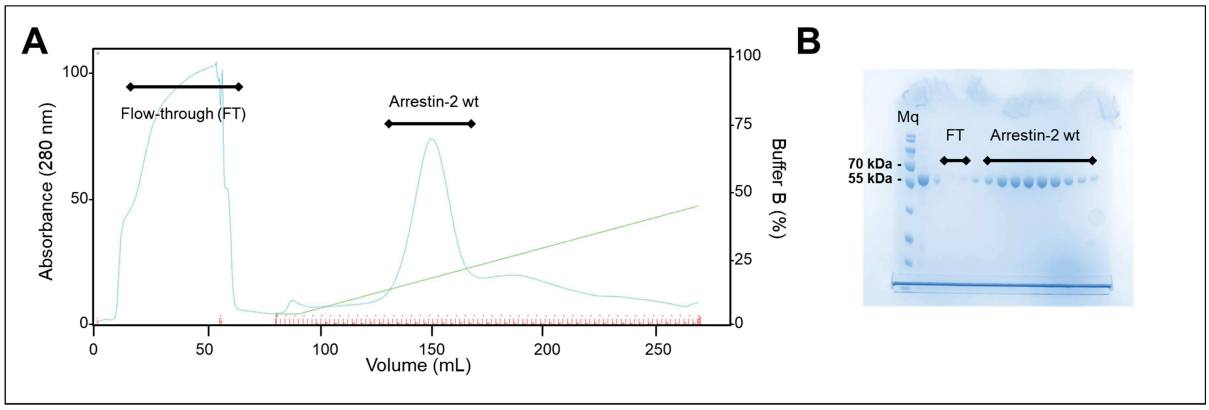


Figure 21 : Purification et pureté de l'arrestine-2. A) Chromatogramme de l'arrestine-2 sauvage (wt) issu de la chromatographie échangeuse d'anion. Un pic important est observé en début de chromatogramme, qui correspond à ce qui n'a pas été retenu sur la colonne (non-retenu, FT). Curieusement, ce pic ne donne pas de signal sur gel SDS-PAGE 12 % (**B**). Il peut potentiellement correspondre à des protéines trop faiblement concentrées pour être visualisées sur le gel. Les fractions correspondantes aux FT et à l'arrestine-2 ont été déposées sur gel où elles sont représentées par les barres noires.

Suite à la deuxième chromatographie d'affinité, l'arrestine est diluée jusqu'à 50 mM en NaCl pour être injectée sur une colonne échangeuse d'anion (HiTrap Q HP, GE Healthcare) équilibrée en tampon B. Après lavage, l'arrestine est éluée par un gradient en sel de 5 à 50 % en tampon B contenant 1 M de NaCl.

3.1.1.3.4 Chromatographie d'exclusion de taille (SEC)

Les fractions issues des chromatographies précédentes et contenant les protéines d'intérêts sont mélangées et concentrées à 2 mL pour être injectées en chromatographie d'exclusion. Les peptides de ~ 3 et ~ 7 kDa des RCPGs sont injectées sur une colonne HiLoad 16/60 Superdex S75 (GE Healthcare) car elle permet la séparation des protéines allant de 3 à 70 kDa. À l'inverse, l'arrestine-2 de 47 kDa est injectée sur une colonne HiLoad 16/60 Superdex S200 (GE Healthcare) qui sépare les protéines allant de 10 à 600 kDa.

Il est intéressant de noter que les domaines C-terminaux des RCPGs éluent plus rapidement que ce qui est attendu pour une protéine globulaire de même masse molaire (Figure 22). Ce phénomène est représentatif du comportement des protéines ou des régions intrinsèquement désordonnées (IDPs/IDRs). En effet, les IDPs/IDRs contiennent peu ou pas de structures secondaires et tertiaires, ils présentent donc un rayon hydrodynamique plus élevé que les protéines globulaires de même masse et

élueront plus vite lors d'une chromatographie d'exclusion pour des masses molaires similaires (Uversky, 2012).

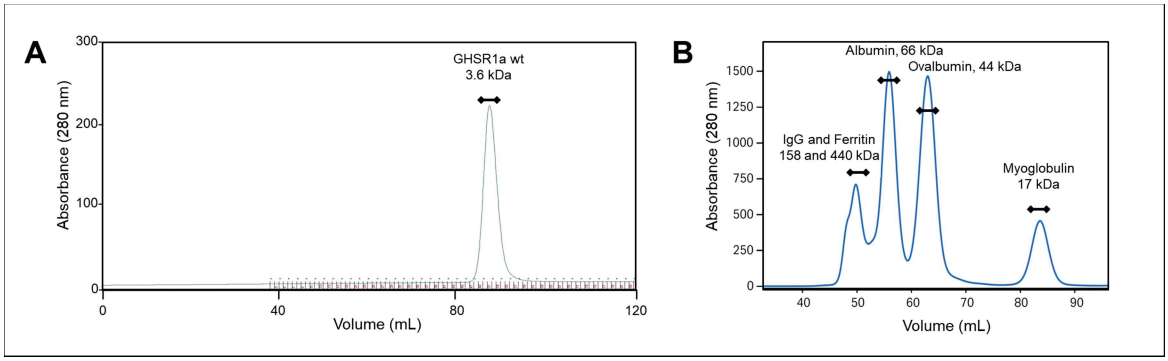


Figure 22 : En SEC, le domaine C-terminal de GHSR élue plus rapidement que ce qui est attendu pour une protéine de 3.6 kDa. A) Profile de la chromatographie d'exclusion (GF) sur colonne HiLoad 16/60 Superdex S75 de la protéine GHSR-Cter. B) Profil de calibration de la colonne HiLoad 16/60 Superdex S75 (Figure extraite du manuel constructeur disponible en ligne à l'adresse https://cdn.cytivalifesciences.com). Les protéines de calibration et leurs tailles sont indiquées sur l'image. Selon la courbe de calibration, une protéine de 3.6 kDa doit éluer vers ~ 96 mL. Cependant, GHSR-Cter élue à ~ 87 mL, soit vers la myoglobuline (17 kDa). GHSR-Cter se comporte donc comme une région intrinsèquement désordonnée en chromatographie d'exclusion. Des résultats similaires sont obtenus pour β2AR-Cters et V2R-Cters.

Pour en revenir à la purification, les fractions protéiques pures sont ensuite concentrées à 400 µM pour les domaines C-terminaux des RCPGs et à 250 µM pour l'arrestine-2. Les concentrations protéiques sont déterminées par mesure de l'absorbance à 280 nm (Abs), excepté pour V2R-Cter. En effet, V2R-Cter ne contient pas d'acides aminés aromatiques (ε = 0) et sa concentration est estimée par réfractométrie (incrément d'indice de réfraction : dn/dc = 0.18 $g.mL^{-1}$). Les rendements de culture des protéines sont présentés Table 7 :

Protéine	wt-V2R	pm-V2R	wt-GHSR	pm-GHSR	wt-β2AR	pm-β2AR	arrestine
E ($M^{-1}.cm^{-1}$)	0	0	5 500	5 500	4 470	4 470	19 370
mg/L de culture	2.3	0.9	8.6	4	10	6.7	0.3

Table 7 : Rendement de production des peptides des RCPGs marqués ^{15}N et de l'arrestine-2 en fin de purification. Les peptides des RCPGs sont notés 'wt' pour la forme non-phosphorylée et 'pm' pour la forme phospho-mimétique. Les coefficients d'extinction molaire (ε, $M^{-1}.cm^{-1}$) ayant servi au calcul des absorbances ont été obtenus à partir de ProtParam (https://web.expasy.org/protparam/).

Les protéines pures sont conservées au congélateur à -80 °C. Pour les RCPG-Cters, dont le désordre les rend très accessibles au solvant, les protéines sont congelées après addition de 1 mM de TCEP (Tris(2-CarboxyEthyl)Phosphine) et 0.5 mM

d'EDTA (EthyleneDiamineTetraAcetic acid). Ces molécules servent respectivement à limiter la formation de ponts disulfures inter-moléculaires et à limiter leur dégradation par les potentielles protéases encore présentes après purification.

3.1.1.3.5 Cas de β2AR-Cter

Initialement, le domaine C-terminal de β2AR était produit à partir du plasmide pETM33 comme pour les autres peptides. Cependant, la chromatographie d'exclusion montrait deux pics au lieu d'un et l'analyse RMN de ces pics a révélé la présence d'une espèce clivée après Q391 de β2AR-Cter (Figure 23). La séquence de β2AR-Cter autour de Q391 étant similaire à la séquence reconnue par la protéase 3C (Figure 23), nous avons changé de plasmide pour passer en pET1a. Dans ce plasmide, la séquence de clivage du tag est celle reconnue par la protéase TEV (Figure 20). La protéine produite à partir de ce plasmide ne donne qu'un seul pic sur le chromatogramme, et n'est pas clivée.

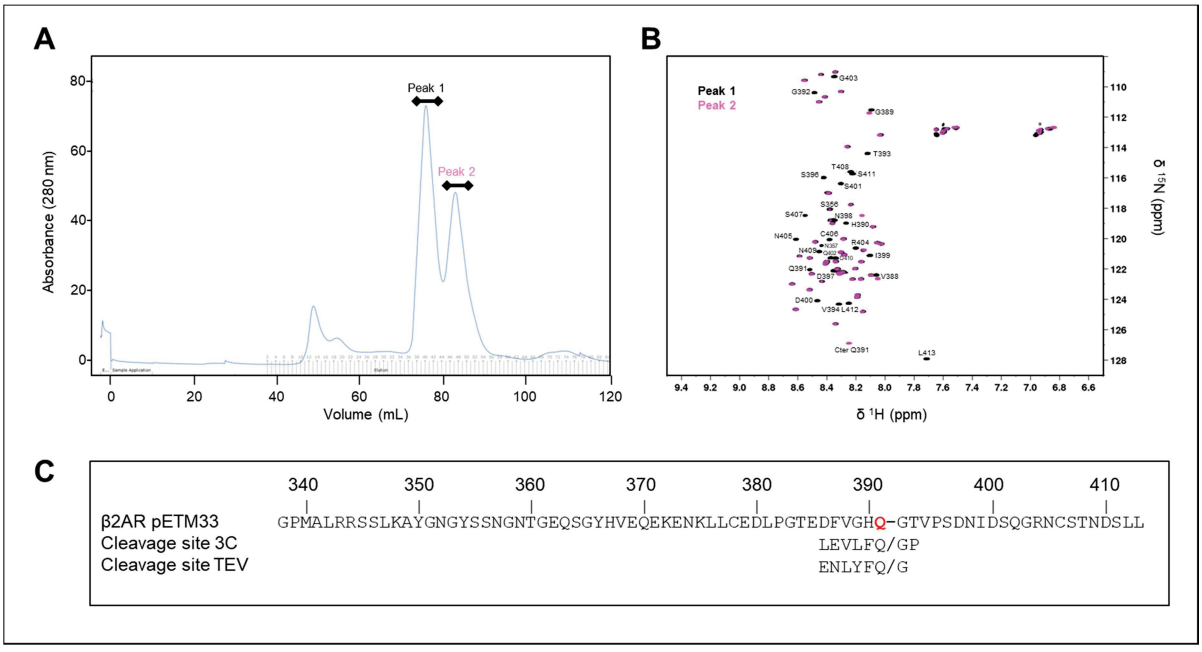


Figure 23 : Clivage du domaine C-terminal de β2AR par la 3C protéase. Lorsque le peptide de β2AR était produit à partir du plasmide pETM33, deux pics étaient visualisés sur le chromatogramme (**A**). L'analyse RMN de ces deux pics a mis en évidence la présence d'une espèce clivée de β2AR-Cter (pic deux, rose) avec un clivage après Q391 car les résidus suivants ne sont plus visibles sur le spectre (spectre rose) comparé au spectre du pic 1 (noir) (**B**). Après comparaison de la séquence de β2AR avec le site de clivage de la protéase 3C (**C**), nous avons émis l'hypothèse que la 3C clivait le peptide. Nous avons donc changé de plasmide pour passer en pET1a, ce qui a résolu le problème.

3.1.2 Production des kinases GRK2 et GRK6

3.1.2.1 Présentation du système de vecteur d'expression de Baculovirus

Durant ma thèse, nous avons mis en place au laboratoire l'expression des kinases GRK2 et GRK6 en système de vecteur d'expression de Baculovirus (BEVS), avec l'aide de Thomas Robert. Ce système repose sur l'intégration d'un gène d'intérêt dans le génome du virus de la polyédrose *Autographa californica* (AcNPV). Ce virus infecte les cellules d'insectes, dont les cellules ovariennes de papillons *Spodoptera frugiperda* (*Sf9*). Le gène d'intérêt est cloné au sein du gène codant pour la polyhèdrine. Cette protéine virale est produite en grande quantité de 24 à 72 h suite à l'infection virale (Ciccarone *et al*, 1998) mais n'est pas essentielle pour la réplication virale (Smith *et al*, 1983).

Nous avons utilisé le système d'expression Bac-to-Bac pour cloner nos gènes d'intérêts dans le génome du baculovirus (Luckow *et al*, 1993). Dans ce système, le gène d'intérêt est introduit par transposition spécifique TN-7 dans un vecteur transportant l'ADN du baculovirus, nommé bacmide (Figure 24). Le bacmide contient une cassette qui lui permet de se répliquer chez *E. coli* tout en gardant sa capacité à infecter des cellules d'insectes. Cette cassette est insérée à l'emplacement du gène (locus) de la polyhédrine et est constituée de plusieurs éléments : une origine de réplication bactérienne (mini-F), un gène de résistance à la kanamycine (Kan) et une séquence codante pour le peptide α de la β-galactosidase, LacZα (Figure 24). La séquence d'attachement du transposon bactérien Tn7 (mini-attTn7) est présente dans le gène LacZα. Cette insertion n'interrompt pas le cadre de lecture du peptide LacZα. Ainsi, après transformation, le bacmide confère aux bactéries *E. coli* DH10B$_{AC}^{TM}$ une résistance à la kanamycine et il permet la complémentation de la β-galactosidase par ajout du peptide α, rendant la β-galactosidase active. Le clivage d'un substrat chromogène, comme le Bluo-Gal, par la β-galactosidase permet l'obtention de colonies bleues. Ce système permettra par la suite une sélection blanc/bleu des bactéries contenant le bacmide recombinant.

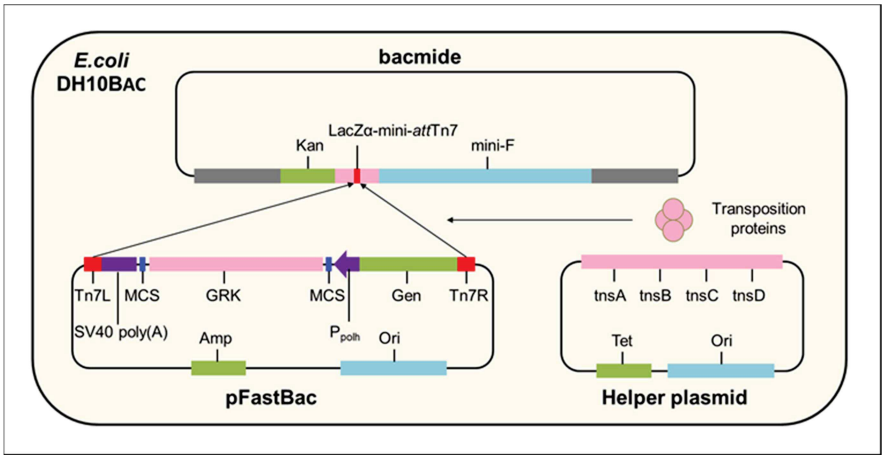


Figure 24 : **Production du bacmide recombinant exprimant les protéines GRKs chez *E. coli* DH10Bac** (Figure modifiée à partir de (Ciccarone *et al*, 1998))**.**Après clonage du gène GRK dans le plasmide pFastBac, il est transformé dans les cellules *E. coli* DH10Bac qui contiennent le bacmide et le plasmide helper. La cassette délimitée par les sites de transposition (Tn7R/L en rouge) du pFastBac est transposée par les protéines de transposition (tnsA à D en rose) au site d'attachement mini-attTn7 (en rouge) du bacmide. Les éléments principaux des plasmides sont indiqués comme suit : gènes de résistance, vert ; origines de réplication, bleu clair ; gènes codant pour les protéines GRK, tns et pour le peptide LacZα, rose. Dans le pFastBac, le promoteur (P_{polh}, violet), le terminateur (SV40 poly (A), violet) et les sites de clonages multiples (MCS, bleu foncé) sont aussi indiqués sur la figure.

L'intégration du gène d'intérêt dans le bacmide se fait par transposition des éléments Tn7 provenant d'un plasmide donneur (pFastBac^{TM}) au site cible mini-*att*Tn7 présent dans le bacmide. La transposition est effectuée par quatre protéines (tnsA à D) produites à partir d'un plasmide helper présent dans les cellules DH10B$_{AC}^{TM}$. Ce plasmide helper contient aussi une origine de réplication (Ori) et un gène de résistance à la tétracycline (Tet) (Figure 24). Le plasmide donneur pFastBac est composé d'une origine de réplication Ori, d'un gène de résistance à l'ampicilline (Amp), et de la cassette de transposition. Cette cassette de transposition est bordée par les éléments de transposition (Tn7R et Tn7L) et contient : un gène de résistance à la gentamicine (Gen), le promoteur de la polyhédrine du virus AcNPV (P_{polh}), des sites de clonage multiple (MCS) permettant l'intégration du gène d'intérêt (GRK par exemple) et le terminateur de transcription SV40 poly(A) (Figure 24). Ainsi, l'expression du gène d'intérêt est sous le contrôle du promoteur de la polyhédrine. L'intégration de la cassette de transposition dans le bacmide interrompt l'expression du peptide α, rendant ainsi la β-galactosidase non fonctionnelle. En présence de Bluo-gal, ce dernier n'est pas clivé et les colonies appararaissent blanches au lieu de bleues (sélection blanc/bleu). Le bacmide purifié à partir des colonies blanches est

ensuite utilisé pour infecter les cellules d'insectes et obtenir des particules virales qui, après amplification, sont utilisées pour produire la protéine d'intérêt.

3.1.2.2 Clonage et cultures des GRK2 et GRK6 humaines en cellules d'insectes

Les étapes de production de GRK2 et GRK6 sont identiques et elles sont présentées ci-dessous.

3.1.2.2.1 Clonage en système d'expression des vecteurs baculovirus (BEVS)

Le clonage des ADNs codants pour les GRK2 et GRK6 se fait en deux étapes : **1**) Clonage de l'ADN des GRKs dans le plasmide donneur, pFastBac™ ; **2**) Transposition du plasmide donneur au bacmide.

1) Les ADNs codants pour les protéines GRK2 et GRK6 ont été conçus à partir des séquences protéiques humaines auxquelles nous avons ajouté un tag hexa-histidine côté C-terminal, et clivable à la 3C pour GRK2 (Figure 25).



Figure 25 : Séquence des protéines recombinantes GRK2 et GRK6 produites au laboratoire. Les séquences protéiques des GRK2 et GRK6 humaines ont été extraites à partir d'UniProt (https://www.uniprot.org) (code P2598, P43250 respectivement). Un tag histidine (rouge) et un site de clivage à la preScission protéase 3C (bleu) ont été ajoutés en C-terminal. Les cystéines connues pour être palmitoylées dans GRK6 sont indiquées en vert (dernière ligne de la séquence).

Les ADNs ont été commandés chez Fisher et clonés en vecteur pFastBac™ (Life Technologie) selon le protocole de clonage In-Fusion (In-Fusion HD cloning Plus, Takara). L'intégration du plasmide dans les cellules *E. coli* (souche Stellar) est effectuée par choc thermique à 42 °C. Ensuite, 500 µL de milieu riche SOC (milieu Luria Broth super optimal avec répression des catabolites) est ajouté et les cellules sont incubées 2 h à 37 °C. La suspension bactérienne est ensuite étalée sur gélose

LB-agar contenant 50 µg/mL d'ampicilline, permettant de sélectionner les bactéries transformées par le vecteur pFastBac™ qui contient un gène de résistance à l'antibiotique. Puis les cellules sont incubées ON à 37 °C. Le clonage a été vérifié par séquençage (Genewiz).

2) Le pFastBac™ contenant le gène de la GRK2 ou de la GRK6 est ensuite transformé par choc thermique à 42 °C, dans bactéries compétentes *E. coli* DH10BAC™ (Fisher), où la transposition a lieu. Ensuite, le mélange est laissé 1 min à 20 °C puis 950 µL de milieu riche SOC est ajouté et les cellules sont incubées 4 h à 37 °C. Le mélange est étalé sur boîte de pétri LB-agar contenant 50 µg/mL de kanamycine, 7 µg/mL de gentamicine, 10 µg/mL de tetracycline, 100 µg/mL de Bluo-gal et 40 µg/mL d'IPTG. Les antibiotiques permettent de sélectionner les bactéries qui contiennent le plasmide pFastBac, le plasmide helper et le bacmide (Figure 24). Le Bluo-gal permet de distinger à l'œil nu les bactéries d'intérêt (blanche), où la transposition du gène de la GRK dans le bacmide a eu lieu. Les cellules sont mises à incuber 2 jours à 37 °C puis 2 jours à 4 °C.

Après sélection de cinq colonies blanches (clone 1 à 5) et d'une colonie bleue (clone 6, contrôle négatif de transposition), les bacmides de ces colonies sont purifiés à l'aide du kit Qiagen (Qiagen Spin Miniprep Kit) pour être vérifiés. La vérification du clonage par séquençage n'est pas viable techniquement à cause de la taille du bacmide (~ 130 kb). Par conséquent, les bacmides ont été vérifiés par réaction de polymérisation en chaîne (PCR) avec différentes amorces (Table 8).

Amorces	Séquences
Amorce sens externe PUC/M13 F	5′-GGATTATTCATACCGTCCCA-3′
Amorce anti-sens externe PUC/M13 R	5′-AATCAGCCATACCACATTTG-3′
Amorce sens interne GRK2 F	5′-TTGATGAAGAGGACACTAAGGGC-3′
Amorce anti-sens interne GRK2 R	5′-TTTTGCCCGTATCGGCCTTTCTA-3′
Amorce sens interne GRK6 F	5′-CTGTCTGGTTCTCACGCTCATG-3′
Amorce anti-sens interne GRK6 R	5′-GGCATCCCAAAGCCCACCAA-3′

Table 8 : Séquence des amorces utilisées pour vérifier par PCR l'intégration du gène GRK dans le bacmide.

Trois sets d'amorces sont utilisés : le couple 1 composé des amorces sens et anti-sens PUC/M13, le couple 2 composé de l'amorce sens externe PUC/M13 et de

l'amorce anti-sens interne spécifique à la GRK, et le couple 3 composé de l'amorce sens interne spécifique à la GRK et de l'amorce anti-sens externe PUC/M13 (Figure 26). Les produits PCR ont ensuite été analysés par électrophorèse sur gel d'agarose à 0.7 % et les bandes ont été visualisées à la lumière Ultra Violette (UV).

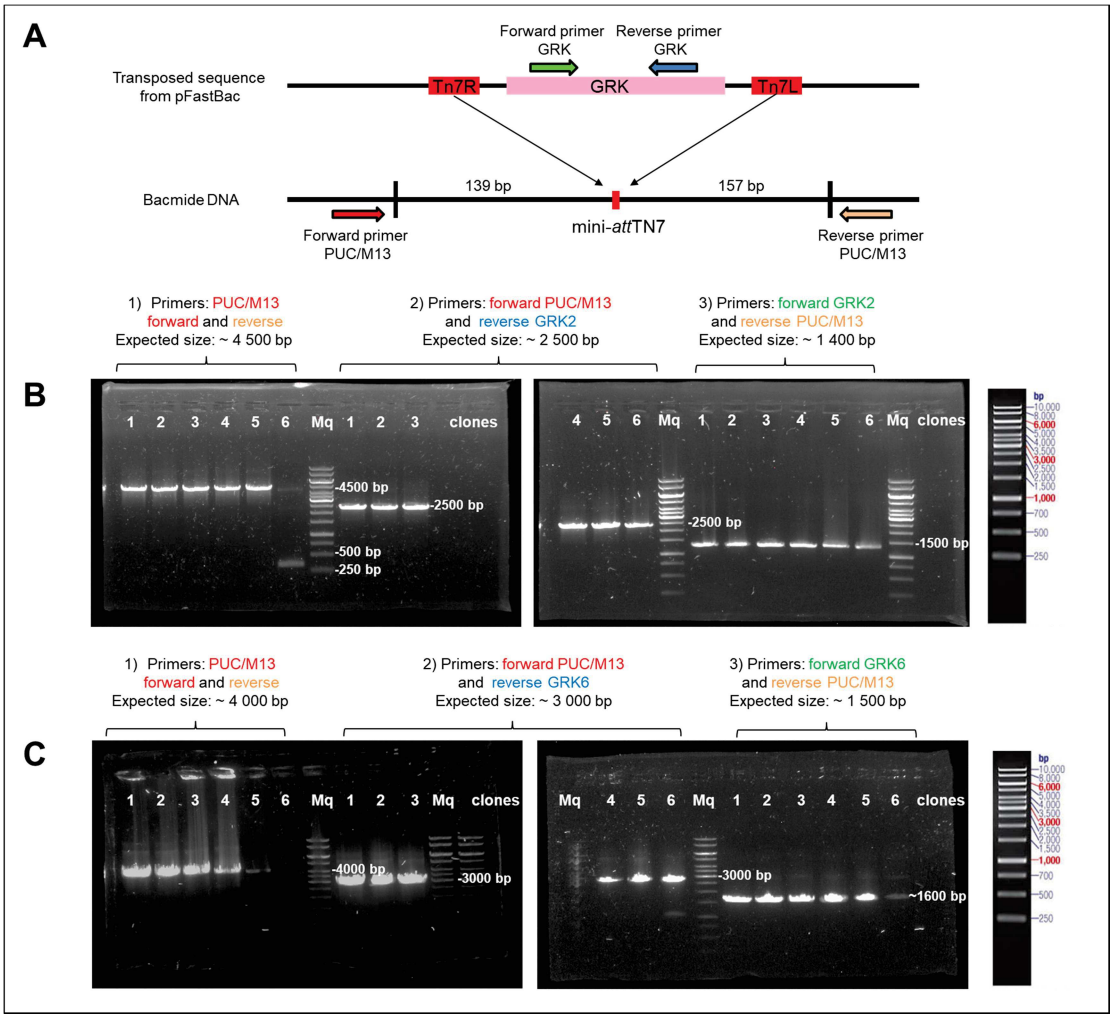


Figure 26 : Vérification par PCR de l'intégration du gène GRK dans le bacmide. A) Représentation schématique de l'insertion de la cassette de transposition (en haut) dans le bacmide (en bas). Les élements importants pour la transposition sont colorés comme suit : sites spécifiques de transposition (Tn7R/L), rouge) ; séquence ADN de la GRK, rose, séquence d'attachement du transposon bactérien Tn7 (mini-attTn7), rouge. Les amorces spécifiques au gène GRK (sens en vert et anti-sens en bleu) et au bacmide (amorce PUC/M13 sens en rouge et anti-sens en orange) sont dessinées au-dessus des séquences ADNs. Les gels agarose pour **B)** GRK2 et **C)** GRK6 montrent les résultats PCR réalisées avec les trois couples d'amorces. Les couples d'amorces utilisés, ainsi que les tailles des produits PCR attendus pour chaque PCR, sont représentés au-dessus des gels et numérotés de 1 à 3. Sur les gels, les échantillons déposés dans chaque puits sont indiqués (clones 1 à 6 et marqueur de taille) ainsi que la taille des bandes obtenues. La taille des bandes composant le marqueur de taille est indiquée sur la gauche.

Pour GRK2, les bacmides 1 à 5 présentent des bandes à ~ 4 500, ~ 2 500 et ~ 1 400 pb avec les couples amorces 1, 2 et 3 respectivement (Figure 26B). Les tailles attendues étant de 4 500, 2 400 et 1 300 pb ces bacmides ont intégré le gène de GRK2. Le bacmide contrôle présente une bande attendue à ~ 300 pb avec le couple d'amorce 1, taille attendue en l'abscence d'intégration dans ce gène. Cependant des bandes sont aussi visibles avec les couples d'amorces 2 et 3. L'hypothèse la plus probable est que le bacmide de la colonie 6 de GRK2 est été contaminé par l'ADN d'une colonie blanche. Pour GRK6, les bacmides 1 à 4 présentent des bandes à ~ 4 000, ~ 3 000 et entre 1 500 – 2 000 pb pour les couples 1, 2 et 3 respectivement (Figure 26). Pour le bacmide 5, la bande à ~ 4 000 pb est très faible. Les bactéries dont est issu ce bacmide ont peut-être peu intégré le bacmide. Concernant le bacmide contrôle 6, la résolution du gel n'était pas suffisante pour visualiser des signaux à faible poid moléculaire à cause d'un temps de migration trop élevé. Un nouveau gel agarose a donc été effectué (Figure 27).

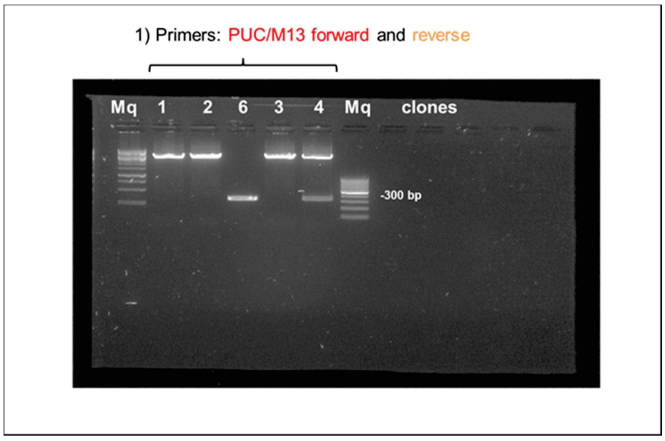


Figure 27 : Vérification par PCR de l'intégration du gène GRK6 dans le bacmide avec les amorces sens et anti-sens PUC/M13. Sur le précédant gel (Figure 26), les bandes à ~ 300 pb n'étaient pas visibles. Nous avons donc fait migrer moins longtemps un nouveau gel et utilisé 2 marqueurs de tailles. Sur la gauche, un marqueur de tailles élevées a été déposé (GeneRuler 1 kb, Thermo Scientific) et sur la droite un marqueur de petites tailles a été utilisé (Gene ruler 100 pb, Thermo Scientific). Entre les marqueurs, les ADNs amplifiés à partir des bacmides 1 à 6 sont indiqués (à l'exception du bacmide 5, non déposé sur le gel).

Tout comme pour GRK2, le profil du bacmide 6 (contrôle) est particulier. Il présente bien une bande à ~ 300 pb avec le couple d'amorce 1 (Figure 27) mais des bandes sont visibles avec les couples d'amorce 2 et 3 alors qu'aucune bande n'est attendue (Figure 26). Comme pour GRK2, l'hypothèse la plus probable est que le bacmide de

la colonie 6 de GRK6 ait été contaminé par l'ADN d'une colonie blanche. Par ailleurs, le bacmide 4 présente aussi une bande à ~ 300 pb, ce qui indique que le bacmide de la colonie 4 de GRK6 est été contaminé par l'ADN d'une colonie bleue.

Les bacmides non-contaminés et présentant les tailles attendues sont conservés à – 20 °C.

3.1.2.2.2 Cultures de cellules d'insectes Sf9

Durant les cultures de cellules *Sf9*, la viabilité et le nombre de cellules sont suivis et déterminés par comptage automatique (compteur EVETM, Avantor) après mise en évidence des cellules mortes au bleu de trypan (0.4 %).

Un bacmide, dont la séquence a été vérifiée par PCR, est sélectionné et utilisé pour la transformation de 15 millions de cellules d'insectes *Sf9* à l'aide d'un agent de transfection (CellFectin reagent, Mirus/Euromedex) dans 10 mL de milieu de culture (EX-CELL 420 Serum free, Sigma). Ensuite, les cellules sont incubées 3 jours à 28 °C. Trois jours plus tard, 50 à 70 % des cellules sont mortes (exemple Figure 28A) et le surnageant de culture contenant les virus est récupéré par centrifugation à 600 g. Ce premier stock de virus recombinants, nommé P1, est un stock à petit volume et à faible quantité virale. S'en suivent des étapes d'amplifications où le stock P1 est uitilisé pour réaliser de nouvelles infections des cellules *Sf9* (Figure 28B). Lors des étapes d'amplification, 1 mL du stock P1 a servi à infecter 200 millions de cellules *Sf9*. Après trois jours d'incubation à 28 °C, les cellules sont centrifugées à nouveau et un stock P2 est obtenu. Pour GRK6, après 3 jours d'incubation les cellules présentaient un taux de mortalité de 20-30 % contre 50-70 % attendu. Cela peut être dû à une quantité virale insuffisante. C'est pourquoi une étape additionnelle d'amplification a été effectuée à partir du stock P2 pour obtenir un stock P3 (Figure 28B).

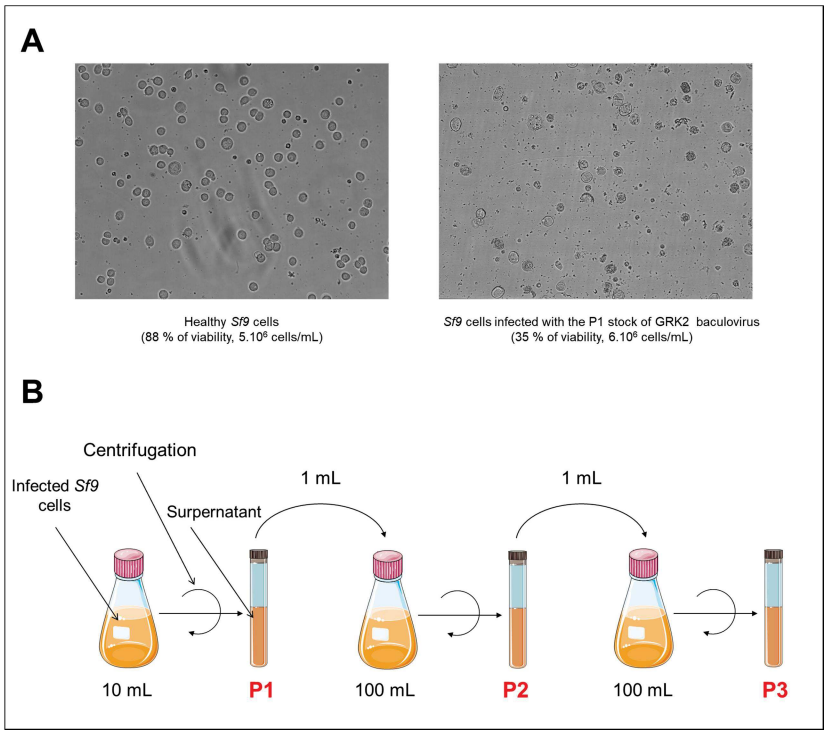


Figure 28 : Infection des cellules *Sf9* par les virus recombinants. A) Obtention du stock viral P2 de GRK2. Les cellules *Sf9* saines (panel de gauche) sont infectées à ratio 1/100 (volume/volume, virus P1/*Sf9*) et incubées trois jours à 28 °C, sous agitation. Trois jours plus tard, ~ 65 % des cellules sont mortes (panel de droite) à cause de l'infection virale. Les virus sont récupérés par centrifugation des cultures et récupération du surnageant (obtention du stock P2). **B)** Cycles d'amplification virale. Les cellules *Sf9* saines sont infectées par un bacmide recombinant et incubées trois jours à 28 °C, sous agitation. Trois jours plus tard, la culture est centrifugée et le surnageant contenant les virus est récupéré (stock P1). Ce stock à petit volume est utilisé pour infecter de nouvelles cellules saines de manière à augmenter le stock viral (étape d'amplification). Pour les étapes d'amplification, 1 mL de stock P1 ou P2 sert à infecter 100 mL de cellules *Sf9* à 2 millions/mL. Les stocks viraux sont récupérés par centrifugation.

Lors de la production des protéines GRK2 et GRK6, de grands volumes de culture sont effectués : 1 000 millions de cellules sont infectées par 5 mL de virus P2 de GRK2 (dilution 1/100ème) ou 4 000 millions de cellules sont infectées par 40 mL de P3 de GRK6 (dilution 1/50ème) (le rationel de ces ratios est expliqué dans le paragraphe suivant). Après deux jours d'incubation à 28 °C, les cellules sont centrifugées à 2 500 g pendant 15 min à 20 °C. Le culot de GRK2 est repris dans 40 mL de tampon contenant 50 mM Tris pH 7.5, 300 mM NaCl, 0.02 % Triton X100 (volume/volume, v/v), 2 mM DTT, 1 mM benzamidine et une pastille d'anti-protéase (Complete, Roche). Celui de GRK6 est repris dans 60 mL de tampon contenant 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM DTT, 1 mM benzamidine et une pastille d'anti-protéase

3.1.2.3 Test d'expression des protéines GRK2 et GRK6

L'expression de protéines en système baculovirus nécessite de trouver le bon équilibre entre la quantité de virus produit et la mort des cellules hôtes *Sf9*. Pour ce faire, plusieurs ratios P2:*Sf9* (volume/volume) ont été testés : 1/50ème, 1/100ème et 1/200ème Des cultures en petits volumes (10 mL) ont été effectuées et une première étape de purification sur billes His-trap a été effectuée. Des aliquotes issues des lavages et des élutions de l'His-trap en batch ont été conservées et analysées par Western blot, avec un anticorps anti-histidine couplé à l'HRP (HorseRadish Peroxidase). Après addition des substrats chimioluminescents de l'enzyme HRP (SuperSignal West Pico PLUS, Thermo scientific) sur la membrane de Western blot, la chimioluminescence est mesurée par exposition automatique grâce à l'appareil Amersham Imager 600.

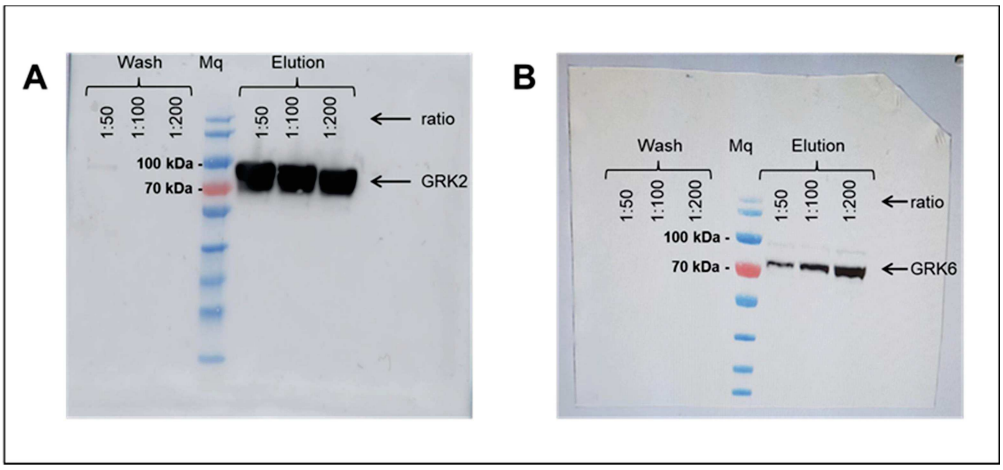


Figure 29 : Test d'expression des protéines GRK2 et GRK6 à différents ratios bacmide:*Sf9*. Western blot des aliquotes de **A)** GRK2 et **B)** GRK6 récupérées après lavage et élution de la résine His-trap. Les échantillons tagués histidine ont été marqués par un anticorps anti-His couplé à la protéine HRP. Les signaux correspondant aux protéines GRK2 et GRK6 sont indiqués par des flèches et les tailles sont indiquées sur la gauche du marqueur (Mq). Les ratios bacmide:*Sf9* sont aussi notés dans les puits.

Pour GRK2 et GRK6, aucun signal n'est obtenu dans les fractions lavées (Figure 29). Pour les fractions correspondant à l'élutions de GRK2, on observe des signaux entre 70 et 100 kDa (Figure 29A). Cela est en accord avec la taille théorique de GRK2 qui est de 81.2 kDa. De plus GRK2 peut présenter des modifications post-traductionnelles, comme des phosphorylations, pouvant faire augmenter son poids moléculaire (Pitcher *et al*, 1999). L'intensité des signaux semble similaire entre les

trois dilutions de virus employées. Cependant le signal étant saturant, l'analyse de l'intensité des signaux n'est pas fiable. Dans les trois cas, la protéine GRK2 est bien exprimée. Pour GRK6, les fractions d'élution présentent des signaux légèrement supérieurs à 70 kDa (Figure 29B). Tout comme GRK2, GRK6 peut présenter des modifications post-traductionnelles, ce qui peut expliquer la légère différence avec le poids moléculaire théorique attendu de 66.8 kDa.

Par ailleurs, l'intensité des signaux varie beaucoup entre les différentes dilutions. Le ratio 1/50ème est celle qui donne la bande la plus intense et donc l'expression la plus forte de GRK6. On utilisera donc les ratios 1/100ème et 1/50ème pour GRK2 et GRK6, respectivement.

3.1.2.4 Purification de GRK2 et GRK6

Les purifications de GRK2-His et GRK6-His ont été mises au point à partir des protocoles standards de purification des GRKs (Sohlemann *et al*, 1993; Lodowski *et al*, 2006; Sterne-Marr *et al*, 2013). La purification repose sur la réalisation successive de différentes chromatographies dans l'odre suivant : chromatographie d'affinité histidine, chromatographies échangeuses d'ions en tandem et enfin une chromatographie d'exclusion. Les tampons utilisés à chaque étape de la purification sont présentés en Table 9.

Etape de purification	Tampon de GRK2
Lyse et Affinité (tampon A)	50 mM Tris pH 7.5, 300 mM NaCl, 0.02 % Triton X100 (v/v), 2 mM DTT, 1 mM benzamidine, pastille d'anti-protéase (Roche)
Echangeuse d'ions (tampon B)	50 mM Tris pH 7.5, 50 mM NaCl, 0.02 % Triton X100 (v/v), 2 mM DTT, 1 mM benzamidine
Exclusion (tampon C)	50mM Tris pH 7.5, 150mM NaCl, 0.02 % Triton X100 (v/v), 2 mM DTT, 1 mM benzamidine

Table 9 : Tampons de purification de GRK2. Les tampons utilisés pour la purification de GRK6 sont identiques à l'exception que le Tris a été remplacé par de l'HEPES et que le triton a été supprimé car il inhibe l'activité de GRK6 (Loudon & Benovic, 1997).

3.1.2.4.1 Lyse cellulaire

Après décongélation des culots à température ambiante, les cellules *Sf9* sont lysées par pression ('French press') à ~ 700 bar. Les débris cellulaires sont éliminés par

centrifugation à 47 000 g pendant 20 min, à 10 °C. Le surnagent est récupéré et filtré sur une membrane de 0.22 µM.

Les étapes suivantes de la purification sont effectuées à température ambiante. Les fractions issues de chaque étape sont analysées par électrophorèse sur gel SDS-PAGE ainsi que par Western blot.

3.1.2.4.2 Chromatographie d'affinité histidine

La solution protéique est injectée sur une colonne d'affinité histidine (5 mL HisTrap HP column, GE Healthcare) équilibrée en tampon A. Après lavage de la colonne, les protéines sont éluées par gradient linéaire de tampon A contenant 500 mM d'imidazole pour GRK2 et 300 mM d'imidazole pour GRK6.

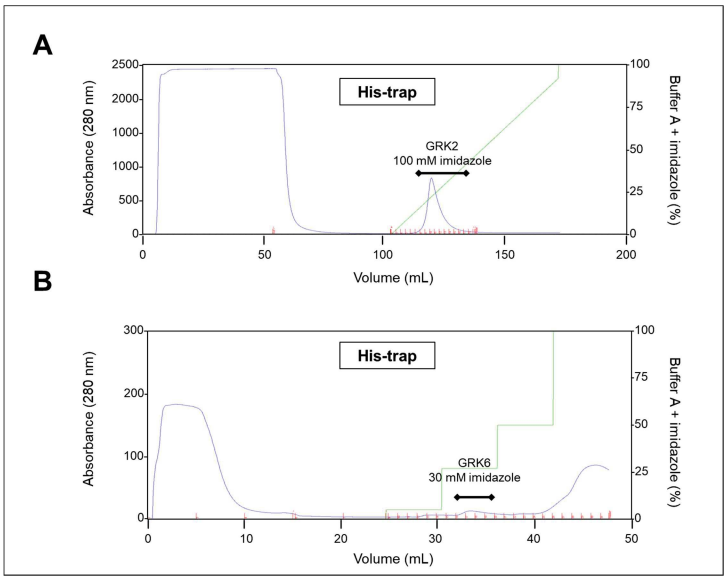


Figure 30 : Chromatogrammes issus de la chromatographie d'affintié de GRK2 (A) et GRK6 (B). Sur chaque chromatogramme, l'absorbance à 280 nm (bleu) et la concentration en tampon d'élution (vert) sont représentées en fonction du volume. Les fractions (rouge) contenant les protéines GRKs sont représentées par un segment noir.

Il est intéressant de noté que, même si GRK2 et GRK6 possèdent le même tag de purification composé de 6 histidines, elles ont des volumes d'élutions différents sur l'His trap : GRK2 élue à 100 mM en imidazole alors que GRK6 élue à 30 mM en imidazole (Figure 30). Même si la chromatographie d'affinité histidine permet une sélection des protéines contenant un tag histidine, des interactions ioniques non-spécifiques peuvent avoir lieu avec la matrice (atome de nickel chargé positivement, Ni^{2+}). A pH 7.5, GRK2 présente une charge globale négative suppérieure à celle de

GRK6 (point isoélectrique théorique de 6.84 pour GRK2 et 8.33 pour GRK6). Si des interactions ioniques ont lieu avec la matrice, cela peut potentiellement expliquer que GRK2 élue plus tardivement que GRK6.

3.1.2.4.3 Chromatographie échangeuse d'ions

Cette étape de purification est composée d'un montage en série de deux colonnes : une colonne Q-sépharose (HiTrap Q HP, GE Healthcare) suivie d'une colonne SP-sépharose (HiTrap SP HP, GE Healthcare). Les protéines sont injectées sur les colonnes, équilibrées en tampon B. Après lavage, la colonne Q-sépharose est démontée et l'élution est effectuée par gradient en tampon B contenant 500 mM de NaCl pour GRK2 et GRK6.

En sortie de colonne d'affinité, les échantillons sont dans un tampon riche en sel (300 mM NaCl). Cependant, les colonnes échangeuses d'ions sont sensibles à la force ionique et l'échantillon doit être injecté à une faible concentration saline. Par conséquent, avant injection sur les colonnes échangeuses d'anions (Q-sépharose) et de cations (SP-sépharose) montées en tandem, l'échantillon protéique est dilué afin d'obtenir une faible concentration en sel (50 mM NaCl). Dans le cas de GRK6, cette dilution brutale de la concentration en sel a entrainé la précipitation de la protéine. Ainsi l'ordre des chromatographies d'affinité et échangeuses d'ions a été inversé.

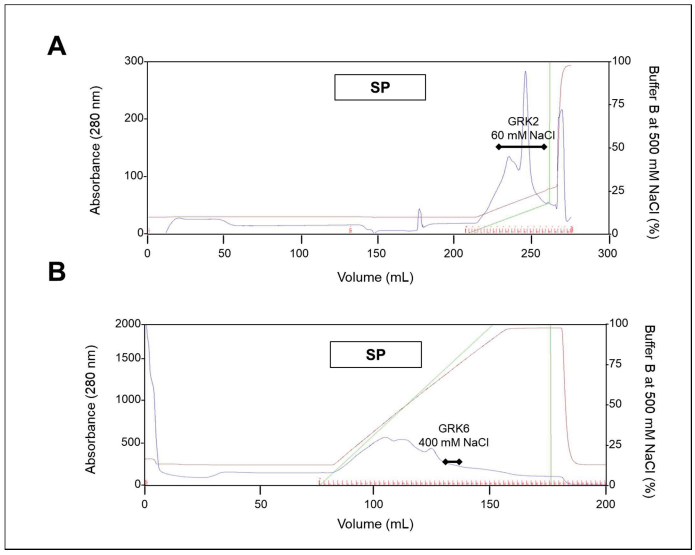


Figure 31 : Chromatogrammes issus de la chromatographie échangeuse de cation de GRK2 (A) et GRK6 (B). Sur chaque chromatogramme, l'absorbance à 280 nm (bleu) et la concentration en tampon d'élution (vert) et la conductimétrie (marron) sont représentées en fonction du volume. Les fractions (rouge) contenant les protéines GRKs sont représentées par un segment noir.

La différence de charge discutée précédement est aussi mise en évidence par l'ordre d'élution des protéines en sortie de colonne échangeuse de cations (SP-sépharose). En effet, GRK2 qui est chargé plus négativement que GRK6 élue à 60 mM NaCl, alors que GRK6 élue à 400 mM NaCl (Figure 31).

3.1.2.4.4 Chromatographie d'exclusion de taille (SEC)

Pour GRK2, les fractions issues de la chromatographie échangeuse de cations, et contenant la protéine, sont mélangées et concentrées pour être injectées en chromatographie d'exclusion (HiLoad 16/60 Superdex S200, GE Healthcare) après équilibration en tampon C. Pour finir, les fractions protéiques pures sont concentrées à 100 µM et conservées au congélateur à -80 °C. La production de GRK2 a un rendement ~ 7 mg/L de culture.

Dans le cas de GRK6, à la fin de l'étape précédente, peu de protéines a été récupérée et GRK6 n'a pas été injectée sur la colone de chromatographie d'exclusion. Le faible rendement obtenu pour GRK6 (~ 0.3 mg/L de culture) s'explique par le fait que les cultures utilisées pour sa production était contaminées par des levures. On peut donc s'attendre à obtenir un meilleur rendement avec l'utilisation d'anti-phongiques lors des cultures.

3.1.2.5 Mesure de la température de dénaturation de GRK2

La température de dénaturation a été obtenue à l'aide du ratio 350/330 nm mesuré par le Tycho (Nanotemper). Les échantillons de GRK2 à 28 µM sans glycérol et à 18 µM contenant 25 % glycérol ont été placés dans des capillaires. La fluorescence intrinsèque des tryptophanes et des tyrosines a été mesurée à 350 nm et 330 nm respectivement, en fonction de la température. La température de dénaturation est obtenue par dérivation d'ordre 1 de la courbe précédente dans GraphPad Prism.
Les résultats sont présentés en Annexe 2, p. 278.

3.2 CARACTERISATION BIOPHYSIQUE

3.2.1 Caractérisation globale des IDPs

Les protéines intrinsèquement désordonnées (IDPs) sont des protéines particulières caractérisées par une faible teneur en structures secondaires et une grande flexibilité (Chapitre 1, revue, p. 19). Leur étude structurale et fonctionnelle a nécessité la combinaison de plusieurs techniques biophysiques. Les méthodes ayant permis la caractérisation globale des RCPG-Cters sont présentées ci-dessous. Par la suite, un paragraphe sera dédié à la RMN, méthode qui a permis l'obtention d'informations structurales au niveau résiduel.

3.2.1.1 Dichroïsme circulaire (CD)

3.2.1.1.1 Principe

La matière est capable d'absorber la lumière. Cependant, les molécules asymétriques vont absorber différemment les composantes droite et gauche de la lumière polarisée circulaire. Cette différence d'absorptivité est appelée dichroïsme circulaire (CD) ou ellipticité molaire et dépend de la conformation des molécules.

Les différentes structures secondaires des protéines possèdent des spectres CD caractéristiques qui forment une combinaison de bandes spectrales au sein d'une protéine (exemple Figure 32). Par conséquent, la mesure du CD dans le domaine de l'ultraviolet lointain (190-260 nm) est un moyen facile d'évaluer rapidement le contenu en structures secondaires présentes dans une protéine. Dans le cas des protéines contenant des régions/protéines désordonnées (IDR/IDP), le spectre CD présente une bande fortement négative près de 198 nm et un épaulement négatif faible à 220 nm (exemple Figure 32) (Woody, 2010). Le CD est donc facilement utilisable pour identifier les protéines contenant du désordre.

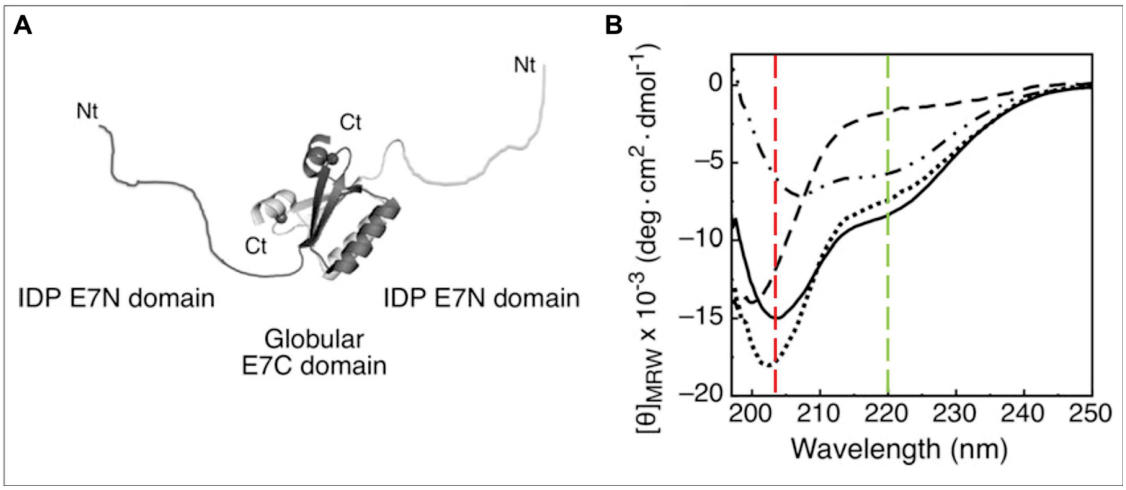


Figure 32 : **Caractérisation spectrale des domaines désordonnés et globulaires de la protéine HPV E7** (Figure modifiée à partir de (Chemes *et al*, 2012)). **A**) Représentation schématique de la protéine HPV E7 comme un dimère étendu composé de deux monomères représentés respectivement en gris foncé et en gris clair. Chaque monomère est composé d'un domaine N-terminal (E7N) intrinsèquement désordonné (IDP) et un domaine C-terminal globulaire (E7C). **B)** Spectres CD de la protéine HPV E7 complète (ligne noire), du domaine E7N intrinsèquement désordonné (ligne en tiret) et du domaine globulaire E7C (ligne pointillée et en tirets). La somme des spectres des deux domaines individuels (E7N + E7C, ligne pointillée) est similaire au spectre de la protéine complète (ligne complète). La protéine HPV E7 présente les caractéristiques par CD d'une protéine désordonnée en présence de structuration avec une bande fortement négative un peu au-delà de 200 nm (ligne pointillée rouge) et un épaulement faiblement négatif à 220 nm (ligne pointillée verte) caractéristique de la présence de structures secondaires.

3.2.1.1.2 Expérimentation

Le spectre CD des domaines C-terminaux des RCPGs préparés à 0.08 mg/mL a été enregistré dans des cuvettes en quartz (longueur optique 0.1 cm) dans l'eau et à 20 °C avec le logiciel Chirascan. L'ellipticité a été enregistrée de 190 nm à 260 nm avec un incrément de 0.5 nm, un temps d'intégration de 3 s et une bande passante constante de 1 nm. Les données et la soustraction du signal du tampon ont été traitées avec Chirascan. De manière à ne plus dépendre de la concentration des protéines et de la longueur de la cuve, l'ellipticité molaire [θ] a été convertie en ellipticité moyenne des résidus ($[\theta]_{MRW}$) par l'équation (1) :

$$[\theta]_{MRW} = \frac{[\theta]}{(nombre\ de\ liaison\ amide\ du\ squelette\ protéique)} \qquad (1)$$

Les résultats sont présentés au Chapitre 4, Articles 1 et 2, p. 129 et p. 160.

3.2.1.2 SEC-MALS

3.2.1.2.1 Principe

Un appareil de SEC-MALS est composé d'une colonne de chromatographie d'exclusion (SEC) couplée à : un détecteur de diffusion de la lumière à plusieurs angles (MALS), un détecteur de l'indice de réfraction (dRI) et parfois un détecteur UV (Figure 33A).

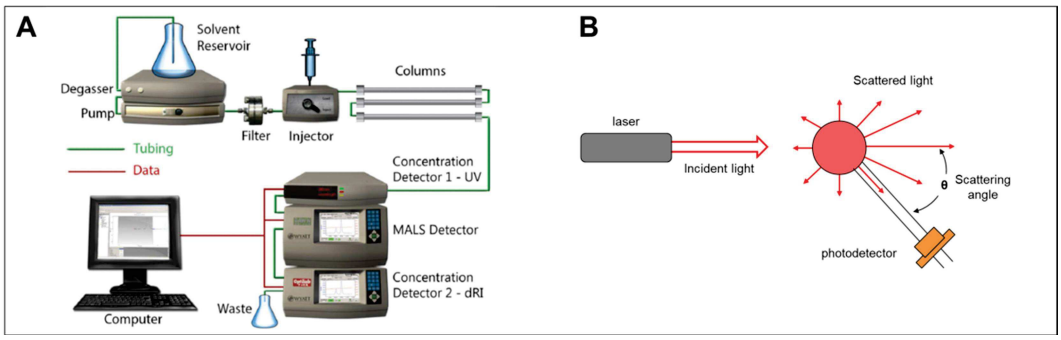


Figure 33 : **Fonctionnement d'un SEC-MALS** (Figure modifiée à partir de (Minton, 2016)). **A)** Représentation schématique d'un système de SEC-MALS. La solution protéique passe successivement à travers un détecteur UV (optionnel), un détecteur MALS et un détecteur de l'indice de réfraction. **B)** Schéma d'un appareil mesurant la diffusion statique de la lumière. Un faisceau de lumière monochromatique frappe l'échantillon et les photons diffusés sont enregistrés sur un détecteur. L'acquisition des données à plusieurs angles peut se faire à partir d'un seul détecteur dont la position peut varier ou à partir de plusieurs détecteurs positionnés à différents angles.

Le détecteur MALS mesure l'intensité moyenne diffusée par une protéine à une certaine concentration et à différents angles (θ) (Figure 33B). L'instrument mesure ensuite le rapport de Rayleigh R(θ), qui correspond à l'intensité due à la protéine (Minton, 2016). Ce rapport R(θ) est relié à la masse molaire (Mw) de la protéine grâce à une approximation de l'équation de diffusion de la lumière de Rayleigh (équation (2) ;(Wen *et al*, 1996)) :

$$\frac{K * c}{R(\theta)} = \frac{1}{Mw} \qquad (2)$$

où c est la concentration de la protéine (en mg/mL) et K une constante optique associée à l'incrément d'indice de réfraction (*dn/dc* ; variation de l'indice de réfraction n à la concentration c).

Comme expliqué au Chapitre 3, p.79 les IDPs/IDRs éluent plus rapidement que les protéines globulaires lors d'une chromatographie d'exclusion pour des masses

molaires similaires (Uversky, 2012). Ainsi, le couplage SEC et MALS permet d'identifier les IDPs via leur volume d'élution tout en confirmant leur masse molaire.

3.2.1.2.2 Expérimentations

L'analyse SEC-MALS est effectuée par couplage d'une colonne Superdex S75 10/30 (GE Healthcare) à un détecteur minidawn treos (Wyatt). Les échantillons sont préparés à 6.3, 2.2 et 5.8 mg/mL pour les peptides V2, GHSR et β2AR wild-type (wt) et à 4.2, 4.9 et 2.8 mg/mL pour leurs variants phosphomimétiques (pm) respectifs. Les données ont été enregistrées à 20 °C en tampon 50 mM Bis-Tris pH 6.7, 50 mM NaCl et analysées à l'aide du logiciel Astra (Wyatt).

Les résultats sont présentés au Chapitre 4, Articles 1 et 2, p. 129 et p. 160.

3.2.1.3 Diffusion des rayons X aux petits angles (SAXS)

3.2.1.3.1 Principe

Le SAXS est une technique basée sur la diffusion des rayons X par un échantillon, comme une protéine. L'intensité *I(s)* de diffusion des rayons X est mesurée à plusieurs angles allant de 0.1 à 20 degrés. L'intensité est ensuite moyennée radialement et tracée en fonction du vecteur de diffusion *s* (ou 'momentum transfert') pour obtenir un profil de diffusion. (Figure 34). Le vecteur de diffusion *s* est relié à l'angle de diffusion θ et à la longueur d'onde λ par l'équation (3) (Blanchet & Svergun, 2013):

$$s = 4\pi.\frac{\sin(\theta)}{\lambda} \qquad (3)$$

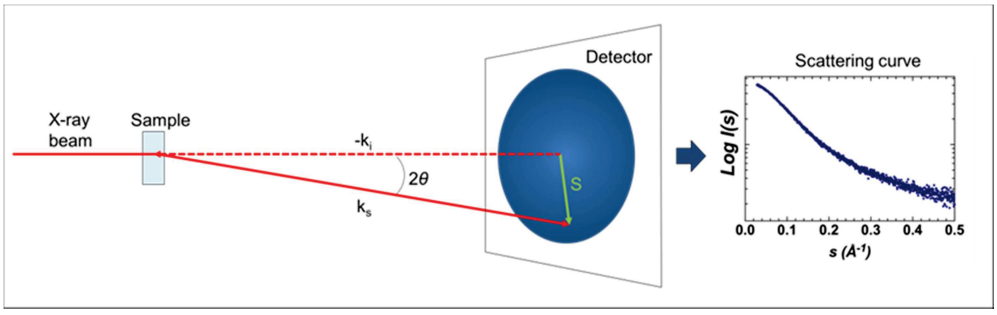


Figure 34 : Représentation schématique de la technique de SAXS (Figure modifiée à partir de (Blanchet & Svergun, 2013)). Un faisceau de rayon X monochromatique frappe l'échantillon et l'intensité des rayons X diffusés est enregistrée sur un détecteur en fonction de l'angle de diffusion. L'intensité *I(s)* est moyennée radialement et tracée en fonction du vecteur de diffusion *s* (qui est la différence entre le vecteur d'onde diffusée k_s et le vecteur d'onde incidente k_i) pour donner un profil de diffusion.

Après soustraction de la diffusion du solvant, le profil de diffusion de la protéine permet d'obtenir un large éventail d'information sur la structure et les propriétés des molécules à basse résolution (1-2 nm). La première information que le SAXS peut donner est le rayon de giration (R_g), qui représente la distance moyenne entre le centre massique d'une molécule et ses extrémités. Par conséquent, le R_g renseigne sur la taille des molécules. Le R_g est obtenu par la courbe de Guinier, qui représente ln[*I(s)*] en fonction de s^2 mais en prenant uniquement les valeurs faibles de *s* (typiquement dans la gamme $s.R_g < 1.3$ pour les protéines structurées). La courbe obtenue est linéaire et la pente donne la valeur de R_g (Figure 35).

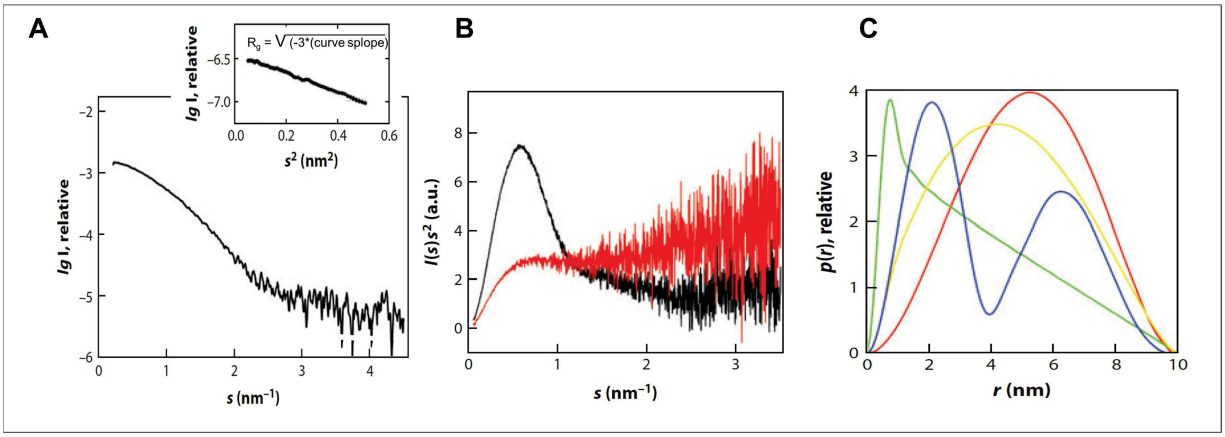


Figure 35 : Les différentes courbes pouvant être obtenues par SAXS (Figure modifiée à partir de (Blanchet & Svergun, 2013)). **A)** Exemple de courbe SAXS d'une protéine avec la courbe de Guinier en encadré. L'équation permettant d'obtenir le R_g à partir de la pente de la courbe est aussi indiquée dans l'encadré. **B)** Exemple de Kratky plot d'une protéine repliée (noire) et d'une protéine désordonnée (rouge). **C)** Exemple de fonctions de distribution des distances pour de simples formes géométriques : une sphère (rouge), un haltère (bleu), un cylindre (vert) et un disque (jaune).

Le SAXS permet aussi d'obtenir des informations sur l'état de repliement des protéines, grâce au Kratky plot. Ainsi, le Kratky plot est traditionnellement utilisé pour identifier des protéines ou régions désordonnées (Cordeiro *et al*, 2017a). Dans cette représentation ($I(s).s^2$ en fonction de *s*), les protéines globulaires présentent une courbe en forme de cloche alors que les protéines désordonnées présentent un plateau suivi d'une augmentation régulière (Figure 35). Enfin, la fonction de distribution des distances, qui est obtenue par transformation de Fourier de la courbe SAXS, donne des informations sur la forme des molécules (exemple en Figure 35) et permet l'obtention du D_{max}. Ce D_{max} est la distance maximum de la particule.

3.2.1.3.2 Expérimentations

Analyse des domaines C-terminaux des RCPGs

Les données de SAXS concernant les RCPG-Cters ont été enregistrées sur la ligne SWING au synchrotron SOLEIL vers Paris (Gif-sur-Yvette, France). La longueur d'onde λ des rayons X était de 1.03 Å et la distance échantillon-détecteur de 1.99 m. L'acquisition des données SAXS est effectuée sur des échantillons à 10 mg/mL, à 15 °C et en tampon 50 mM BisTris pH 6.7, 50 mM NaCl et 2 mM DTT. Avant exposition aux rayons X, 45 µL d'échantillon est injecté sur une colonne de 3 mL Superdex 75 5/150 GL (GE HealthCare) avec un débit de 0.2 mL/min, équilibrée dans le même tampon que celui de l'échantillon. L'intensité est mesurée en fonction de l'amplitude du vecteur de diffusion, *s* d'après l'équation (3). Le profil de diffusion du tampon seul a été enregistré avant le volume mort de la colonne (1 mL). Les profils de diffusion mesurés couvraient une plage de diffusion de $0.002 < s < 0.5$ $Å^{-1}$. La soustraction du solvant est faite automatiquement par le logiciel disponible de la ligne SWING. Le rayon de giration a été obtenu en utilisant la courbe de Guinier, avec les faibles valeurs de *s* (*s*.Rg < 1.1, pour les IDPs).

Les résultats sont présentés au Chapitre 4, Articles 1 et 2, p. 129 et p. 160.

Analyse des arrestines-2 sauvage et variant R169, seules

Les données de SAXS concernant les arrestines seules ont été enregistrées sur la même ligne SWING au synchrotron SOLEIL. L'acquisition des données SAXS est effectuée sur des échantillons à 5 mg/mL, à 15 °C et en tampon 50 mM BisTris pH 6.7, 50 mM NaCl et 2 mM DTT. Avant exposition aux rayons X, 45 µL d'échantillon est injecté sur une colonne de 3 mL Superdex 200 5/150 GL (GE HealthCare) avec un débit de 0.2 mL/min, équilibrée dans le même tampon que celui de l'échantillon. L'intensité de diffusion est mesurée en fonction de l'amplitude du vecteur de diffusion, *s* d'après l'équation (3). Le rayon de giration, R_g, a été mesuré à de très petits angles ($s.R_g < 1.3$) en utilisant l'approximation de Guinier. La fonction de distribution des distances (P(r)) est proportionnelle au nombre moyen d'atomes à une distance donnée, r (en Å). La distance maximale (D_{max}) entre deux atomes au sein de la protéine est mesurée lorsque la P(r) tend vers 0.

Les résultats sont présentés au Chapitre 4, Partie 4, p 226.

Analyse de l'arrestines-2 en complexe avec les RCPG-Cters

Les données de SAXS concernant les arrestines seules et en complexes ont été enregistrées sur la même ligne SWING que précédemment au synchrotron SOLEIL. L'acquisition des données SAXS est effectuée sur des échantillons d'arrestine wt à 80 µM (~ 4 mg/mL) seule ou en complexe avec le variant phosphomimétique de GHSR-Cter (à un ratio GHSR:arrestine-2 de 1:2.5 en molarité). Avant exposition aux rayons X, ces échantillons de 45 µL ont été injectés sur une colonne de 3 mL Superdex 200 5/150 GL (GE HealthCare) équilibrée soit en tampon 50 mM BisTris pH 6.7, 50 mM NaCl et 2 mM DTT, soit dans ce même tampon auquel a été ajouté le peptide phosphomimétique de GHSR-Cter à 180 µM. Les acquisitions et les analyses SAXS ont été effectuées comme indiqué dans le paragraphe précédent.

Par ailleurs, des expériences identiques ont été effectuées avec l'arrestine-2 WT ou R169E 80 µM (~ 4 mg/mL) seule ou en complexe avec GHSR-Cter sauvage (wt), mais sans acquisition de données SAXS.

Les résultats sont présentés au Chapitre 4, Partie 4, p. 228.

3.2.2 Estimation de la constante de dissociation

Nous avons étudié les interactions RCPG-Cter:arrestine avec différentes méthodes : la fluorimétrie différentielle (DSF), la calorimétrie par titrage isotherme (ITC), l'anisotropie de fluorescence, la thermophorèse à micro-échelle (MST) et la RMN. La RMN sera présentée dans une partie qui lui est dédiée. La DSF a déjà permis d'étudier l'interaction de l'arrestine avec un phospho-peptide synthétique du domaine C-terminal de V2R, cependant les auteurs ont utilisé l'anticorps Fab30 pour stabiliser le complexe (He *et al*, 2021). Lors de nos essais, la DSF présentait un manque de reproductivité entre les différentes expériences. Il est possible qu'une trop grande dynamique soit impliquée dans la formation/dissociation de ces complexes, ce qui expliquerait le manque de reproductibilité sans anticorps. Pour l'ITC, nos premiers essais ne donnaient pas un signal protéique suffisant par rapport au tampon. Après discussion avec Aymeric Audfray (expert ITC de chez Malvern), il semblerait que 400 µL à 1.6 mM de RCPG-Cter soit nécessaire pour une constante de dissociation (K_d) estimée à ~ 100 µM. Ces quantités étant trop élevées, nous avons mis l'ITC de côté. Nous avons donc essayé de déterminer la constante de dissociation (K_d) entre les

peptides des RCPGs et l'arrestine-2 par deux autres méthodes : l'anisotropie de fluorescence et la thermophorèse à micro-échelle (MST).

3.2.2.1 Anisotropie de fluorescence

3.2.2.1.1 Principe

L'anisotropie de fluorescence (ou polarisation de la fluorescence) est un phénomène où un fluorophore émet différentes intensités en fonction de l'axe de polarisation de la lumière. Cette caractéristique est utilisée pour détecter et mesurer la liaison entre des peptides marqués avec une sonde fluorescente et une protéine d'intérêt (Moerke, 2009). Les peptides sont marqués car leur petite taille permet une réorientation rapide en solution, et donc une faible polarisation basale. Le principe de la polarisation de la fluorescence adapté à l'étude des interactions protéine-protéine (PPI) est décrit en Figure 36. En bref, lorsque le fluorophore est lié au peptide libre en solution, son excitation par une lumière polarisée donne lieu à une émission de lumière dépolarisée due à sa réorientation pendant la durée de vie de son état excité. À l'inverse, l'interaction entre protéines réduit les mouvements Browniens du peptide et le fluorophore va émettre principalement une lumière polarisée après excitation.

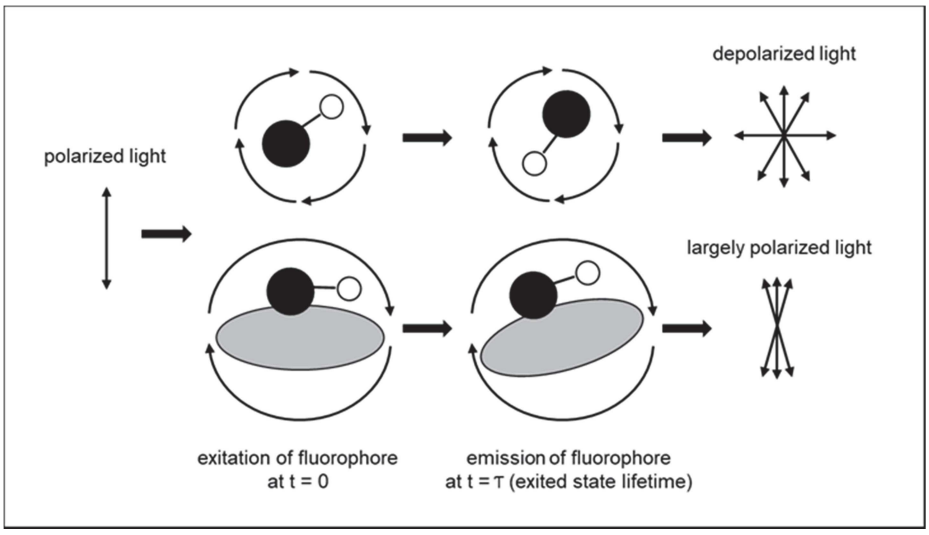


Figure 36 : Représentation schématique du principe général de la polarisation de la fluorescence (Figure extraite de (Moerke, 2009)). Lorsque la sonde fluorescente (rond blanc) attachée à un peptide (rond noir) est excitée par une lumière polarisée, elle va émettre une lumière dépolarisée. Cette dépolarisation est due à la réorientation du peptide en solution. Si le ligand est en interaction avec une protéine (ovale gris), cela va limiter sa rotation et permettre à la lumière émise de conserver sa polarisation.

L'anisotropie étant le ratio entre l'intensité polarisée et l'intensité totale, elle va augmenter proportionnellement à la fraction de peptides liés. Tracer l'anisotropie en fonction de la concentration en protéine partenaire permet l'obtention d'une courbe de liaison qui permet de déterminer la constante de dissociation (K_d) par régression non-linéaire.

3.2.2.1.2 Expérimentation

L'anisotropie de fluorescence a été utilisée pour suivre l'interaction entre le peptide phosphomimétique de GHSR-Cter marqué à l'alexa Fluor® 488 succinimidyl ester (Alexa488 SE, ThermoFisher) sur son premier groupe amine (GHSR-Cter*) et l'arrestine-2.

Le marquage de GHSR-Cter (50 µM) est effectué en tampon sans DTT et à pH 7 (50 mM Bis-Tris pH 7, 50 mM NaCl and 1 mM TCEP) par ajout de deux équivalents molaires de sonde. A pH 7, seulement la fonction amine de l'acide aminé N-terminal est réactive, ce qui permet le marquage uniquement sur le N-terminus (Figure 37).

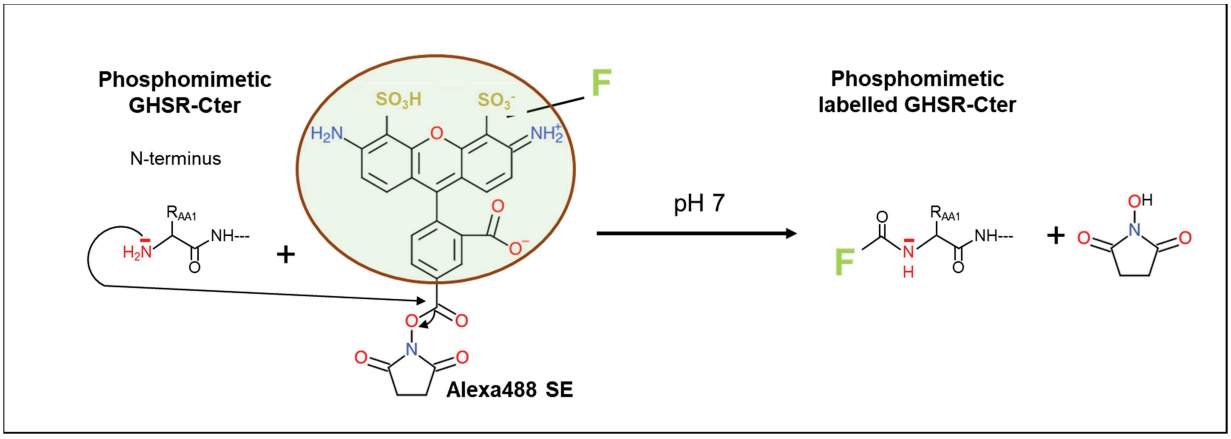


Figure 37 : Marquage du domaine C-terminal de GHSR phosphomimétique par l'alexa Fluor® 488 succinimidyl ester (Alexa488 SE). La fonction amine de l'acide aminé N-terminal est un bon groupement nucléophile à pH 7 (Brinkley, 1992). Il va aller attaquer la fonction carboxylique de l'Alexa et ainsi lier le fuorophore (F). Cette réaction est simple et permet l'obtention d'un produit stable, non réversible. Le choix du pH est très important car au-dessus de 8, les fonctions amines des lysines et arginines deviennent réactives et pourraient aussi liées le fluorophore.

Après 1 h d'incubation à 20 °C à l'abri de la lumière, la réaction est stoppée par addition de 10 µL de Tris pH 7 à 1 M, pendant 15 min d'incubation. Le Tris est une molécule contenant de nombreux groupements NH_2 qui pourront interagir avec la sonde arrêtant ainsi la réaction. L'excès de sonde est supprimé par passage de

l'échantillon sur une PD MiniTrapTM G-10 (GE Healthcare) équilibrée en tampon 50 mM Bis-Tris pH 6.7, 50 mM NaCl et 1 mM TCEP. La concentration de GHSR-Cter* est déterminée par mesure de l'absorbance à 280 nm en utilisant un coefficient d'extinction de 71 000 $M^{-1}.cm^{-1}$ et un facteur de correction de 0.11 (contribution de la sonde à 280 nm). La concentration de la sonde seule est déterminée par absorbance à 495 nm.

Les complexes sont préparés en mélangeant 1 nM de GHSR-Cter* à différentes concentrations d'arrestine-2 allant de 0 à 670 µM. Après 30 min d'incubation, la plaque noire à fond plat (Corning 384) est placée dans le spectromètre de fluorescence (Tecan Safire II). Le fluorophore est excité à 470 nm et les intensités émises sont mesurées à 530 nm. L'anisotropie de fluorescence est ensuite tracée en fonction de la concentration en arrestine pour permettre l'obtention de la constante de dissociation (K_d) par régression non-linéaire (GraphPad Prism).

Les résultats sont présentés au Chapitre 4, Article 2, p. 160.

3.2.2.2 Thermophorèse à micro-échelle (MST)

3.2.2.2.1 Principe

La thermophorèse est le nom du phénomène décrivant le mouvement dirigé des molécules dans un gradient de température. Ce phénomène dépend des propriétés moléculaires, comme la charge, la taille, la solvatation et la conformation des molécules et est donc souvent utilisé pour étudier les interactions entre protéines (Jerabek-Willemsen *et al*, 2014). Le principe de la MST et son application pour l'étude des PPI sont présentés Figure 38. En bref, une sonde fluorescente attachée à une protéine permet de suivre le mouvement de ladite protéine au cours de l'expérience de MST. Durant cette expérience, un laser infra-rouge est allumé pour créer un gradient de température local entraînant un mouvement différencié des protéines selon si elles sont libres ou liées à une protéine partenaire. La différence de fluorescence ΔF en début et en fin de chaque expérience MST (avant l'allumage et avant l'extinction du laser) est ensuite calculée et tracée en fonction de la concentration en protéine partenaire. La courbe de liaison obtenue peut être utilisée pour obtenir le K_d de l'interaction par régression non-linéaire.

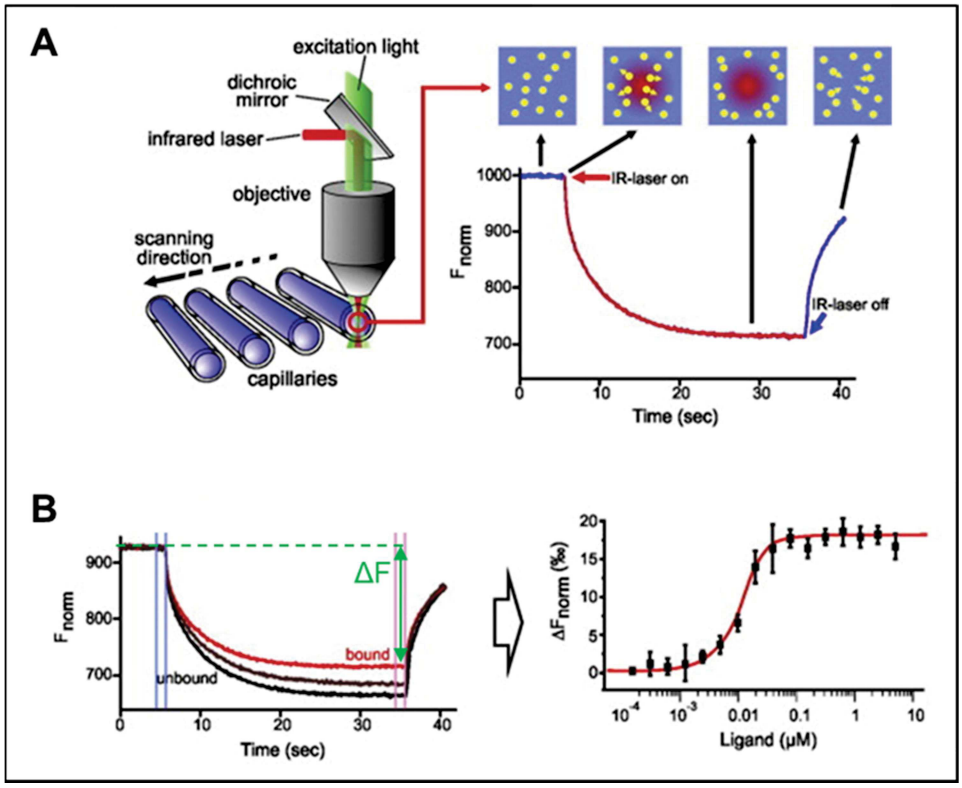


Figure 38 : **Principe de la MST** (Figure modifiée à partir de (Jerabek-Willemsen *et al*, 2014). **A)** Représentation schématique d'une expérience de MST. La MST est mesurée dans des capillaires contenant une concentration constante en protéines marquées par une sonde fluorescente et une concentration croissante en protéines partenaires. La sonde fluorescente est excitée et détectée par le même objectif que le laser infra-rouge (IR) utilisé pour chauffer localement la solution. Lorsque le laser est allumé (IR-laser on), les molécules fluorescentes vont se déplacer et sortir ou s'accumuler dans le volume chauffé. Le déplacement des molécules est suivi via le signal de fluorescence durant le temps de l'expérience de MST (~ 30 s). Après extinction du laser (IR-laser off), les molécules diffusent à nouveau librement dans le capillaire. **B)** Représentation des résultats typiques d'une interaction. Les mouvements des molécules fluorescentes (courbe noire, 'unbound') changent après interaction avec un partenaire (courbe rouge, 'bound'), donnant différentes courbes de MST en fonction des concentrations de 'libre' et de 'lié'. Le changement de fluorescence (ΔF) de chaque expérience de MST est tracé en fonction de la concentration en protéine partenaire donnant une courbe de liaison qui peut être utilisée pour obtenir le K_d.

3.2.2.2.2 Expérimentation

La MST a été utilisée pour suivre l'interaction du peptide phosphomimétique de GHSR-Cter avec l'arrestine-2 marquée à l'alexa (Arr*) Fluor® 488 Maleimide C5 (ThermoFisher).

Le marquage de l'arrestine (50 µM) est effectué dans un tampon sans DTT et à pH 7.5 (50 mM Bis-Tris pH 7.5, 50 mM NaCl and 1 mM TCEP) par ajout de sonde en excès (ratio 1:10 arrestine:sonde en mole/L). A pH 7.5, les fonctions thiols des cystéines sont réactives, ce qui permet le marquage de l'arrestine (Figure 39).

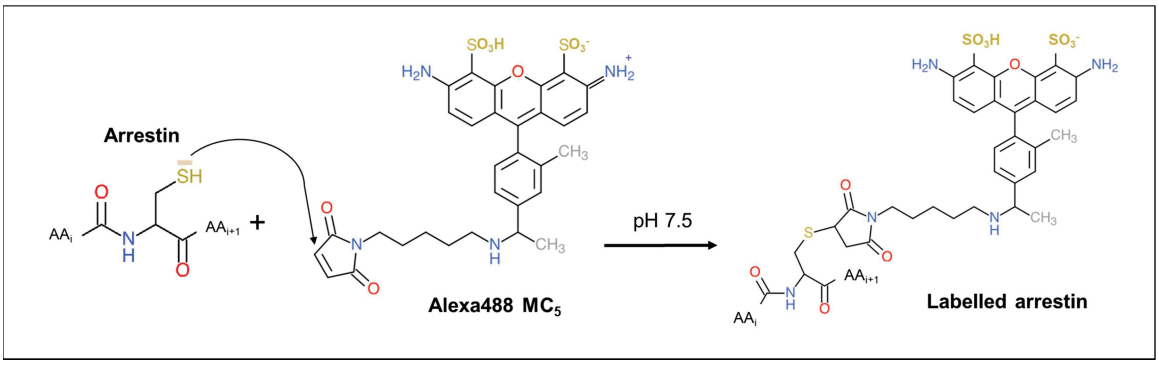


Figure 39 : Marquage des cystéines de l'arrestine par l'alexa Fluor® 488 maleimide C5 (Alexa488 MC$_5$). La fonction thiol des cystéines de l'arrestine-2 est réactive à pH 7.5 et elle va attaquer la fonction maleimide de l'Alexa, liant ainsi le fluorophore (Brinkley, 1992). Là encore, le choix du pH est important car au-dessus de 8, les fonctions maléimides peuvent être hydrolysées et devenir des acides maléamiques non réactifs.

Après 1 h d'incubation sur glace et protégée de la lumière, la réaction est stoppée par ajout de 100 mM de L-cystéine sur laquelle la sonde va être fixée, avec 15 min d'incubation à 20 °C. L'excès de sonde est supprimé par passage sur une PD MiniTrapTM G-10 (GE Healthcare) équilibrée en tampon 50 mM HEPES pH 7.5, 50 mM NaCl. La concentration de l'arrestine-2 marquée est déterminée par mesure de l'absorbance à 280 nm en utilisant un coefficient d'extinction de 71 000 $M^{-1}.cm^{-1}$ et un facteur de correction de 0.1470 (contribution de la sonde à 280 nm). La concentration de sonde libre est calculée par mesure de l'absorbance à 495 nm.

Les complexes sont préparés en mélangeant 5 nM de Arr* avec différentes concentrations de peptide GHSR-Cter (de 0.01 à 450 µM). Après 30 min d'incubation à 20 °C, les solutions protéiques sont injectées dans différents capillaires standards (NanoTemper Technologies) placés ensuite dans l'instrument de MST (Monolith NT.115 MST). La fluorescence émise est mesurée à ~ 520 nm, après excitation à ~ 470 nm. La différence de fluorescence ΔF des expériences MST est tracée en fonction de la concentration en peptide GHSR-Cter donnant une courbe de liaison. Enfin, la constante de dissociation est obtenue par régression non-linéaire (GraphPad Prism).

Les résultats sont présentés au Chapitre 4, Article 2, p. 160.

3.3 LA RESONANCE MAGNETIQUE NUCLEAIRE (RMN)

3.3.1 Principe de la RMN

La RMN est une technique de spectroscopie qui étudie les propriétés des noyaux atomiques de molécules biologiques possédant un 'spin nucléaire' non nul, placés dans un champ magnétique (Figure 40). Ce spin nucléaire peut être assimilé à un 'petit aimant' capable d'absorber l'énergie d'un rayonnement électromagnétique (impulsion radiofréquence) puis de la relâcher lors de la relaxation. Ce phénomène de résonance a lieu à une fréquence précise pour chaque atome, nommée fréquence de résonance, qui dépend de la nature et de l'environnement des atomes et du champ magnétique. Son observation permet donc l'obtention d'informations sur la structure, l'environnement et la dynamique des molécules.

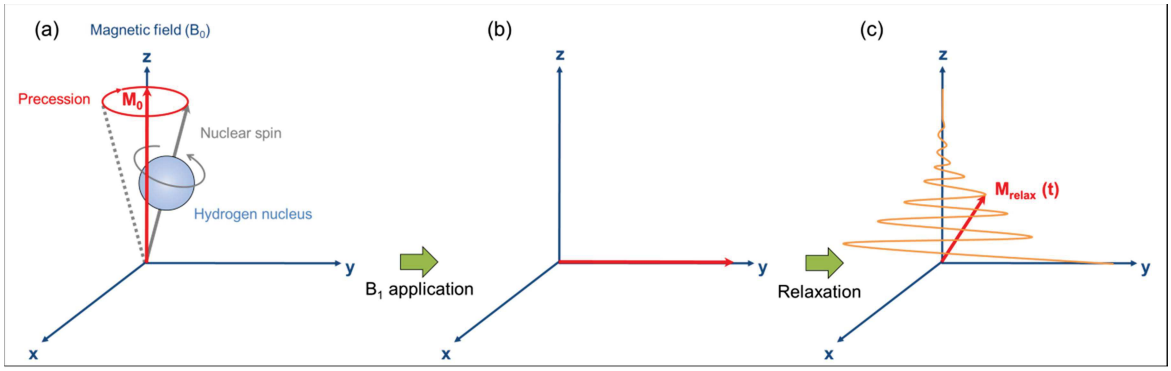


Figure 40 : **Représentation schématique du principe de la RMN.** Lorsqu'un noyau possédant un spin nucléaire (en gris) est placé dans un champ magnétique (B_0), il va précesser autour de l'axe du champ magnétique B_0 (z). Ce mouvement est représenté par le vecteur d'aimantation M_0 (en rouge) (a). L'application d'un deuxième champ magnétique B_1 (onde radiofréquence, RF) un court instant modifie le mouvement du spin dans l'axe x-y (b) qui va ensuite retourner à l'équilibre (phénomène de relaxation, en orange) en émettant de l'énergie sous forme d'ondes RF que l'on détecte (c). C'est le signal mesuré en RMN.

Pour étudier les propriétés des protéines par RMN, il est souvent nécessaire de les marquer isotopiquement et uniformément. En effet, les noyaux de carbone 12 (^{12}C) et d'azote 14 (^{14}N), qui compose les protéines, sont des isotopes abondants naturellement mais ils ne sont pas observables par RMN (spin 1). À l'inverse, leur isotope carbone 13 (^{13}C) et azote 15 (^{15}N) sont observables par RMN (spin ½) mais leur abondance naturelle respective est de 1.1 et 0.4 %. C'est pourquoi du glucose

marqué ^{13}C et/ou du chlorure d'ammonium marqué ^{15}N ($^{15}NH_4Cl$) ont été utilisés comme unique source de carbone et/ou d'azote lors de la production des protéines.

Les applications de la RMN pour l'étude des protéines sont très nombreuses. Au cours de ces travaux nous nous en sommes servi pour : obtenir des informations structurales et dynamiques sur les peptides des RCPGs, étudier les interactions entre les domaines C-terminaux des RCPGs et l'arrestine, ou encore étudier la phosphorylation des RCPGs par les GRKs.

3.3.2 Expérimentation

Tous les échantillons utilisés en RMN ont été préparés dans un tampon contenant 50 mM de Bis-Tris pH 6.7, 50 ou 150 mM de NaCl, 1 mM d'EDTA, 0.5 mM de TCEP. Ce tampon sera appelé tampon RMN dans le reste du manuscrit. Après ajout de 5 % de D_2O (Eurisotop) contenant 5 mM de DSS-d_6 (Sigma) servant de référence interne (Wishart & Sykes, 1994b), les expériences RMN ont été enregistrées à 20 °C, sur des spectromètres Bruker Bruker Avance III à 700 ou 800 MHz. Les spectromètres du laboratoire sont équipés d'une sonde cryogénique à triple résonance (^1H, ^{15}N et ^{13}C) et ils sont blindés. Les séquences d'impulsions proviennent pricipalement de la base fournie par Bruker Topspin 3.2. Concernant le traitement des données, un doublement du nombre de point ('zero-filling') est appliqué puis le signal a été traité par une fonction d'apodisation en cosinus carré dans les dimensions ^1H, ^{15}N et ^{13}C avant d'être passé en fréquence par une transformée de Fourier dans TOPSPIN (version 4.0.6, Bruker). Le traitement des spectres a été effectué dans NMRFAM-SPARKY (version 1.414, (Lee *et al*, 2015)).

3.3.3 Attribution 3D des régions C-terminales des RCPGs

Quelles que soient les futures applications d'une étude par RMN d'une protéine, il est généralement nécessaire d'attribuer les résonances des atomes la constituant. L'attribution consiste à faire la correspondance des pics observés sur un spectre RMN ^{15}N-HSQC aux acides aminés de la séquence de la protéine.

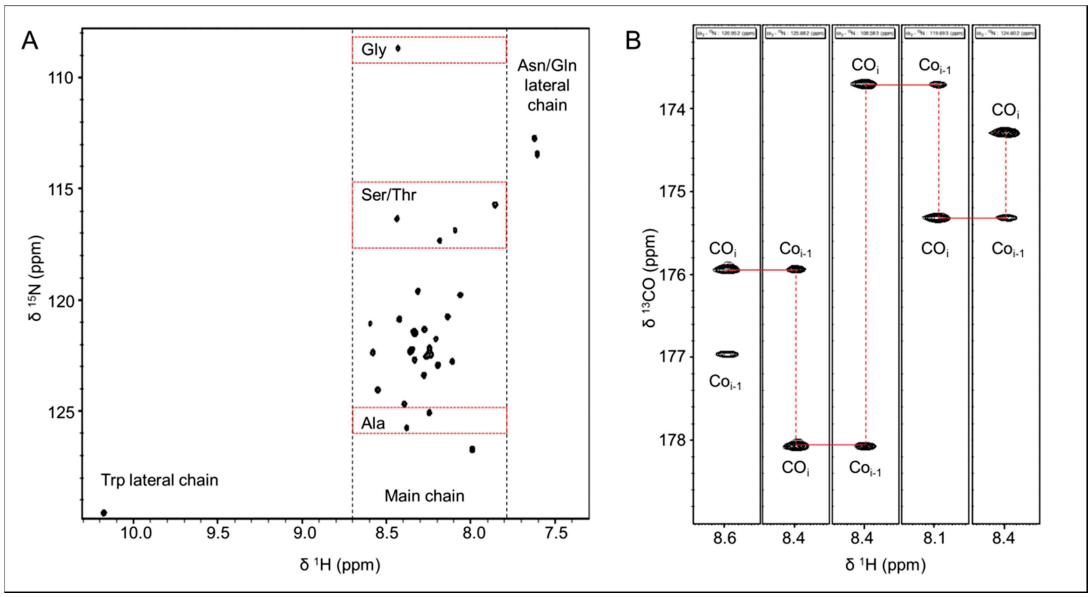


Figure 41 : Stratégie d'attribution RMN. A) Spectre 2D ^{15}N-HSQC du peptide GHSR phosphomimétique. Ce spectre montre les corrélations ^{1}H-^{15}N et correspond à l'empreinte digitale de la protéine avec 1 pic = 1 liaison amide = 1 acide aminé. La région des liaisons amides de la chaîne principale est encadrée par des traits noirs pointillées et les régions spécifiques aux résidus glycine, sérine/thréonine et alanine sont encadrées en rouge pointillé. Les fonctions amides des chaînes latérales des tryptophanes et des asparagine/glutamine sont aussi visibles. **B)** Extractions des plans ^{1}H-^{13}C issus de l'HN(CA)CO du peptide GHSR phosphomimétique. Chaque plan montre la corrélation ^{1}H-^{15}N d'un acide aminé (un pic de l'HSQC) avec le ^{13}CO du résidu i (pic le plus intense) ou du ^{13}CO du résidu i-1 (pic moins intense). L'attribution séquentielle est effectuée par correspondance des corrélations CO entre les acides aminés (trait rouge).

Le spectre ^{15}N-HSQC (Corrélation Hétéronucléaire à Simple Quantum) est un spectre à deux dimensions qui corrèle les résonances du proton et de l'azote relié par une liaison amide ^{1}H-^{15}N. Par conséquent, un spectre HSQC présente un pic par liaison peptidique amide (soit un pic par acide aminé) et des pics correspondant aux chaînes latérales des résidus glutamine et asparagine (NH_2 ~ 112 ppm en ^{15}N), tryptophane (NεH ~ 130 ppm en ^{15}N), arginine (NεH ~ 90 ppm en ^{15}N) et lysine (NH_3 ~ 50 ppm en ^{15}N, visible qu'à pH très acide) (Figure 41A). Les prolines ne sont pas visibles car leur chaîne latérale forme un cycle avec l'azote de la chaîne principale et ne porte donc pas de proton. Ainsi, le spectre HSQC sert d'empreinte digitale de la protéine étudiée et est utilisé comme base pour de nombreuses expériences RMN.

Dans le cas des protéines intrinsèquement désordonnées, une faible dispersion en dimension ^{1}H est observée (Figure 41A, (Sibille & Bernadó, 2012; Delaforge *et al*, 2018)). Cela est dû au fait que le déplacement chimique ^{1}H dépend majoritairement de l'environnement des atomes qui sont dans le cas des IDPs moins impliqués dans des structures secondaires et/ou tertiaire, ils voient donc tous le même

environnement chimique qui est principalement le tampon dans lequel la protéine est diluée. En dimension ^{15}N, le déplacement chimique dépend majoritairement du type d'acides aminés. Par exemple, les glycines sont regroupées vers ~ 108 ppm en haut du spectre, les sérines et thréonine sont juste en dessous (~ 116 ppm) et les alanines sont en bas du spectre (~ 125 ppm) (Figure 41A). Cependant, cette répartition des déplacements chimiques ^{15}N n'est pas suffisante pour l'attribution de spectres HSQC. L'attribution sans ambiguïté requière des expériences 3D triple résonance ^{1}H, ^{15}N, ^{13}C afin de reconstituer le puzzle des résonances des acides aminés composant la protéine.

Spectre RMN	Atomes observés
2D-^{15}N-HSQC	^{1}H-^{15}N
3D-HNCO	$^{13}CO_{i-1}$
3D-HN(CA)CO	$^{13}CO_{i-1}$ et $^{13}CO_{i}$
3D-HN(CO)CA	$^{13}C_{\alpha-1}$
3D-HNCA	$^{13}C_{\alpha-1}$ et $^{13}C_{\alpha}$
3D-CBCA(CO)NH	$^{13}C_{\alpha-1}$ et $^{13}C_{\beta-1}$
3D-HNCACB	$^{13}C_{\alpha-1}$, $^{13}C_{\alpha}$, $^{13}C_{\beta-1}$ et $^{13}C_{\beta}$

Table 10 : Expériences utilisées pour l'attribution de la chaîne principale des peptides des RCPGs.

Ainsi, des expériences RMN à trois dimensions ont été effectuées (Table 10). Selon l'expérience 3D, la liaison amide ^{1}H-^{15}N du squelette peptidique est corrélée avec différents atomes de carbones (CO, Cα et/ou Cβ) de l'acide aminé i et/ou de l'acide aminé précédent i-1 (Table 10). Par exemple, le spectre 3D-HNCO présente un pic par acide aminé avec une corrélation entre l'atome $^{1}H_N$ et ^{15}N de la liaison amide du résidu i et le groupement carbonyle (CO) du résidu i-1 (Figure 41A). Ainsi, pour le 'peak picking' initial permettant de démarrer le processus d'attribution, nous avons projeté ces pics de l'HNCO (détectés automatiquement dans Sparky) en 2D sur l'HSQC ^{1}H-^{15}N, afin de lever des dégénérescences de superposition de pics dues à la résolution inférieure des expériences HSQC 2D par rapport à des expériences 3D. Ceci nous sert de base pour aligner les couples i et i-1 qui permettent de reconstituer le puzzle de l'attribution. Le spectre 3D-HN(CA)CO présente les mêmes corrélations que la 3D-HNCO avec en plus une corrélation entre les atomes $^{1}H_N$, ^{15}N et CO du résidu i, donnant un second pic par acide aminé.

Ainsi, l'attribution séquentielle des atomes de la chaîne principale des protéines et des C_β a été effectuée pour les domaines C-terminaux des RCPGs dans leur forme wild-type (wt), phosphomimétique (pm) et phosphorylée par GRK2 (phopsho). Les spectres ont été enregistrés dans un tampon RMN contenant 150 mM NaCl, excepté pour la forme phosphorylée par GRK2, enregistrée dans un tampon RMN contenant 50 mM NaCl.

Les acquisitions 3D CBCA(CO)NH/HNCACB du peptide pm-V2R (50 µM) ou phospho-GHSR (50 µM) en présence d'arrestine (250 µM) ont été enregistrées en tampon RMN contenant 50 mM en sel. En effet, la baisse en sel a été nécessaire afin de minimiser les écrantages des interactions électrostatiques qui pouvaient apparaître entre les RCPG-Cters et l'arrestine-2. Ainsi, nous avons vérifié si le sel avait un effet sur la conformation des RCPG-Cters. Pour cela, nous avons enregistré les ^{15}N-HSQCs des ^{15}N-RCPG-Cter sauvages et phosphomimétiques préparés en tampon RMN contenant 50 mM NaCl ou 150 mM NaCl, ainsi que par les expériences 3D CBCA(CO)NH/HNCACB de pm-V2R (50 µM) préparé en tampon RMN contenant 50 mM NaCl (résultats au Chapitre 4, Articles 2, p. 160). Comme précédemment, ces expériences ont permis l'obtention sans ambigüité des noyaux H_N, N, CO, C_α et C_β de tous les résidus à l'exception des prolines et du premier résidu N-terminal qui sont invisibles dans un spectre RMN ^{15}N-HSQC et de l'alanine A339 du tag de β2AR-Cter.

3.3.4 Caractérisation structurale et dynamique des RCPG-Cters

Les domaines C-terminaux des RCPGs sont désordonnés. La description du comportement moléculaire d'une région désordonnée revient à devoir décrire un ensemble de conformations transitoires et dynamiques. Cette flexibilité structurale rend leur caractérisation structurale plus difficile par RMN, aussi elle empêche leur cristallisation ou leur visualisation dans des structures cristallographiques aux rayons X ou de cryo-microscopie électronique. La RMN reste la méthode de choix pour l'étude des IDPs (Sibille & Bernadó, 2012; Delaforge *et al*, 2018). Elle permet d'obtenir des informations structurales et dynamiques grâce à un jeu de descripteurs expérimentaux composés de : l'index de déplacement chimique (CSI), le couplage scalaire (^{3}J), le couplage dipolaire résiduel (RDC), les paramètres de relaxation (R_1, R_2, ^{15}N{^{1}H}-NOE hétéronucléaire) et l'augmentation de la relaxation paramagnétique (PRE ; augmentation de la relaxation paramagnétique). Ces paramètres seront

présentés dans les paragraphes suivant en mettant l'accent sur leur utilisation dans le cadre de l'analyse des protéines IDPs.

En pratique, les spectres permettant l'obtention de ces descripteurs ont été enregistrés sur le spectromètre à 700 MHz et en tampon RMN contenant 150 mM NaCl sur des échantillons dont les concentrations sont listées en Table 11. Les résultats de la caractérisation RMN des RCPG-Cters est disponible au Chapitre 4, Articles 1 et 2, p. 129 et p. 160.

Protéine	wt-V2R	pm-V2R	wt-GHSR	pm-GHSR	wt-β2AR	pm-β2AR
CSI	170	50	312	377	255	206
$^{3}J_{HNHA}$	170	93	314	263	487	476
RDC hexanol	310	93	390	263	487	476
RDC Pf1	139	138	285	250	220	212
relaxation	373	93	124	263	487	476

Table 11 : Concentration en peptides des RCPGs utilisées pour la mesure des différents paramètres RMN. Les peptides des RCPGs sont notés 'wt' pour la forme sauvage, non-phosphorylée et 'pm' pour la forme phosphomimétique. Les concentrations indiquées sont en µM pour les expériences RMN de : CSI (acquisition 3D), couplages scalaires, RDC (mesuré en milieu PEG/hexanol ou Pf1), et de relaxation (R1, R2 et hetNOE).

L'effet de la concentration sur la conformation des RCPG-Cters a été vérifié par acquisition de spectres ^{15}N-HSQCs sur des échantillons à 300 µM et 100 µM sur le spectromètre de 800 MHz. La superposition parfaite de ces spectres nous a permis de conclure que la concentration n'affecte pas leur conformation (Chapitre 4, Articles 1 et 2, p. 129 et p. 160).

La caractérisation RMN des domaines C-terminaux wt-V2R et wt-GHSR a été faite avant mon arrivée et vérifiée par mes soins en cas d'erreur.

3.3.4.1 L'index de déplacement chimique (CSI)

3.3.4.1.1 Principe

Les déplacements chimiques des carbones $^{13}C_{\alpha}$ et $^{13}C_{\beta}$ des protéines sont très sensibles à la structure secondaire du squelette polypeptidique et l'environnement local des acides aminés. Par conséquent ils sont souvent utilisés pour identifier et localiser les structures secondaires (hélice α, brin β) au sein des protéines (Wishart & Sykes, 1994a, 1994b).

Pour se faire, les déplacements chimiques expérimentaux (δ_{EXP}) sont comparés aux déplacements chimiques 'random-coil' (δ_{RC}) obtenus à partir d'une base de données. Cela donne l'indice de déplacement chimique (δ_{CSI}) (équation (4)) :

$$\delta_{CSI} = \delta_{EXP} - \delta_{RC} \qquad (4)$$

Les déplacements chimiques 'random-coil' sont des déplacements chimiques expérimentaux, très souvent obtenus à partir de peptides synthétiques. Ces peptides sont considérés être suffisamment courts (5 acides aminés) pour rester complètement déstructurés et pour pouvoir adopter toutes les conformations possibles selon le diagramme de Ramachandran (Wishart *et al*, 1995). Historiquement, ces bases de données peptidiques ont été construites à partir du peptide Ac-GGXGG-NH_2, Ac étant une acétylation du N-terminus causé par la méthode de production des peptides synthétiques (Wishart *et al*, 1995; Schwarzinger *et al*, 2001). Cependant, les glycines (G) sont des résidus très flexibles qui peuvent représenter des conformations particulières des protéines. L'équipe de Poulsen a donc créé une base de données à partir du peptide Ac-QQXQQ-NH_2, car la glutamine (Q) a une chaîne latérale composée d'un atome γ lourd (groupement $CONH_2$) similaire à celui de nombreux résidus, comme la lysine ou l'arginine (Kjaergaard & Poulsen, 2011; Kjaergaard *et al*, 2011). Malgré ces améliorations, prédire les structures secondaires résiduelles des IDPs par le déplacement chimique restait un défi car elles explorent des conformations transitoires, faiblement peuplées. Avec l'augmentation du nombre d'IDPs et d'IDRs identifiées dans la littérature, Mulder et son équipe ont développé une base de données 'random-coil' basée sur le déplacement chimique expérimental de protéines désordonnées (Tamiola *et al*, 2010; Nielsen & Mulder, 2018). Ils ont montré que leur base de données POTENCI, permet d'améliorer l'identification des structures secondaires transitoires des IDPs à partir du déplacement chimique.

3.3.4.1.2 Expérimentation

Etant donné que les domaines C-terminaux des RCPGs se comportent comme des IDRs, nous avons utilisé la base de données POTENCI (Nielsen & Mulder, 2018) pour calculer les CSI des atomes de carbones C_α et C_β. Les C_α présentent un CSI négatif lorsqu'ils sont situés en hélices et un CSI positif lorsqu'ils sont localisés dans

des régions étendues (brin bêta), et inversement pour les C_β. La combinaison de ces deux CSI permet de faire un consensus de leur CSI respectifs, le Ca etant plus sensible à l'hélicité et le Cb aux brins, ceci améliore la prédiction des structures secondaires (Wishart & Sykes, 1994a). Ainsi, pour une meilleure détermination des structures secondaires, nous avons combiné les CSI des carbones $^{13}C_\alpha$ et $^{13}C_\beta$ en les soustrayant l'un à l'autre (CSI C_α – CSI C_β). Des valeurs positives de ce CSI sont associées à une conformation en hélice alors que des valeurs négatives sont associées à une conformation étendue.

Dans le cas des sérines phosphorylées, les déplacements chimiques expérimentaux ont été soustraits aux déplacements chimiques 'random-coil' de sérines phosphorylées décrits par Kunze et son équipe (Hendus-Altenburger *et al*, 2019).

3.3.4.2 Le couplage scalaire $^3J_{HNHA}$

3.3.4.2.1 Principe

Les spins nucléaires des atomes interagissent entre eux au sein des molécules. L'interaction à travers les électrons des liaisons est appelée couplage scalaire spin-spin J.

Le couplage scalaire séparé par trois liaisons (^{3}J) entre les noyaux ^1H_N et $^1H_\alpha$ du squelette peptidique (appelé $^3J_{HNHA}$) dépend de l'angle de torsion φ entre les protons couplés d'après la loi de Karplus (équation (5) ; Figure 42 ; (Karplus & Anderson, 1959)) :

$$^3J_{HNHA} = A\,cos^2(\varphi - 60) + B\cos(\varphi - 60) + C \qquad (5)$$

où A, B et C sont des constantes de paramètrages souvent obtenues à partir de structures cristallographiques de protéines structurées.

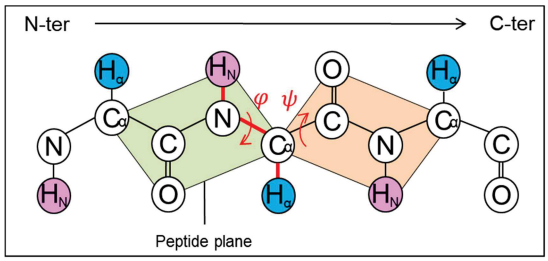


Figure 42 : Relation entre le couplage scalaire à trois liaisons $^3J_{HNHA}$ et l'angle φ du squelette protéique. La valeur de couplage entre les noyaux homonucléaires H_N (rose) et H_α (bleu) dépend de l'angle dièdre φ du squelette peptidique.

Les valeurs d'angles φ et ψ étant reliées aux structures secondaires par le diagramme de Ramachandran (Ramakrishnan & Ramachandran, 1965), les valeurs de couplage sont couramment corrélées aux structures secondaires. Les régions random-coil ont des valeurs de couplages $^{3}J_{HNHA}$ allant de 6 à 8 Hz environ. Sachant que les structures secondaires des IDPs sont transitoires, des valeurs de couplages inférieures et supérieures aux valeurs random-coil indiquent la présence d'hélice alpha ou de brin bêta, respectivement (Table 12, (Delaforge *et al*, 2018)). Par ailleurs, la conformation étendue des polyprolines hélices de type 2 (PPII) leur donne un profile CSI et RDC similaire au brin bêta. Par conséquent, uniquement le couplage scalaire peut permettrent de distinger les PPII des brins bêta grâce à leur valeurs de couplages proches de ~ 6 Hz au lieu de valeurs supérieures à 8 Hz (Oh *et al*, 2012).

Structure secondaire	Hélice α et PPII	Brin β	Random-coil
$^{3}J_{HNHA}$ (Hz)	≤ 6	≥ 8	6-8

Table 12 : Corrélation entre les valeurs de couplage $^{3}J_{HNHA}$ et les structures secondaires transitoires des IDPs. PPII signifie polyproline de type 2.

Pour améliorer l'identification des structures secondaires transitoires des IDPs, il est possible de comparer les valeurs de couplages expérimentales aux valeurs 'random-coil' obtenues à partir d'une base de donnée. Ces valeurs de couplages random-coil $^{3}J_{HNHA}$, peuvent avoir été obtenues experimentalements, à l'aide d'agent dénaturant, ou peuvent être prédits à partir des régions non structurées des structures cristallographiques (Shen *et al*, 2018).

3.3.4.2.2 Expérimentation

Nous avons mesuré les constantes de couplage $^{3}J_{HNHA}$ des domaines C-terminaux des RCPGs par enregistrement d'une 3D HNHA (Vuister & Bax, 1993) qui corrèle les résonances ^{15}N, $^{1}H_{N}$ et $^{1}H_{\alpha}$ du squelette protéique. Dans le spectre 3D, deux pics sont visibles par acide aminé : un pic diagonal (diag) qui correspond au HN et un pic croisé (cross) qui correspond au couplage H_N-H_α. Le ratio des intensités de ces deux pics donne directement la valeur de couplage $^{3}J_{HNHA}$ selon l'équation suivante (6) (Vuister & Bax, 1993)) :

$$\frac{S_{cross}}{S_{diag}} = -tan^2(2\pi J_{HNHA}\zeta) \qquad (6)$$

où S_{cross} et S_{diag} correspondent aux intensités du pic diagonal (diag) et du pic croisé (cross) respectivement et 2ξ est le temps de déphasage total (26.1 ms).

Ces valeurs expérimentales ont été comparées aux valeurs random-coil prédites par la base de données de (Shen *et al*, 2018), dont les prédictions ont montré une bonne corrélation aux valeurs expérimentales de protéines désordonnées.

3.3.4.3 Les couplages dipolaires résiduels (RDCs)

3.3.4.3.1 Principe

Le couplage dipolaire ou couplage dipôle-dipôle est l'interaction entre spins à travers l'espace. Les interactions dipolaires entre les noyaux ^{15}N et $^{1}H_N$ (D_{NH}) du squelette protéique sont régies par l'équation (7) (Chen & Tjandra, 2011) :

$$D_{NH} = -\frac{\mu_0 \hbar \gamma_N \gamma_H}{4\pi^3 r_{NH}^3} \left\langle \frac{(3cos^2\boldsymbol{\theta} - 1)}{2} \right\rangle \qquad (7)$$

où θ est l'angle entre le vecteur internucléaire ($\overrightarrow{NH}$) et le champ magnétique $\overrightarrow{\boldsymbol{Bo}}$, γ_N et γ_H sont les rapports gyromagnétiques des spins des noyaux azotes (~ -27.10^6 rad.s^{-1}.T^{-1}) et protons (~ 267.10^6 rad.s^{-1}.T^{-1}), r_{NH} est la distance entre les deux spins (1.04 Å), μ0 est la constante de perméabilité du vide (μ0 = $4\pi.10^{-7}$.m^{-1}) et $\hbar$ est la constante de Planck divisée par 2π ($\hbar$=1.05457266.10^{-34} J.s)

Le couplage dipolaire est naturellement moyenné à 0 en solution dû à la réorientation isotrope des molécules. Cependant, il est possible de le réintroduire par utilisation d'un milieu d'alignement (ou milieu orientant ou milieu anisotrope), qui confère une orientation préférentielle à la protéine. Dans le cas des IDPs, les RDCs permettent de mesurer directement l'orientation du vecteur $\overrightarrow{NH}$ par rapport à B_0.Il est ainsi possible d'en déduire les structures secondaires associées (Figure 43) (Jensen et al, 2009).

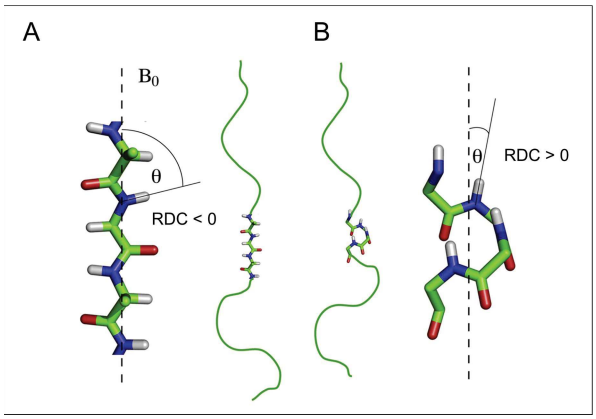


Figure 43 : Relation entre les RDCs et l'orientation du vecteur $\overrightarrow{NH}$ dans les structures secondaires transitoires des IDPs. (Figure modifiée à partir de (Jensen *et al*, 2009))**. A)** Dans une conformation étendue, comme les brins β, les vecteurs $\overrightarrow{NH}$ (bleu-gris) sont généralement allignés plus ou moins perpendiculairement à $\overrightarrow{Bo}$ (angle θ ≈ 90 °C) donnant une valeur négative de RDC. **B)** À l'inverse, dans une hélice α les vecteurs $\overrightarrow{NH}$ sont plus ou moins parallèles à $\overrightarrow{Bo}$ (angle θ ≈ 0) donnant une valeur positive de RDC.

3.3.4.3.2 Les milieux orientants

Les milieux permettent de créer un environnement préférentiellement anisotrope en alignant les molécules grace à des interactions stériques ou électrostatiques.

Nous avons utilisé deux milieux orientants : un milieu composé d'un mélange d'alkyl-poly(éthylène glycol) (PEG/$C_{12}E_5$) et d'hexanol à ratio 0.85 et un milieu composé de bactériophages filamenteux (Pf1) (Figure 44).

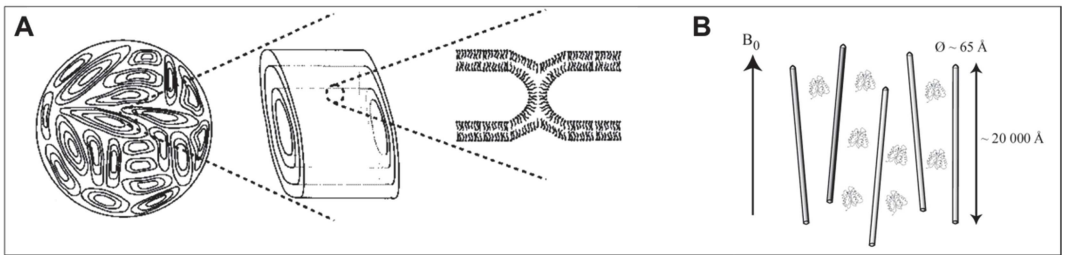


Figure 44 : Milieux orientants $C_{12}E_5$/hexanol et phage Pf1 par rapport au champ magnétique B_0. A) Les chaînes carbonées du PEG se regroupent pour former des bicouches planes qui forment une superstructure lamellaire (Figure extraite de (Gaemers & Bax, 2001)). L'espacement entre chaque bicouche dépend du ratio PEG:hexanol. **B)** Les phages Pf1 s'alignent en longueur le long du B_0 grâce à leur susceptibilité magnétique (Figure extraite de la thèse de Nathalie Sibille).

En milieu $C_{12}E_5$/hexanol, les protéines n'ont généralement pas tendance à interagir avec la protéine. Rückert et Otting notent que la surface des bicouches qui forme une phase lamellaire n'est pas chargée et doit donc provoquer un alignement

stérique (Figure 44, (Rückert & Otting, 2000)). À l'inverse, les phages sont chargés négativement en surface ce qui peut induire des interactions électrostatiques avec les protéines. Ces charges disposéees le long du phage filamenteux leur permettent de s'aligner le long du champ magnétique B_0 (Bax *et al*, 2001) (Figure 44).

L'alignement et la stabilisation dans le temps du milieu orientant sont suivis par la mesure du couplage quadrupolaire qui induit un dédoublement de la raie du deutérium (ΔD) de l'eau lourde (D_2O) (Figure 45).

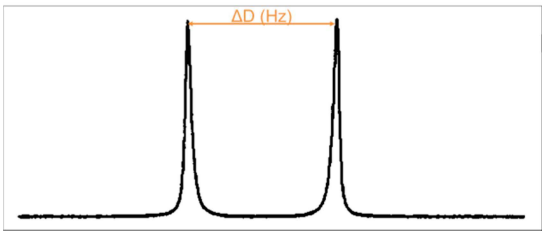


Figure 45 : Observation du dédoublement quadrupolaire de la raie du deutérium (ΔD) sur un spectre RMN ^{2}H enregistré en milieu $C_{12}E_5$/hexanol (Figure extraite de (Rückert & Otting, 2000)).

Afin de vérifier si des interactions non désirées entre le milieu orientant et la protéine surviennent, nous suivons les variations de déplacements chimiques de l'échantillon dilué en milieu isotrope et anisotrope avec des spectres ^{15}N-HSQCs. Il est important qu'aucune interaction spécifique n'apparaisse afin de pouvoir mesurer des RDCs de façon précise et non biaisée.

3.3.4.3.3 Expérimentation

En pratique, nous avons utilisé la séquence IPAP de (Ottiger *et al*, 1998) pour l'enregistrement des expériences qui permettent d'extraire les RDCs. Cette séquence permet d'enregistrer deux expériences séparées qui contiennent respectivement les deux composantes du doublet : un spectre avec une composante en phase (IP) et l'autre en anti-phase (AP). Ces expériences sont additionnées ou soustraites afin d'obtenir les deux composantes du doublet dans deux spectres séparés. La valeur de couplage est ensuite déterminée par soustraction de la position des pics (δ) du doublet entre les spectres 1 = IP + AP et 2 = IP – AP en dimension ^{15}N (équation 8) :

$$Couplage = |\delta_1 - \delta_2| \qquad (8)$$

En milieu isotrope, le couplage correspond uniquement au couplage scalaire (J). Alors qu'en milieu anisotrope, le couplage correspond au couplage scalaire auquel s'ajoute la composante anisotrope du couplage qui est le couplage dipolaire (J+D). Par conséquent, une simple soustraction des couplages obtenus dans chaque condition permet l'obtention des couplages dipolaires résiduels (D).

3.3.4.4 Relaxation de la liaison amide ^{1}H-^{15}N

Les mouvements locaux des protéines dépendent de leur structuration. Des régions possédant des conformations même transitoires seront plus rigides et moins dynamiques que des régions complètement 'random-coil'. Ainsi, la dynamique peut nous renseigner sur la présence de structures secondaires dans les IDPs. Les mouvements des protéines, compris dans l'échelle de temps de la picoseconde (ps) à la milliseconde (ms), peuvent être détectés par leur effet sur les paramètres de relaxation R_1 et R_2 et sur les NOE hétéronucléaires (hetNOE) de la liaison amide ^{1}H-^{15}N du squelette protéique (Kleckner & Foster, 2011). En traçant ces paramètres en fonction de la séquence en acides aminés, il est possible d'identifier des régions possédant des mouvements spécifiques dont les valeurs seront éloignées de la moyenne. Lors du traitement des données, nous avons supprimé les trois résidus N- et C-terminaux du calcul des moyennes étant donné leur flexibilité intrinsèque.

3.3.4.4.1 Le phénomène de relaxation

Le phénomène de relaxation correspond au retour à l'équilibre des spins excités par une impulsion radiofréquence B_1 (chapitre 3, p. 106, Figure 40). Le temps de relaxation T_1 ou spin-réseau est le temps nécessaire pour le retour à l'équilibre thermodynamique de ces spins le long de l'axe z. Le temps de relaxation T_2 ou spin-spin traduit la perte de cohérence des spins responsables de la disparition de l'aimantation transversale dans le plan des axes x et y (Figure 46). Plus généralement, on mesure leurs constantes de vitesses respectives ($R_1=1/T_1$ et $R_2=1/T_2$).

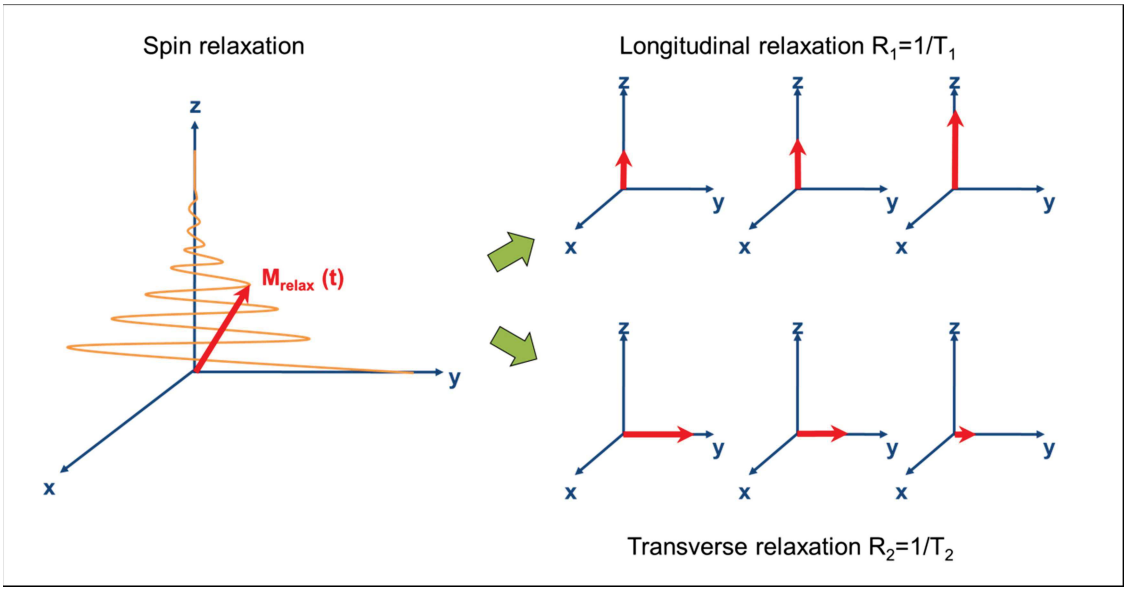


Figure 46 : Phénomène de relaxation des spins. Dans le temps T_1, les spins reviennent à l'équilibre le long de l'axe z et dans le temps T_2 l'aimantation disparaît du plan x/y. Ces deux temps sont souvent décrits via leurs constantes de vitesses respectives R_1 et R_2.

Le phénomène de relaxation est régi par deux interactions principales qui fluctuent dans le temps selon le mouvement des molécules : l'anisotropie de déplacement chimique (CSA) et le couplage dipolaire (D). L'anisotropie de déplacement chimique est due au fait que l'environnement électronique (tenseur) entourant un noyau varie en fonction de l'orientation de la molécule par rapport au champ magnétique B_0. Une fluctuation dans le temps de ce tenseur va influer le phénomène de relaxation. Le couplage dipolaire dépend de la distance entre les noyaux ^{1}H et ^{15}N de la liaison amide. Les fluctuations de cette distance internucléaire et la réorientation du vecteur internucléaire due aux mouvements moléculaires vont avoir pour conséquence de faire varier le champ magnétique local et donc la relaxation du noyau ^{15}N.

3.3.4.4.2 Expérimentation

Les paramètres de relaxation R_1 et R_2 ont été mesurés par acquisition d'une série de spectres ^{15}N-HSQCs avec différents délais de relaxation (t) : 10, 50, 100, 200, 400, 600, 800, 1 000 ms pour R1, et 16, 32, 64, 96, 160, 240, 480, 640 ms pour R2. Ensuite, les intensités de chaque pic sont utilisées pour déterminer les paramètres de relaxation R_1 et R_2 selon l'équation (9) (Kleckner & Foster, 2011) :

$$I(t) = I(0).(1 - \exp(-R_{1\,ou\,2} * t)) \qquad (9)$$

où I(t) et I(0) sont les intensités à un délai de relaxation (t) et au temps t=0.

Cette étape a été faite dans le programme Sparky (Lee *et al*, 2015) qui trace les intensités en fonction du temps de relaxation et ajuste les courbes par décroissance exponentielle pour obtenir les vitesses de relaxation T_1 et T_2.

3.3.4.4.3 Effet Oerhauser hétéroNucléaire (hetNOE)

Les NOE hétéronucléaires ^{1}H-^{15}N (ou $^{15}N\{^{1}H\}$-NOE) résultent du phénomène de relaxation croisée, qui témoigne du couplage entre deux aimantations. La vitesse de relaxation croisée tire son origine de l'interaction dipolaire entre les noyaux ^{1}H et ^{15}N, modulée par les mouvements moléculaires. Lorsque les protons $^{1}H_N$ sont saturés, on observe un transfert de polarisation du proton vers l'azote-15 en fonction de la vitesse de la relaxation croisée. Cette efficacité de transfert peut être mesurée par la différence de l'intensité des pics de deux expériences décrites ci-après.

3.3.4.4.4 Expérimentation

Les hetNOEs ont été mesurés expérimentalement en enregistrant deux spectres 2D de type (^{1}H-^{15}N)-HSQC avec et sans saturation de 5 s du noyau ^{1}H.

Les intensités dans chaque spectre ont été mesurées et les NOE ont été calculées avec l'équation suivante (10) :

$$hetNOE = \frac{I_{sat}}{I_{ref}} \qquad (10)$$

où I_{sat} et I_{ref} sont, repsectivement, les intensités des pics dans le spectre de saturation du noyau ^{1}H, et dans le spectre de référence (sans saturation proton mais où une saturation hors résonance a été effectuée pour compenser d'éventuels effets d'échauffement).

3.3.4.5 Augmentation de la Relaxation Paramagnétique (PRE)

3.3.4.5.1 Principe

L'utilisation des PREs (augmentation de la relaxation paramagnétique) en RMN permet l'obtention d'informations sur les interactions à longue distance entre les résidus d'une protéine (Clore, 2015). Pour cela, la protéine est marquée sur une cystéine avec une sonde paramagnétique, molécule avec un électron non-apparié, qui peut interagir avec les atomes situés jusqu'à ~ 20 Å de distance (Teilum *et al*, 2002). La présence de cette sonde va entraîner l'apparition d'interactions dipolaires entre l'électron non-apparié et les spins qui l'entourent. Ces interactions vont

augmenter la vitesse de relaxation des spins et donc diminuer l'intensité des signaux RMN. Par conséquent, une manière simple de mesurer les PREs est de comparer les intensités entre deux spectres ^{15}N-HSQCs avec et sans l'effet paramagnétique (spectre dit paramagnétique et diamagnétique).

3.3.4.5.2 Expérimentation

Les PREs ont été mesurés sur des échantillons de 50 µM et de 100 µM pour les formes sauvage et phosphomimétique de β2AR-Cter, respectivement. Les peptides de β2AR mutés sur la cystéine C378 (C378A) ou C406 (C406A) en alanine sont marqués avec une sonde 3-(2-Iodoacetamodi)-proxyl (Merck) sur l'unique cystéine restante, C406 ou C378 (Figure 47). Le marquage est effectué par ajout de 1 mM de proxyl pour 100 µM de β2AR en tampon 50 mM BT pH 6.7, 150 mM NaCl, 0.5 mM EDTA et 1 mM TCEP. Après 4 h d'incubation à 20 °C sans lumière, la réaction est stoppée par ajout de 5 mM de DTT. L'excès de sonde est enlevé par passage sur une PD MiniTrapTM G-10 (GE Healthcare). Après marquage, le spectre paramagnétique est enregistré puis 5 mM de vitamine C est ajouté dans le tube RMN (volume négligeable de ~ 2 µL) pour enregistrer le spectre diamagnétique. La vitamine C est capable de donner un électron au proxyl, supprimant son effet paramagnétique (Figure 47). Les spectres sont enregistrés avec un délai de répétition d1 de 2 s pour s'assurer que le retour de l'aimantation à l'équilibre sur z soit le même dans les échantillons paramagnétiques et diamagnétiques.

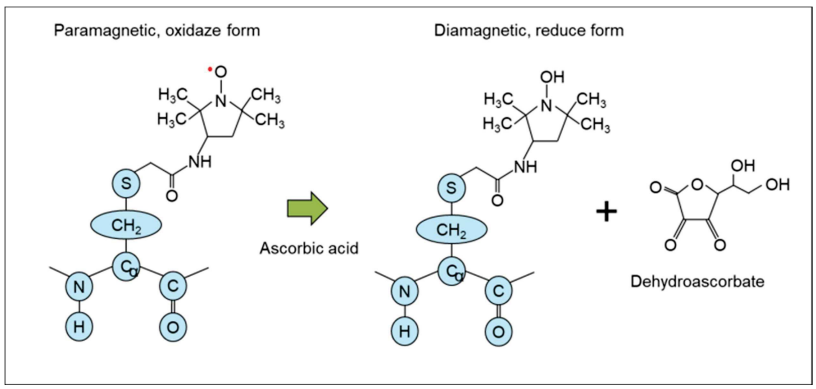


Figure 47 : Obtention des formes paramagnétiques et diamagnétiques. La cystéine C378 (en bleu) porte une sonde paramagnétique (proxyl) dont un électron n'est pas apparié (en rouge). Après ajout de vitamine C (acide ascorbique), la sonde est réduite donnant la forme diamagnétique et du dehydroascorbate.

Les PREs sont ensuite extraits par ratio des intensités entre les spectres paramagnétiques (I_{para}) et diamagnétiques (I_{dia}) (équation 11) :

$$PRE = \frac{I_{para}}{I_{dia}} \qquad (11)$$

Pour les profiles théoriques des PREs, Poulsen et son équipe ont créé une équation qui permet de déterminer la contribution qu'aurait la sonde paramagnétique sur la relaxation R_2 d'une protéine complètement déstructurée, R_{2P} (équation 12) (Teilum *et al*, 2002) :

$$R_{2P} = \frac{K}{r^6}(4\tau_c + \frac{3\tau_c}{1+\omega_H^2\tau_c^2} \qquad (12)$$

où K est une constante ($1.23.10^{-32}$ cm^6s^{-2}), r est la distance entre les noyaux N-H (3.1 Å), τ_c est le temps de corrélation d'une interaction dipole-dipole electron-noyau (4.0 ns) et ω_H est la fréquence de résonance du proton (700 MHz).

Grâce à ce paramètre R_{2P}, il est ensuite possible de déterminer les PREs qu'aurait la protéine désordonnée avec l'équation (13) (Teilum *et al*, 2002) :

$$\frac{I_{para}}{I_{dia}} = \frac{R_2 \,.\, \exp(-R_{2P}.t)}{R_2 + R_{2P}} \qquad (13)$$

où R_2 est la constante de vitesse de relaxation transversale moyenne (3.5 s^{-1} pour β2AR-Cter, le paramètre R_{2P} a été calculé pour chaque résidu selon l'équation (13) (Teilum *et al*, 2002) et t est un paramètre d'acquisition RMN (10 ms, il représente le temps où la polarisation du proton reste dans le plan transversale lors de la séquence INEPT (Renforcement de transfert de polarisation entre deux noyaux).

3.3.5 Étude de l'interaction RCPG-Cter:arrestine-2 par CSP

Comme présenté précédemment (Chapitre 3, p. 100), nous avons étudié les interactions RCPG-Cter:arrestine avec différentes méthodes, dont la RMN. Par RMN nous avons suivi l'interaction par perturbation du déplacement chimique (CSP) et de l'intensité. Les interactions des domaines C-terminaux des RCPGs avec l'arrestine sont de nature plutot electrostatique donc elles peuvent être inhibées par des concentrations élevées en sel (Gurevich *et al*, 1995). Par conséquent, les expériences d'interaction ont été effectuées à 50 mM en sel et enregistrées sur le spectromètre de 800 MHz.

3.3.5.1.1 Expériences de CSP (perturbation du déplacement chimique) ou de perturbation d'intensité

Lors d'une interaction avec une protéine partenaire, l'environnement des spins va être modifié. Cela se traduit par une modification du déplacement chimique. De plus, la formation du complexe protéique entraîne des changements d'intensités des pics. Ceci est associé à une augmentation du temps de corrélation de la protéine compléxée, qui augmente la vitesse de relaxation R_2, et entraine un élargissement des raies de résonnance des résidus en interaction (baisse d'intensité).

La variation de déplacement chimique et d'intensité peut être obtenue sur des spectres ^{15}N-HSQCs en absence et en présence du partenaire. La mesure de leur variation est effectuée avec les équations (14) et (15) (Grzesiek *et al*, 1996; Williamson, 2013) :

$$\Delta\delta = \sqrt{\frac{1}{2}\left[(\Delta_{HN})^2 + (\frac{\Delta_N}{5})^2\right]} \qquad (14)$$

$$\Delta I = \frac{I_{complexe}}{I_{libre}} \qquad (15)$$

où Δδ est la combinaison de la variation des déplacements chimiques ^{1}H (Δ_{HN}) et ^{15}N (ΔN) du groupe amide avec et sans partenaire (équation 14) et où ΔI est le ratio des intensités extraites des HSQCs sans (I_{libre}) et avec ($I_{complexe}$) le partenaire (équation 15).

3.3.5.1.2 Expérimentation

Les interactions entre les peptides des RCPGs (à 10 µM) et l'arrestine-2 ont été analysées avec des spectres ^{15}N-HSQC à différents ratios (RCPG:arrestine) = 1:1, 1:2, 1:5 et 1:10. Les variations de déplacements chimiques et d'intensités calculées à l'aide des équations (14) et (15) ont été tracées en fonction de la séquence en acides aminés pour identifier les régions impliquées dans la liaison à l'arrestine. La déviation de déplacement chimique n'a pas été calculée sur les résidus superposés suivants : 346/367 et 363/364 pour wt-V2R; 345/363 et 350/362 pour pm-GHSR; 345/364, 346/356, 347/381, 352/359/398, 370/372, 367 et 391 pour wt-β2AR; et 343/372, 347/381, 358/361, 364/371/391, 370/373, 401/410 pour pm-β2AR. La valeur seuil d'interaction a été calculée d'une part avec l'écart-type du déplacement chimique multiplié par 4 pour les CSP et d'autre part avec la moyenne des intensités soustraite

à l'incertitude pour les ΔI (Williamson, 2013). Les résidus en interaction ont été exclus des calculs de seuils.

3.3.6 Étude de la phosphorylation des RCPG-Cters

Les régions désordonnées C-terminales des RCPGs sont souvent phosphorylées sur plusieurs sites par les GRKs. Dans ce contexte, la RMN permet d'obtenir des informations quantitatives et qualitatives pour chaque acide aminé comme la localisation des phospho-résidus, le taux de phosphorylation et leur cinétique de phosphorylation (Landrieu *et al*, 2006). Ce paragraphe est développé dans le Chapitre 4, partie 3, p. 203). Ainsi, uniquement les protocoles utilisés pour l'étude des RCPGs-Cters phosphorylés sont présentés ci-dessous.

3.3.6.1 Réaction de phosphorylation

Dans le cas des domaines C-terminaux des RCPGs, la phosphorylation par GRK2 est effectuée *in vitro* (résultat au Chapitre 4, p. 200). Le suivi de phosphorylation directement dans le tube RMN n'a pas pu être effectué car la kinase ne fonctionne pas au pH plus bas nécessaire pour l'enregistrement des spectres RMN, en effet à pH 7.5 l'échange eau:proton amide est trop important pour les IDPs. Avant d'être incubés avec la kinase, les domaines C-terminaux des RCPGs sont dialysés ON à 4 °C en tampon de phosphorylation (50 mM Tris pH 7.5, 150 mM NaCl, 2 mM DTT). Le lendemain, la phosphorylation est lancée par ajout de 40 µM de GRK2 pour 10 µM de peptide (ratio 4:1 en kinase:RCPG), en présence de 10 mM de $MgCl_2$ et 10 mM d'ATP (en excès de 250 par rapport à la kinase). Le mélange est mis à incuber 9 h à 30 °C sous agitation. La phosphorylation est stoppée par chauffage de la solution à 75 °C pendant 10 min, puis l'échantillon est centrifugé 15 min à 12 000 g et 20 °C. Pour finir, la protéine phosphorylée est dialysée ON à 4°C en tampon RMN (50 mM Bis-Tris pH 6.7, 150 ou 50 mM NaCl, 1 mM TCEP et 0.5 mM EDTA).

3.3.6.2 Optimisation des conditions de phosphorylation

Le protocole présenté précédemment est le protocole final, obtenu après de nombreuses mises au point. Dans ce paragraphe, seuls les éléments qui varient par rapport à ce protocole seront précisés.

Pour la mise au point de la phosphorylation des peptides des RCPGs par GRK2, nous avons testé l'effet du glycérol et de la concentration en ATP sur l'activité de GRK2. Pour cela la réaction de phosphorylation a été effectuée à 10 µM ou 40 µM de GRK2, en absence ou en présence de 25 % de glycérol (10 µM β2AR:10 µM GRK2), ou par utilisation de 10 mM ou 40 mM d'ATP (10 µM β2AR:40 µM GRK2). Les spectres ^{15}N-HSQCs ont été enregistrés sur le spectromètre de 700 MHz pour le test du glycérol et sur le spectromètre de 800 MHz pour le test de l'ATP. Le taux de phosphorylation dans les différentes conditions a été suivi par mesure de l'intensité normalisée (I_{norm}) des pics phosphorylés ($I_{phospho}$) par rapport à l'intensité d'un pic non perturbé (I_{ref}) (L413 pour β2AR, M337 pour GHSR et S371 pour V2R) (équation 16 ; (Hendus-Altenburger *et al*, 2016) :

$$I_{norm} = I_{phosp} \quad * \frac{100}{I_{ref}} \qquad (16)$$

3.3.6.3 Identification des sites phosphorylés

L'identification des sites phosphorylés a été partiellement faite par cinétique de phosphorylation. Pour GHSR-Cter phosphorylé par GRK2, l'attribution a été confirmée par acquisition du spectre RMN 3D HNCACB (800 MHz).

La cinétique de phosphorylation a été faite à la paillasse en utilisant 10 µM (ratio 1:1) ou 40 µM (ratio 1:4) de GRK2 pour 10 µM de β2AR ou 10 µM de GHSR, respectivement. La réaction est stoppée t = 0 min, 15 min, 30 min, 1 h, 2 h, 4 h ou 16 h et les spectres ^{15}N-HSQCs des solutions de réaction sont enregistrés sur le spectromètre à 800 MHz. La cinétique est suivie par mesure de la différence de déplacement chimique (équation 15, p.111 et par suivi des intensités des pics. Les intensités sont normalisées par rapport à des résidus non perturbés : L413 pour β2AR-Cter, M337 pour GHSR-Cter et S371 pour V2R-Cter selon l'équation 17. Les intensités normalisées sont tracées en fonction du temps et les constantes apparentes de vitesse (K_{app}) sont extraites par régression non-linéaire, grâce à la méthode des moindres carrés, en utilisant GraphPad Prism.

3.3.7 Analyse statistique

En biologie, on ne connait généralement pas les valeurs réelles, dites « vraies » des grandeurs mesurées. On parle donc d'incertitudes. Statistiquement, le terme « incertitude (de mesure) » est le paramètre qui caractérise la dispersion des valeurs qui pourraient raisonnablement être associé aux grandeurs mesurées (selon la norme ISO : JCGM 100:2008). L'incertitude est notée (SD) mais elle ne correspond pas à un écart à la moyenne car les expériences n'ont été effectuée qu'une seule fois.

Dans ce manuscrit, les incertitudes des mesures RMN (SD) ont été calculées selon la loi de propagation des erreurs (Ku, 1966).

3.3.7.1 Incertitude sur les intensités

L'incertitude sur les ratios d'intensité ($SD_{I1/I2}$) a été calculée par l'équation (17) :

$$SD_{I1/I2} = \frac{I_1}{I_2} \times \sqrt{\left[\left(\frac{1}{SN_{I1}}\right)^2 + \left(\frac{1}{SN_{I2}}\right)^2\right]} \qquad (17)$$

Avec I_1 et I_2 les intensités des pics souhaités et SN_{I1} et SN_{I2} le ratio signal sur bruit calculé par Sparky pour les intensités I_1 et I_2 respectives.

3.3.7.2 Incertitude sur les déplacements chimiques

L'incertitude sur les déplacements chimiques (SD_δ) a été calculée par l'équation (18) (Bax *et al*, 2001) :

$$SD_\delta = \frac{LW}{SN} \qquad (18)$$

Où LW est la largeur de raie et SN le signal sur bruit du pic.

Par la loi de propagation de l'erreur, l'incertitude sur la variation de déplacement chimique entre deux conditions 1 et 2 devient (équation 19) :

$$SD_{\Delta\delta} = \frac{\sqrt{\frac{[625(\Delta_H)^2.(SD_{\delta H})^2 + 625(\Delta_H)^2.(SD_{\delta H2})^2] + [\Delta_N{}^2.((SD_{\delta N1})^2 + (SD_{\delta N2})^2)]}{25\delta_{H1}^2 - 50\delta_{H1}\delta_{H2} + 25\delta_{H2}^2 + (\delta_{N1} - \delta_{N2})^2}}}{5\sqrt{2}} \qquad (19)$$

Où Δ_H et Δ_N sont les variations de déplacement chimique ^1H et ^{15}N, δ_{H1} et δ_{H2} sont les déplacements chimiques ^1H du résidu i dans les conditions 1 et 2 souhaitées et $SD_{\delta H1}$, $SD_{\delta H2}$ leurs incertitudes respectives. À l'inverse, δ_{N1} et δ_{N2} sont les

déplacements chimiques ^{15}N du résidu i dans les conditions 1 et 2 souhaitées et $SD_{\delta N1}$, $SD_{\delta N2}$ leurs incertitudes respectives.

Par la loi de propagation de l'erreur, l'incertitude sur les RDCs (SD_{RDC}) devient (équation 20) :

$$SD_{RDC} = \sqrt{SD_{J1}^{2} + SD_{J2}^{2} + SD_{(J+D)1}^{2} + SD_{(J+D)2}^{2}} \qquad (20)$$

où SD_{J1}, SD_{J2}, $SD_{(J+D)1}$ et $SD_{(J+D)2}$ sont les incertitudes de déplacement chimique d'un résidu i dans les spectres IPAP isotropes J1 et J2 et IPAP anisotropes (J+D)1 et (J+D)2, respectivement.

4 RÉSULTATS ET DISCUSSION

4.1 ARTICLE 1 : PROBLEMATIQUE ET RESUME

Caractérisation structurale et dynamique des domaines C-terminaux des RCPGs non-phosphorylés

Le domaine C-terminal des RCPGs a un rôle crucial dans la signalisation dépendante de l'arrestine. Après sa phosphorylation par les GRKs, il permet le recrutement de l'arrestine au récepteur (Seyedabadi *et al*, 2021). La formation du complexe RCPG:arrestine mène à l'activation de l'arrestine via des changements conformationnels. Cependant, les potentiels changements structuraux que subit le domaine C-terminal des RCPGs, après la phosphorylation par les GRKs ou par liaison à l'arrestine, ne sont toujours pas décrits. En effet, ces régions sont difficilement cristallisables ou visualisables dans les cartes de densité électronique. Ces domaines sont très flexibles, prédits pour être désordonnés et pour se comporter comme une région intrinsèquement désordonnée, régions qui contiennent peu de structures secondaires (IDRs) (Jaakola *et al*, 2005; Venkatakrishnan *et al*, 2014).

Dans l'objectif de mieux comprendre les potentiels changements que subit le domaine C-terminal des RCPGs dans la signalisation dépendante de l'arrestine, nous avons caractérisé structuralement la conformation des régions C-terminales de trois RCPGs modèles (RCPG-Cter) : le récepteur à la vasopressine V2 (V2R), le récepteur à la ghréline (GHSR) et le récepteur β2-adrénergique (β2AR). La conformation des domaines C-terminaux cytosoliques de ces récepteurs a été étudiée en utilisant principalement la Résonance Magnétique Nucléaire (RMN) couplée à des analyses de dichroïsme circulaire (CD), de diffusion de la lumière aux petits angles (MALS) et de diffusion des rayons X aux petits angles (SAXS). La caractérisation globale (CD, MALS et SAXS) a permis de confirmer la nature désordonnée de ces régions. Ensuite, l'analyse des données RMN a permis de mettre en évidence l'existence de courtes régions contenant des structures secondaires transitoires, à l'intérieur des domaines C-terminaux désordonnés des RCPGs. Nous avons montré que le domaine C-terminal de V2R-Cter contient une région repliée en hélice, alors que cette même région forme un brin bêta dans les complexes V2R:arrestine décrit dans la littérature (Zhou *et al*, 2017; Nguyen *et al*,

2019; Staus *et al*, 2020). Ainsi, nos résultats suggèrent qu'une transition structurale a lieu, après phosphorylation et/ou interaction avec l'arrestine. Si la phosphorylation est à l'origine de la transition structurale, on peut s'attendre à ce qu'elle régule la signalisation des RCPGs en modulant le repliement de leur domaine C-terminal. Cette hypothèse permettrait d'expliquer le mécanisme moléculaire impliqué dans le phospho-barcode modèle.

En résumé :

Les domaines C-terminaux désordonnés des RCPGs contiennent des structures secondaires transitoires qui peuvent être impliqué dans l'interaction à l'arrestine.

V2R-Cter contient une région repliée en hélice alors que cette région forme un brin bêta dans les complexes V2R:arrestine, suggérant une transition structurale.

Cette transition structurale pourrait être régulée par la phosphorylation qui, ainsi, dicterait la signalisation des RCPGs, ce qui expliquerait le mécanisme moléculaire impliqué dans le phospho-barcode modèle.

Dans cet article, j'ai contribué à la production des protéines et à leur caractérisation globale (CD, MALS, SAXS). J'ai réalisé et analysé les expériences RMN pour l'étude du domaine C-terminal du récepteur β2-adrénergique. J'ai également complété la caractérisation RMN des domaines C-terminaux des récepteurs V2R et GHSR, commencé avant mon arrivée.

La partie 'matériel et méthode' de l'article reprend les protocoles présentés précédement dans le chapitre 3, matériel et méthode, il n'est donc pas nécessaire de le lire.

Structural insights into the intrinsically disordered GPCR C-terminal region, major actor in arrestin-GPCR interaction

Myriam Guillien[a], Assia Mouhand[a], Aurélie Fournet[a], Amandine Gontier[a,b], Aurélien Thureau[c], Frédéric Allemand[a], Jean-Louis Banères[d], Pau Bernadó[a], Nathalie Sibille[a,*,1]

[a] Centre de Biochimie Structurale (CBS), CNRS, INSERM, Univ Montpellier, Montpellier, France

[b] Sorbonne University, École Normale Supérieure, PSL University, CNRS, Laboratoire des biomolécules (LBM), Paris, France

[c] Beamline SWING, Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91190 Gif-sur-Yvette, France

[d] IBMM, UMR5247 CNRS, Pôle Chimie Balard Recherche, 1919 route de Mende, 34293 Montpellier cedex 5, France

* Corresponding author. Email addresses: nathalie.sibille@cbs.cnrs.fr

[1] Lead contact. Tel: +33 4 67 41 77 05; Fax: +33 4 67 41 79 13; Email: nathalie.sibille@cbs.cnrs.fr

ABSTRACT

G protein-coupled receptors (GPCRs) are the largest family of membrane proteins implicated in signal transduction. Their signaling pathways involve two main cytosolic partners: G proteins and arrestins. Nowadays, the understanding of the second signaling pathway, which is regulated by arrestins, is expanding rapidly but structural data are still scarce. One of the remaining questions concerns the structural impact of GPCR C-terminal disordered regions on arrestin binding. For this purpose, we describe here the basal conformations of the C-temrinal domain of three class A GPCRs: the vasopressin V2 Receptor (V2R), the Growth Hormone Secretagogue or ghrelin Receptor type 1a (GHSR) and the β2-Adernergic Receptor (β2AR). We revealed the presence of transient secondary structures that are potentially involved in the interaction with arrestin. These secondary structure elements differ from those described in the literature in interaction with arrestin. This suggests a mechanism where the secondary structure conformational preferences in the C-terminal regions of GPCRs are modulated by post-translational modifications and/or upon interaction with arrestin in order to optimize its recognition.

Keywords

GPCR; arrestin; Intrinsically Disordered Proteins or Regions (IDPs/IDRs); NMR

INTRODUCTION

G Protein-Coupled Receptors (GPCRs) are integral membrane proteins involved in signal transduction. They are central in the cellular response for a wide range of extracellular ligands, such as hormones, nucleotides, lipids, ions, photons and neurotransmitters (Bockaert & Philippe Pin, 1999). Their signaling outcome regulate a large number of biological functions and therefore their dysfunctions are linked to various pathologies (Schöneberg *et al*, 2004). Consequently, GPCRs are

the target of a third of the current clinical drugs (Hauser *et al*, 2017). Their functional diversity comes from the existence of a large number of GPCRs (~800 human members), classified according to their sequence and phylogenetic analyzes (Fredriksson, 2003). While they share a highly conserved core domain (7TM) composed of seven transmembrane helices connected by three extracellular (ECL) and intracellular (ICL) loops, their extracellular N- and intracellular C-terminus, as well as the loops, are highly variable in both sequence and length (Figure1A) (Lagerström & Schiöth, 2008).
Upon extracellular ligand binding, GPCR conformational rearrangements allow G protein association and G protein dependent signaling initiation. Besides G proteins, GPCRs can trigger other signaling pathways by interaction with arrestin, such as desensitization, internalization and receptor trafficking (Gurevich & Gurevich, 2019a). Arrestin interaction requires GPCR C-terminal domain (GPCR-Cter) phosphorylation by specific G protein-coupled Receptor Kinases (GRKs). The current model of GPCR:arrestin interaction, the so called phospho-barecode model (Ren *et al*, 2005; Nobles *et al*, 2011), states that, upon activation, GPCRs get phosphorylated at specific sites on their C-terminal domain. This impact their interaction with arrestin, in turn affecting its conformation and, ultimately, the intracellular signal triggered by arrestin. However, we are just beginning to grasp the mechanistic bases of this mode of functioning as experimental demonstrations are still scarce even for the best-studied GPCRs.

GPCR:arrestin complex formation leads to arrestin activation through conformational changes (for a review see (Gurevich & Gurevich, 2019b)). However, putative conformational changes of GPCR-Cter induced by GRK phosphorylation or arrestin binding are still not described. These C-terminal regions are predicted to behave as intrinsically disordered regions (IDRs) (Venkatakrishnan *et al*, 2014; Jaakola *et al*, 2005). Intrinsically disordered proteins (IDPs) or IDRs are highly dynamic proteins/regions containing a low content of transient secondary structures, which make their structural characterization difficult (Uversky, 2002). Atomic structures of GPCR-arrestin complex revealed that the C-terminus of vasopressin 2 (V2R) and rhodospin receptors are partially folded on arrestin surface (Kang *et al*, 2015; Zhou *et al*, 2017; Nguyen *et al*, 2019; Staus *et al*, 2020). This suggests an usual mechanism for IDPs by which an Intrinsic Disordered Region (IDR) undergoes conformational changes upon post-translational modification and/or binding to their targets (Wright & Dyson, 2015). Indeed, IDPs/IDRs contain most likely some pre-formed secondary structure elements that are often involved in the recognition of specific partners (Fuxreiter et al, 2004).

In order to better understand the putative conformational changes of the GPCR C-terminal domain associated with GRK phosphorylation and/or arrestin binding, we characterized the structure of the truncated C-terminal regions of three GPCRs by Nuclear Magnetic Resonance (NMR): the vasopressin V2 Receptor (V2R) (from residue 343 to 371), the Growth Hormone Secretagogue or ghrelin Receptor type 1a (GHSR) (from residue 339 to 366) and the β2-Adernergic Receptor (β2AR) (from residue 342 to 413) (Figure 1A). These three class-A receptors are important therapeutic targets (Schöneberg *et al*, 2004) and they are representative of the two different classes of arrestin binders (Oakley *et al*, 2000). The first class, which includes β2AR and GHSR, forms transient complexe with arrestin which dissociates near the plasma membrane. Thus, arrestin does not internalise with the

receptor. In contrast, the second class, wich include V2R, forms more stable complexe allowing the internalisation of the whole complex.

Here, we confirmed the disordered nature of these three GPCR-Cters and the presence of transient secondary structures by a set a biophysical tools: Circular Dichroïsm (CD), Multi Angle Light Scattering (MALS), and Small Angle X-ray Scattering (SAXS). Then, we used NMR to probe the conformational and dynamic preferences at residue level using a set of complementary experiments, such as chemical shift (CSI), scalar (J) and Residual Dipolar Couplings (RDCs), Paramagnetic Relaxation Enhancement (PRE) and relaxation (Senicourt *et al*, 2020; Sibille & Bernadó, 2012; Delaforge *et al*, 2018). We show that the three C-terminal regions of the chosen GPCRs revealed different transient secondary structures which could be involved in arrestin binding. These regions could be short linear motifs (SLiMs), which is the partner recognition of IDPs that are embedded in disordered poorly conserved regions. By comparison with crystallographic structure of synthetic peptides in interaction with arrestin (Shukla *et al*, 2013; Staus *et al*, 2020; Nguyen *et al*, 2019), our results suggest that conformational changes occur either after phosphorylation of the C-terminal region or upon arrestin binding.

RESULTS

The disordered C-termini of V2R, GHSR and β2AR contain transient secondary structures.

As for many C-terminal domains of GPCRs (Tovo-Rodrigues *et al*, 2014), the C-terminal regions of V2R, GHSR and β2AR are predicted to be disordered by various computational tools. In fact, they are composed of more than 50% of disordered-promoting residues (Ala, Arg, Gly, Gln, Ser, Pro, Glu and Lys) (Figure 2A) (Dunker *et al*, 2001). In the Charge-Hydropathy plot (Uversky *et al*, 2000) (Figure S1), V2R-Cter is at the disordered-order boundary, while β2AR-Cter (pink) and GHSR-Cter (green) appeared in the cluster of disordered proteins (yellow). Sequence based disorder prediction by a set of six predictors (Table S1) showed an overall disordered state (values higher than 0.5) for the three C-termini (Figure 2B). More specifically, prediction by PrDOS, DISOPRED3 and Espritz-NMR showed less disordered regions from residues ~ 351 to 365 for V2R-Cter and from residues ~ 343 to 359 for GHSR-Cter (Figure 2B). β2AR-Cter is also predicted to be less disordered from ~ 345 to 356 and from ~ 369 to 405 according to PrDOS, DISOPRED3 and Espritz-NMR predictors (Figure 2B). Interestingly, these regions are predicted to contain helical conformations by a set of distinct secondary structures predictors (Figure 2C, Table S2).

In order to characterize these domains experimentally, we expressed and purified the intrinsically disordered and soluble C-terminal regions of V2R (343-371), GHSR (339-355) and β2AR (342-413). SEC-MALS analysis revealed a single elution pic for each protein at a volume that corresponds to the volume of standard proteins with a molecular mass greater than 13 kDa . However, the masses derived from MALS analysis are 3.9 (± 3.2%), 3.4 (± 3.7%) and 8.6 (± 1.1%) kDa for V2R-Cter, GHSR-Cter and β2AR-Cter, respectively (Figure 1B). This is in agreement with the expected molecular mass of their monomeric forms (3.3, 3.6 and 8.2 kDa), which were confirmed by mass spectrometry (MS) with 3,264, 3,672 and 8,176 kDa obtained for V2R-Cter, GHSR-Cter and β2AR-Cter, respectively.

These results confirmed that V2R-Cter, GHSR-Cter and β2AR-Cter behaved as disordered proteins in size-exclusion chromatography (Uversky, 2012). In addition, far UV CD spectra showed a minimum around 198 nm, characteristic of unfolded proteins (Woody, 2010) (Figure 1C), and a negative shoulder around 222 nm suggesting the presence of residual secondary structures (Woody, 1996) (Figure 1C). The Kratky plots extracted from SAXS data were typical of disordered regions with no clear maximum and a monotonic increase along the momentum transfer range (Figure S2) (Sibille & Bernadó, 2012; Cordeiro *et al*, 2017a, 2017b). Additionally, ^{15}N-HSQC spectra of the three C-terminal regions showed a reduced amide protons spectral dispersion (around 1 ppm) typical of a disordered protein (Figure 1D) (Sibille & Bernadó, 2012; Delaforge *et al*, 2018).

Altogether, bioinformatics analyses and experimental data confirm the disordered nature of the C-terminal regions of V2R, GHSR and β2AR. Furthermore, these results highlight the presence of residual secondary structures, such as in many IDPs/IDRs.

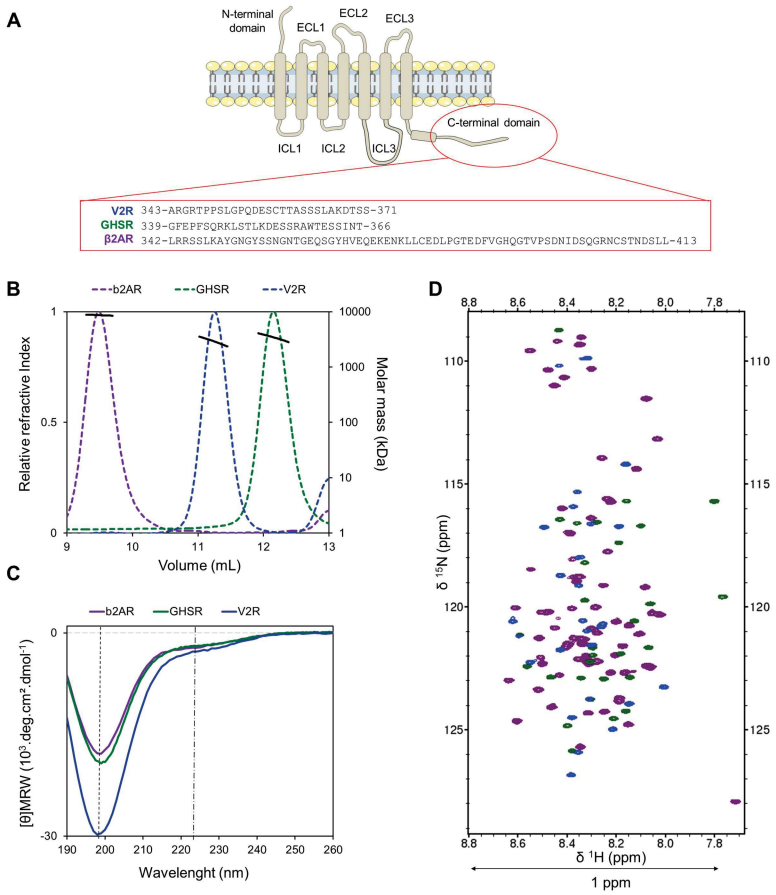


Figure 1: The C-terminal domain of V2R-Cter (blue), GHSR-Cter (green) and β2AR-Cter (purple) are monomeric, disordered with transient secondary structures. This color code is maintained for all panels. **A)** Schematic representation of a GPCR. GPCRs share a common core domain with seven transmembrane helices (7TM) link by extracellular (ECL) and intracellular (ICL) loops. These loops as well as the N- and C-terminal domains are variable in length and sequence. The sequences of the three GPCR-Cters used in this study appeared just below. **B)** SEC-MALS curves (colored dashed lines). Molar mass derived from MALS are indicated by thick black line (right axis). **C)** Far-UV Circular Dichroism (CD) spectra (colored lines) present a minimum around 200 nm (black dashed line) and a shoulder at 220 nm (black dashed and dotted line) characteristic of disorder proteins with transient secondary structure contents. **D)** ^{15}N-HSQC spectra display a low proton spectral dispersion (~ 1 ppm) typical of IDPs. HSQCs were recorded on 300 µM samples at 700 MHz and 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

Location of the transient secondary structures in V2R, GHSR and β2AR C-termini.

Pre-structuration profile of GPCR-Cters. In order to localize the transient secondary structures of these three GPCR C-terminal regions, we performed a NMR study as described in (Senicourt *et al*, 2020). Before NMR investigation, the backbone assignments were performed on the studied C-terminal domains of GPCRs (Figure S3) (CODE BMRB, in progress). Then, we used a consensus of four NMR parameters to highlight the secondary structure content. First, the ^{13}C Chemical Shift Index (CSI), which is highly sensitive to the backbone conformations (Wishart & Sykes, 1994a).. Three-bond H_N-H_a J-coupling constants ($^3J_{HNHA}$), which are related to the φ angles of the polypeptide chain, were compared to random coil scalar coupling using RC_3JHNHa server (Vuister & Bax, 1993; Shen *et al*, 2018). Residual Dipolar Couplings (RDCs), which are related to the orientation of the amide backbone vector to the magnetic field (Bax et al, 2001), give information on the location and secondary structure type. These data were also compared to back-calculated RDCs on random-coil ensembles computed with Flexible-Meccano (Bernadó *et al*, 2005b; Ozenne *et al*, 2012). Finally, dynamic parameters, such as heteronuclear $^{15}N\{^1H\}$-NOEs, longitudinal (R_1) and transverse (R_2) relaxation rates, give information on local backbone mobility on the ps-ns timescale, while R_2 is also sensitive to motions on the µs to ms timescale (global conformation motion). Thus, R_2/R_1 ratio reveals slow conformational motions or conformational/chemical exchange. Depending of the NMR experiments, the GPCR-Cters concentrations varied but did not affect the conformations (Table S3, Figure S4).

Vasopressin V2 receptor C-terminal domain (V2R-Cter)

For V2R-Cter, a central helical region was identified from residues 356 to 364, called V2-1 from now on (Figure 2). Indeed, ^{13}C CSI presented positive values in this region (Figure 2D). The presence of this helix is in agreement with disorder and secondary structure predictions (Figure 2B and -C). Globally, in V2R-Cter, 32 % of the residues displayed $^3J_{HNHA}$ scalar coupling values below 6 Hz, normally assigned to helical conformations, and the rest were between 6 and 8 Hz, consistent with a random coil (RC) polypeptide chain (Figure S5). Moreover, experimental $^3J_{HNHA}$ scalar couplings for the V2-1 region were lower than the predicted ones (Shen *et al*, 2018), substantiating its helical conformation. Further evidence for transient secondary structure elements in V2-1 came from dynamical parameters. V2R-Cter showed low heteronuclear $^{15}N\{^1H\}$-NOE values as expected for a disordered polypeptide. However, in the V2-1 region, heteronuclear NOE values were above the average (-0.09 ± 0.01), suggesting an enhanced rigidity in the region (Figure 2E). Residues of V2-1, adopted larger R_2 values than the average (2.80 ± 0.05 Hz), suggesting the presence of slow dynamic processes. R_1 and R_2/R_1 also displayed slightly higher values for V2-1 than their averages, 1.75 ± 0.03 Hz and 1.59 ± 0.04, respectively (Figure S6). This transient helical secondary structure in V2-1 was also highlighted by $^1D_{NH}$ RDC values. RDCs were measured in alcohol mixture ($C_{12}E_5$/hexanol media) and in filamentous bacteriophage Pf1. However, the quality of alignment was checked on the quadrupolar splitting of D_2O and was higher in alcohol medium (27 Hz) than in Pf1 medium (12 Hz). Thus, we used $^1D_{NH}$ RDC data extracted in alcohol mixture to get better accuracy. While RDC values were mainly negative, as observed in disordered proteins, segment-encompassing V2-1 displayed

higher values than these expected for a random coil, which was in agreement with the presence of a helix in this region (Figure 2F).

Growth hormone secretagogue 1a receptor C-terminal domain (GHSR-Cter)

For GHSR-Cter we identified two transient secondary structures: from residues 345 to 348 (called GH-1) and from residues 356 to 361 (called GH-2) (Figure 2). These regions are predicted to form helices (Figure 2C) and CSI showed positive values, suggesting helical conformations (Figure 2D). The CSI profile also presented high positive values for 354-355 that could be related to the presence of a turn as predicted by the secondary structure predictor SOPMA. $^{3}J_{HNHA}$ scalar coupling confirmed that GHSR-Cter was mostly disordered (50 %) with transient helical conformations (46 %) and a small portion of extended conformation (4 %) (Figure S5). The comparison of experimental and random coil $^{3}J_{HNHA}$ scalar couplings was in agreement with the presence of transient helices in GH-1 and GH-2 (Figure S5). Additionally, in these two regions, heteronuclear NOE, R_1, R_2, and R_2/R_1 showed higher values than their averages, 0.11 ± 0.01; 2.11 ± 0.04 Hz; 3.17 ± 0.06 Hz and 1.49 ± 0.04, respectively (Figure 2E and S6). This indicated less flexibility in these regions, suggesting some local structuration. When measuring RDCs, a small interaction between alcohol mixture and GH-1 region (chemical shift change > 0.01 ppm between ^{15}N-HSQC spectra) was observed (Figure S7), thus we analyzed RDCs measured in phage Pf1 medium (quadrupolar splitting of 18 Hz). $^{1}D_{NH}$ RDCs exhibited lower (GH-1 region) and higher (GH-2 region) values than these expected from a random coil, suggesting an extended and a helical conformation in these regions, respectively (Figure 2F).

βeta 2 adrenergic receptor C-terminal domain (β2AR-Cter)

For β2AR-Cter, we identified two regions forming secondary structures: from residues 349 to 357 (called β2-1) and from 368 to 376 (called β2-2). They were predicted to be more ordered and to contain helical secondary structure (Figure 2A-B). However, CSI presented negative values for β2-1 and positive values for β2-2, corresponding to an extended and a helical conformation, respectively (Figure 2D). $^{3}J_{HNHA}$ scalar couplings were consistent with an overall unstructured protein (73 % of the residues) and smaller content of helical (25 %) and extended conformations (1%) (Figure S5). In β2-2, experimental couplings were lower than these predicted for a random coil, suggesting a helical conformation in this region. Note that glycines are not predicted by the RC_3JHNHa server, thus scalar couplings in β2-1 could not be properly compared to random-coil values, (Shen et al, 2018) (Figure S5). In β2-1 and β2-2, heteronuclear NOE and R_2 values were above their average, 0.11 ± 0.01 and 3.44 ± 0.01 Hz, respectively. This indicated restricted flexibility in these regions (Figure 2E and S6). However, in β2-2, R_1 remained flat. Consequently, R_2/R_1 displayed larger value than the average (1.83 ± 0.03), suggesting conformational fluctuations on the µs-ms time-scale instead of chemical exchange. RDCs measurement in Pf1 medium showed a small interaction with β2-2 (chemical shift change > 0.01 ppm) (Figure S7), thus we used RDC data extracted from the alcohol mixture (quadrupolar splitting of 31 Hz). Compared to random coil values computed with FM, experimental RDCs showed more positive values in β2-1 and more negative values in β2-2 (Figure 2F), suggesting a helical and an extended conformation, respectively. However, RDCs are not only sensitive to local structuration such as CSI, but to long-range conformation (Bernadó *et al*, 2005a; Salmon *et al*, 2010). Indeed, Paramagnetic relaxation enhancement (PRE) data of C406A variant

revealed long-range contacts between β2-1 and a neighboring region of β2-2. The C378 labeled with the paramagnetic probe, which just lies after β2-2 region, induced a reduction of the intensity ratio in β2-1 region (Figure S8). These long-range interactions could affect the overall shape of the molecule in solution and thus, the profile of RDCs. This might explain the absence of consensus between CSI and RDC data.

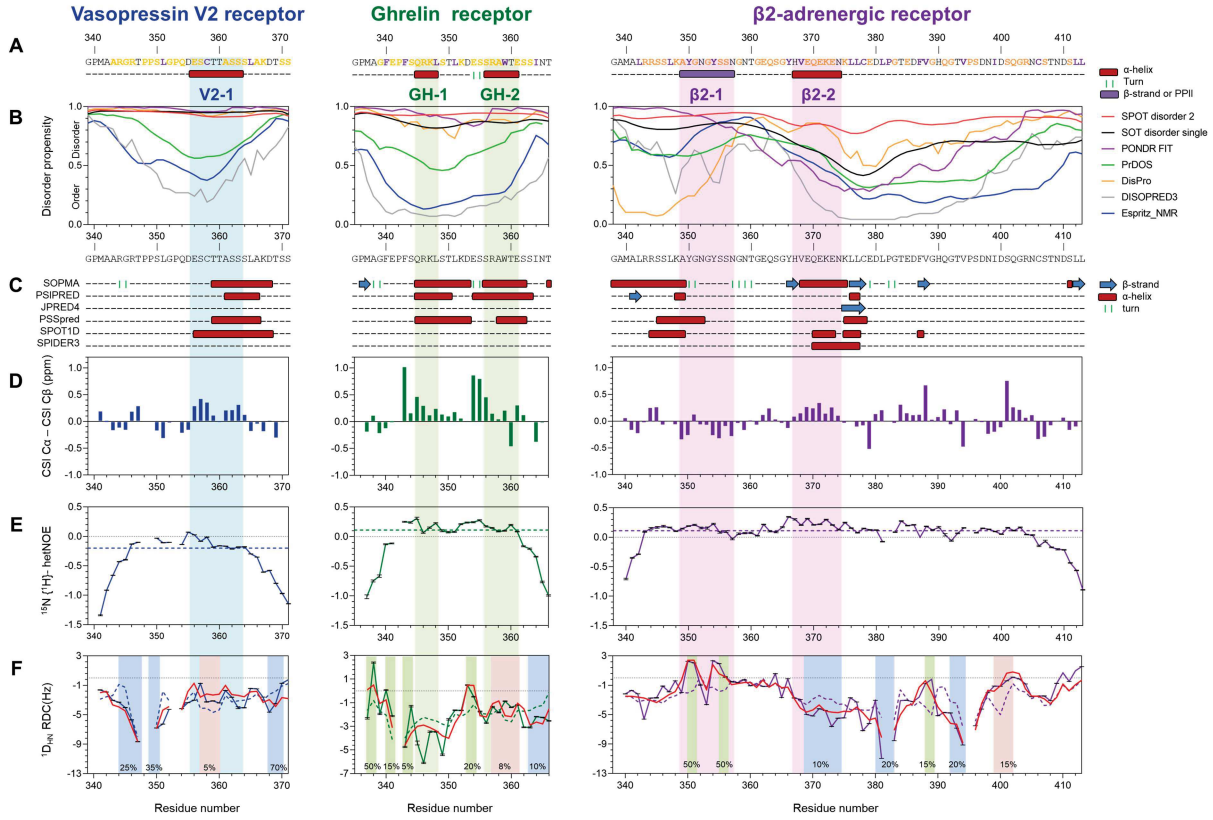


Figure 2: Bioinformatics predictions, secondary structure propensity, backbone dynamics and RDC conformational profile of the disordered V2R-Cter (blue), GHSR-Cter (green) and β2AR-Cter (purple). A) GPCR C-termini are composed of more than 50% disorder-promoting residues (in orange). Ordered-promoting residues are indicated in purple. Secondary structures obtained by NMR secondary structure consensus are indicated under the sequence. Helix, β-strand, extended conformation (β-strand or polyproline helix 2, PPII) and turns are represented as red cylinder, blue arrow, purple box and green bars, respectively. **B)** Disorder prediction by SPOT-Disorder 2 (red), SPOT-Disorder Single (black), PONDR-FIT (purple), PrDOS (green), DisPro (orange), DISOPRED 3 (grey) and Esrpitz-NMR (blue). The disorder/order threshold (0.5) is indicated in black line. **C)** Secondary structure prediction by SOPMA, PSIPRED, JPRED4, PSSpred, SPOT 1D and SPIDER 3 web-servers are represented for helices (red cylinder), strands (blue arrow) and turns (green). **D)** Computed secondary structure CSI Cα – CSI Cβ using random coil chemical shifts from POTENCI. **E)** Heteronuclear $^{15}N\{^{1}H\}$-NOEs are represented with colored curves and the average with dash colored lines. **F)** Back-calculated and experimental $^{1}D_{HN}$ RDCs. Comparison of experimental RDC (colored according to the GPCR-Cter color code) and back-calculated values computed with FM on a random-coil ensemble (dash black line) and biased ensemble (red line) and populated as indicated in the figure for helical structure (highlighted in red), extended structure (highlighted in blue) or turn (highlighted in green). Spectra were recorded at 700 MHz and 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

Conformational ensemble of GPCR-Cters. To further study the presence of transient secondary structures in the three C-terminal domains of GPCRs, we used the pre-structurated elements determined by NMR as a basis to build biased ensembles using Flexible Meccano (FM) (Figure 2F). Ensembles of 50 000 conformers were built for V2R-Cter, GHSR1a-Cter and β2AR-Cter respectively,

and were used to compute back-calculated RDCs (Figure 2F). Quality of these ensembles was evaluated by comparing the back-calculated RDCs with the experimental RDCs measured in alcohol mixture for V2R-Cter and β2AR-Cter or in phage Pf1 medium for GHSR1a-Cter. For V2R-Cter, the best agreement (χ^2 = 1.41) was obtained when 5% of α-helix was imposed in V2-1 and when poly-proline helices II (PPII) were added at 25%, 35% and 70% in position 344-347, 349-350 and 368-370. For GHSR1a-Cter, the best ensemble (χ^2=1.49) was obtained by adding 8% of α-helix in GH-2 and 20% of type 1 β-turn that was predicted for 353 to 354. Additional secondary structures were added (residues 337-338: 50% of type I β-turns, 340-341: 15% of type I β-turns: 343- 344: 5% γ-turn; 363-366: 10% of β-strand). For β2AR-Cter, the lowest χ^2 (χ^2=1.36) was obtained when the long-range contact (15 Å) identified with PRE data was added between regions surrounding β2-1 (from residues 338 to 357) and β2-2 (from residues 367 to 386). In β2-1, two type I β-turns at 50% were added, and in β2-2, a type II Poly-Proline helix (PPII) at 10% was imposed. This ensemble was also constrained by two others PPII in position 380-383 and 391-394, a type I β-turn in position 388-389 and a helix from residues 399 to 402.

For each C-terminal domains, the introduction of these secondary structures resulted in a better description of the conformational ensembles.

DISCUSSION AND CONCLUSION

G protein-coupled receptors (GPCRs) are the largest family of membrane proteins, with ~ 800 different receptors. Due to the large number of GPCRs, they are involved in a wide range of cellular processes and therefore their dysfunctions are linked to various pathologies (Schöneberg *et al*, 2004). Consequently, GPCRs are the target of one third of the drugs placed on the market (Hauser *et al*, 2017) and understanding their signaling is clinically important. These receptors share a commun structural organization consisting of seven transmembrane helices (7TM) (Figure 1A), whereas their extracellular (ECL) and intracellular (ICL) loops and their extracellular N- and intracellular C-terminus, are very variable in length and amico acid composition (Lagerström & Schiöth, 2008). Despite their low conservation, ICL loops and the C-termini of receptors are involved in their functional interaction with the same cytoplasmic partners, G proteins and arrestins.

Despite the importance of arrestin-dependent signaling in GPCR function, we are just beginning to understand the mechanistic basis for it, even for the best-studied GPCRs. Arrestin binds to the phosphorylated C-terminal domain of the receptor and undergoes conformational changes leading to its activation (for a review see (Gurevich & Gurevich, 2019b). However, how only two non-visual arrestins interact with so many, variables, and not-conserved GPCR C-termini remains an open question. The characterization of the structural features of the C-terminal regions of GPCRs in their free state and in complex wit arrestin is crucial to reveal the molecular basis of signaling and cell regulation. Unfortunately, these regions are often truncated or absents from X-Ray crystallography (RX) and cryo-Electron Microscopy (cryo-EM) structures due to their high flexibility. Indeed, these regions are predicted to be intrinsically disordered regions (IDRs) (Venkatakrishnan *et al*, 2014; Jaakola *et al*, 2005), which makes their characterization less straightforward. IDRs explore an

astronomical number of conformations in solution that we assume in fast equilibrium, with pre-formed secondary structure elements. In the majority of cases, these transient secondary structures, or short linear motifs (SLIMs), are involved in the binding process with their partner (Fuxreiter *et al*, 2004), which makes IDPs very well suited agents for signaling processes, such as arrestin:GPCR interaction. Based on this hypothesis, we characterized the free and basal state of three commonly studied GPCRs: the C-terminal domain of vasopressin V2 receptor (V2R-Cter), Growth Hormone Secretagogue or ghrelin Receptor type 1a (GHSR-Cter) and the β2-Adernergic Receptor (β2AR-Cter). These three class-A receptors are important therapeutic targets (Schöneberg *et al*, 2004) and present different affinities for arrestins (Oakley *et al*, 2000).

The structural characterization of these disordered C-termini was challenging and required the combination of biophysical tools. Above all, we confirmed that V2R-Cter, GHSR-Cter and β2AR-Cter are intrinsically disordered regions (IDRs) containing secondary structures by SAXS, MALS and CD. Then, we used NMR, in combination with computationally generated ensembles, to locate and identify these residual secondary structure elements. The C-terminus of vasopressin V2 receptor displayed a central helix from residues 356 to 364 (V2-1). The C-terminus of ghrelin receptor encompassed two helixes from 345 to 348 (GH-1) and from residues 356 to 361 (GH-2). The C-terminus of β2 adrenergic receptor contained two structured regions: an extended conformation from 349 to 357 residues (β2-1) and a helix from 368 to 376 (β2-2). Moreover, the propensities of these secondary structures in biased ensemble are low (~ 25 % on average), which illustrate the high flexibility of GPCR-Cters and their ability to adopt distinct conformations in solution, a classic characteristic of IDPs/IDRs.

V2R-Cter, GHSR-Cter and β2AR-Cter are variable in length and in sequence (Figure 1A). However, it is interesting to note that their residual secondary structure elements encompassed, or are next to residues known to be phosphorylated by GRKs (Figure 3A). It is accepted that the phosphorylation of the C-terminal domains of GPCRs are important for the interaction with arrestin (for a review see (Gurevich & Gurevich, 2019b)). Thus, these transient secondary structures, or SLiMs, could be involved in arrestin binding.

Moreover, V2-1 region is folded in complex with arrestin in RX and cryo-EM structures (Figure 3B) (Shukla *et al*, 2013; Staus *et al*, 2020; Lee *et al*, 2020). These structures were obtained by the use of a fully synthetic phospho-peptide of V2R C-terminal domain (V2Rpp), truncated or attached to other receptors (chimeric receptor) (Figure 3B), and each of these complexes with arrestin were stabilized with Fab30 antibody. In these structures, the central region of V2Rpp forms a β-strand which interacts with the N-domain of arrestin. Here, we showed that, in solution, this region or SLiM, displayed a helical structure instead of an extended conformation. This result suggests a conformation change upon binding to arrestin and/or phosphorylation, a classical characteristic of SLiM, which would favor arrestin recognition. Indeed, the disorder-to-order transition observed in SLiM has a low entropic cost which favor high specific interaction with a low to moderate affinity, allowing fast reversal of cellular event, such as GPCR signaling (Tompa *et al*, 2015). Moreover, this structural transition might be at the basis of the molecular regulation of GPCR:arrestin interaction. Indeed, β-strand folding was proposed to serve as a general mechanism by which arrestins recognize the phosphorylated carboxy-terminal domains of receptors (Shukla *et al*, 2013). In the pre-structuration profile of GHSR-Cter and

β2AR-Cter GH-1, GH-2 and β2-2 are in helical conformations such as V2-1 region. Thus, we can hypothesize that these regions can interact with arrestin with a similar mechanism to the one of the vasopressin V2 receptor, where their basal conformation changes upon binding and/or phosphorylation.

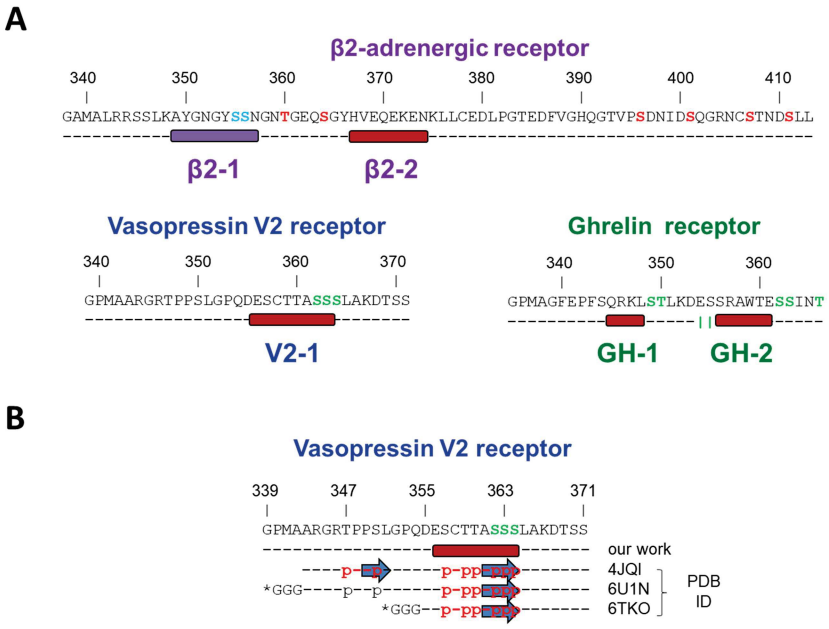


Figure 3: The C-terminal regions of β2AR-Cter (purple), GHSR-Cter (green) and V2R-Cter (blue) revealed different transient secondary structures which could be involved in arrestin binding. **A)** GRK2 (red) and GRK6 (blue) phosphorylated sites according to (Nobles *et al*, 2011) for β2AR, (Bouzo-Lorenzo *et al*, 2016) for GHSR, and (Nobles *et al*, 2007) for V2R are indicated in the sequence. Secondary structures obtained by NMR secondary structure consensus are indicated under the sequence according to figure 2. **B)** Comparison of the binding region identified in complexes between a fully phosphorylated phospho-peptide of vasopressin V2 C-terminus (V2Rpp) and arrestin-2. On the right, PDB ID of V2R:arrestin complexes. On top is represented the sequence of human V2R C-terminus. Residues known to be phosphorylated are colored in green according to (A). Below, V2R-Cter sequences used in each work are represented in dashed line. Binding regions are identified in red and phosphorylated residues are noted as p. Stars (*) indicate if the V2R C-terminus is fused to another receptor (PDB: 6U1N, muscarinic receptor; 6TKO: β1-adrenergic receptor). Helices and β-strands are represented in red cylinder and blue arrow, respectively.

Until now, the free state of GPCR C-termini were unknown due to the high flexibility of these regions. Comparison of our data with GPCRs:arrestin complexes showed that a conformational change is expected after GRK phosphorylation and/or arrestin binding. The phospho-barecode model states that distinct GRK phosphorylation sites at the C-termini of GPCRs lead to distinct arrestin conformations and outcome functions (Ren *et al*, 2005; Nobles *et al*, 2011; Bouzo-Lorenzo *et al*, 2016). Thus, we can anticipate that phosphorylation would dictate the signaling of GPCRs by modulating the folding of their C-termini and/or their folding upon binding. Phosphorylation has already been described as a regulator of IDP folding mechanism for biological function (Bah *et al*, 2015). To test this hypothesis, the characterization of the secondary structure profile for each phosphorylation pattern of GPCR C-termini and the comparison with the basal profile described in this paper would be the key to understand how arrestin dependent signaling pathways are modulated.

MATERIAL AND METHODS

Sample preparation

The (human) cDNA coding for the C-terminus of V2R (343-371), GHSR (339-366) and β2AR (342-413) were ordered to GeneArt® gene synthesis (life technologies). They were cloned into the expression vector pETM33 for GHSR-Cter and V2R-Cter, while β2AR-Cter was cloned into pET1a expression vector. Vectors were transfected into *E. Coli* BL21 DE3 strain (Fisher scientific) and $(His)_6$-GST-tagged proteins were expressed in self-inducible medium N5052 or in minimal medium M9 by adding 2 mM IPTG when the OD_{600} reached 6.0. Labelling with ^{15}N and/or ^{13}C was carried out by growing cells in the previous media with 99 % $^{15}NH_4Cl$ (CORTECNET) and/or 99 % $^{13}C_6$-Glucose (Cambridge Isotope Laboratories) as unique source of nitrogen and carbon respectively. After harvesting, bacteria were resuspended and lysed by sonication in 20 mM Tris 20 pH 7.5, 300 mM NaCl and 2 mM dithiothreitol (DTT) (buffer A) supplemented with anti-protease (*Complete*® EDTA free tablet (Roche)). Cell debris was removed by centrifugation and the soluble fraction was loaded onto 5 mL HisTrap FF column (GE Healthcare). The resin was washed with buffer A containing 10 mM imidazole and proteins were eluted with a linear gradient of buffer A supplemented with 500 mM imidazole. The combined eluate was dialyzed over night at 4 °C in 1 L of buffer A with the 3C PreScission protease (ratio protease:protein 1:100 (w/w)) or TEV protease (ratio protease:protein 1:50 (w/w)) for the pETM33 and pET1a constructs, respectively. Proteases and His-GST tags were removed using gravity column containing 2.5 mL Ni-sepharose (GE Healthcare). Untagged proteins were concentrated using a 2 kDa Millipore concentrator and inject into a HiLoad 16/60 Superdex 75 exclusion column (GE Healthcare) with SEC buffer (50 mM Bis-Tris pH 6.7, 50 mM NaCl, 2 mM DTT). The purified proteins were then concentrated to 200 - 300 µM using Millipore concentrator and aliquots were flash-frozen in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 1 mM EDTA (EthyleneDiamineTetraAcetic acid), 0.5 mM TCEP (Tris(2-CarboxyEthyl)Phosphine) and stored at -80 °C until use.
Concentration of V2R-Cter was determined by refractometry after gel filtration, while concentrations of β2AR-Cter and GHSR-Cter were estimated using absorbance at 280 nm.

Bio-informatic analysis

Disorder prediction. The disorder state of the three C-termini was determine using the Charge-Hydropathy plot computed by PONDR (Romero *et al*, 2001). Disordered propensity of amino-acid were predicted using PONDR-FIT (Xue *et al*, 2010), PrDOS (Ishida & Kinoshita, 2007), SPOT-Disorder (Hanson *et al*, 2018, 2019b), Espritz-NMR (Walsh *et al*, 2012), DisPro (Cheng *et al*, 2005) and DISOPRED3 (Jones & Cozzetto, 2015) predictors.
Secondary structure prediction. Secondary structure predictions were computed with different servers: SOPMA (Geourjon & Deléage, 1995), PSIPRED (Buchan & Jones, 2019), JPRED4 (Drozdetskiy *et al*, 2015), PSSpred (Yan *et al*, 2013) , SPOT1D (Hanson *et al*, 2019a), SPIDER3 (Heffernan *et al*, 2017).

Size Exclusion Chromatography- Multi-Angle Light Scattering (SEC-MALS)

The experiments were performed at 25 °C using a Superdex 75 10/300 GL column (GE HealthCare) connected to a miniDAWN-TREOS light scattering detector and an Optilab T-rEX differential refractive index detector (Wyatt Technology, Santa Barbara, CA). The column was equilibrated in 50 mM BisTris pH 6.7, 50 mM NaCl, 1 mM TCEP and 0.5 mM EDTA buffer filtered at 0.1 µM, and the SEC-MALS system was calibrated with a sample of Bovine Serum Albumin (BSA) at 1 mg/mL. Samples at 1.5 mM, 0.6 mM and 0.7 mM were prepared for V2R-Cter, GHSR-Cter and β2AR-Cter, respectively. For each GPCR-Cter, 40 µL of sample were injected at 0.5 mL/min. Data acquisition and analyses were performed using the ASTRA software (Wyatt).

Circular Dichroism (CD)

Far UV-spectra of the C-termini were recorded in a quartz cuvette (path length 0.1 cm) at 0.08 mg/mL in H_2O at 20 °C using a Chirascan. The ellipticity was scanned from 190 to 260 nm with an increment of 0.5 nm, an integration time of 3 s, and a constant band-pass of 1 nm. Data were treated using Chirascan and, after substraction of the buffer signal, were converted to mean residue ellipticity ($[\theta]_{MRW}$, mdeg.cm².dmole^{-1}) using equation (1) (Chemes *et al*, 2012):

$$[\theta]_{MWR} = [(\theta * Mw)/(L * C * 10)]/(n-1) \qquad (1)$$

where θ is the ellipticity (mdeg), Mw is the molecular weight (g/mol), L is the cuve lengh (cm), C is the protein concentration (mg/mL) and n is the number of peptide bond.

Small-angle X-ray scattering (SAXS) measurement and analysis

Synchrotron radiation small-angle X-ray scattering (SAXS) data were acquired for pm-GPCR-Cters at the SWING beamline at the SOLEIL synchrotron (Saint-Aubin, France) (Bizien *et al*, 2016) using an X-ray wavelength of 1.03 Å and a sample-to-detector distance of 1.99 m. Samples were measured at 15 °C and at two concentrations, 5 mg/mL and 10 mg/mL, for all GPCR-Cters, in 50 mM BisTris pH 6.7, 50 mM NaCl and 2 mM DTT buffer. Before exposure to X-rays, 45 µL of sample were injected into 3 mL Superdex 75 5/150 GL column (GE HealthCare) at 0.2 mL/min, pre-equilibrated into the same buffer as samples. The intensity was measured as function of the magnitude of the scattering vector, s, using equation (2) (Blanchet & Svergun, 2013):

$$s = 4\pi.\sin(\theta)/\lambda \qquad (2)$$

where θ the scattering angle and λ the X-ray wavelength.
The scattering patterns of the buffer were recorded before the void volume of the column (1 mL). The scattering profiles measured covered a momentum transfer range of $0.002 < s < 0.5$ Å^{-1}. Data processing (buffer subtraction) was done automatically with the software available at the beamline.

Nuclear Magnetic Resonance (NMR) spectroscopy

All NMR experiments were performed on a Bruker Bruker Avance III 700 MHz spectrometer, except for the 3D assignment of β2AR-Cter performed on a 800 MHz, and for the $^3J_{HNHA}$ of GHSR-Cter performed on a 500 MHz. The 700 MHz and 800 MHz spectrometers are equipped with a cryogenic

triple-resonance (^{1}H, ^{15}N, ^{13}C) probe and shielded z-gradients. All NMR experiments were recorded at 20 °C in a buffer (named NMR buffer) composed of 50 mM Bis-Tris pH 6.7, 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP, 5% D2O (Eurisotop) and 5 mM DSS-d6 (2,2-dimethyl-2-silapentane-5-sulfonate, Sigma) as internal reference (Wishart & Sykes, 1994b). All experiments used the pulse sequences provided by Bruker Topspin 3.2. Squared cosine apodization was used in indirect dimensions, prior to zero-filling and Fourier transformation using TOPSPIN (version 4.0.6, Bruker) and data processing was performed using NMRFAM-SPARKY (version 1.414, (Lee *et al*, 2015)). For each NMR experiments, concentrations of GPCR-Cters were indicated in Table S3. For all NMR experiments, data were measured for all residues of C-terminus regions excepted proline residues, the residue A339 of β2AR-Cter and the first N-terminal residue.

Backbone assignment. For the sequential assignment of the $^{13}C/^{15}N$ GPCR C-terminus of V2R, GHSR and β2AR, HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCA(CO)NH and HNCACB triple resonance 3D experiments were recorded. H_N, N, CO, C_α and C_β nuclei of all residues were assigned, expected the first N-terminal residue, A339 for β2AR-Cter and proline residues.

Chemical Shift Index (CSI). $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts were used to calculate Chemical Shift Index (CSI) by subtraction of experimental chemical shifts (from the 3D experiment) from random-coil chemical shift computed by POTENCI database (Tamiola *et al*, 2010; Nielsen & Mulder, 2018). CSI were calculated for all residues of C-terminus domains excepted proline residues, the N-terminal residue, A339 for β2AR-Cter and the last residue.

$^{3}J_{HNHA}$ coupling. $^{3}J_{HNHA}$ scalar coupling measurements were obtain accordingly to Vuister and Bax (Vuister & Bax, 1993). Briefly, HNHA experiments were recorded on ^{15}N-labelled GPCR C-termini. Intensity of the cross-peak (S_{cross}) and intensity of the corresponding diagonal peak (S_{diag}) were extracted using Sparky. They were used to calculate the $^{3}J_{HNHA}$ scalar coupling of each amino acid using the equation (2):

$$\frac{S_{cross}}{S_{diag}} = -tan^2(2\pi^3 J_{HNHA}\xi) \qquad (2)$$

Where 2ξ is the total evolution time for the homonuclear $^{3}J_{HNHA}$ coupling, which has been set to 26.1 ms.

$^{3}J_{HNHA}$ scalar coupling were measured for all residues of C-terminus tails excepted proline residues and the first glycine residue. Random coil scalar coupling were predicted using RC_3JHNHa server (Shen *et al*, 2018).

^{1}H-^{15}N Residual Dipolar Coupling (RDC). RDCs were obtained by recording 2D IPAP HSQC spectra (Ottiger *et al*, 1998) in isotropic and anisotropic media. The anisotropic media were obtain by adding a 5 % (w/v) mixture of polyoxyethylene 5-lauryl ether (PEG/C12E5) (Sigma) and 1-hexanol (Sigma) in a molar ratio of 0.85 (Rückert & Otting, 2000) or by adding ~20 mg/mL of filamentous phage Pf1 (Asla biotech) (Hansen *et al*, 1998). Spectra were recorded on ^{15}N-labelled GPCR C-termini in alcohol and phage media. $^{1}D_{NH}$ dipolar couplings were measured from the difference of doublet peak positions in the ^{15}N dimension measured in the anisotropic (J + D) and isotropic (J) spectra. Flexible Meccano (FM) (Ozenne *et al*, 2012) was used to build biased ensemble of 50 000 conformations or random-coil ensemble of 200 000 conformations from which RDCs were back-calculated.

^{15}N relaxation experiments. Relaxation data were measured on ^{15}N-labeled GPCR C-termini for all residues excepted proline residues and the two first N-terminal residues. Heteronuclear ^{15}N{^{1}H}-NOE values were determined from two experiments with on- (saturated spectrum) and off-resonance ^{1}H saturation (unsaturated spectrum) which were recorded in an interleaved manner. The saturation time by 120 ° pulses (~ 10 kHz) was set to 6 s and the recycle delay to 6 s. NOEs values were obtained from the ratio of intensities measured in the saturated (I) and unsaturated (I_0) spectra. Longitudinal (R_1) and transversal (R_2) relaxation rate were measured through acquisition of ^{15}N-HSQC spectra with different relaxation delays: 10, 50, 100, 200, 400, 600, 800, 1 000 ms for R_1, and 16, 32, 64, 96, 160, 240, 480, 640 ms for R_2. For each pick, intensity was fitted to a single exponential decay using Sparky (Lee *et al*, 2015) to obtain the relaxation parameters. For all relaxation parameters, three residues at N- and C-extremity were discarded from the calculation of average values due to their inherent higher flexibility.

Paramagnetic Relaxation Enhancement (PRE). The C378A or C406A variant of ^{15}N β2AR-Cter were labeled on the remaining cysteine using 3-(2-Iodoacetamodi)-proxyl (Merck). Paramagnetic samples were recorded with a recycling delay of 2 s Reference diamagnetic samples were recorded in the same condition after addition of 5 mM fresh ascorbic acid, pH 6.7, in the NMR tube. PRE were analyzed by measuring the peak intensity ratios (I_{para}/I_{dia}) between two ^{15}N-HSQC spectra of paramagnetic and diamagnetic samples. The theoretical profile as in a strictly random coil polymer was calculated according to (Teilum *et al*, 2002).

ACCESSION NUMBERS

The ^{1}H, ^{15}N and ^{13}C assignment of V2R, GHSR and β2AR C-termini are available in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number BMRB (in progress).

SUPPLEMENTARY DATA

CREDIT AUTHOR STATEMENT

Myriam Guillien: Investigation, Methodology, Writing - Original Draft. XXX Investigation, Writing - Review & Editing. XXX: Investigation, Resources. XXX: Software, Formal Analysis. XXX: Investigation, Resources. XXX: Conceptualization, Resources. XXX: Software, Formal Analysis. CCC: Investigation, Writing - Review & Editing. XXX: Conceptualization, Writing - Review & Editing. Nathalie Sibille: Conceptualization, Investigation, Supervision, Writing - Original Draft, Review & Editing.

FUNDING

This work has benefited from support by the Labex EpiGenMed, an « Investissements d'avenir » program, reference ANR-10-LABX-12-01 awarded to MG. N.S. and J.L.B. acknowledge the support of the ANR GPCteR (ANR-17- CE11-0022-01). The CBS is a member of the French Infrastructure for

Integrated Structural Biology (FRISBI) supported by the French National Research Agency (ANR- 10-INSB-05).

ACKOWLEDGMENTS

The authors are grateful to Dr. Baptiste Legrand for CD experiment, Pierre Sanchez and Pr. Gilles Subra for MS support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

IDP, intrinsically disordered protein; IDR, intrinsically disordered region; PTM, post-translational modifications; GPCR, G protein-coupled receptor; GRK, G protein receptor kinase; SEC-MALS, size exclusion chromatography – multi angle light scattering; CD, circular dichroism; SAXS, small angle X-ray scattering; NMR, nuclear magnetic resonance; RDC, residual dipolar coupling; CSI, chemical shift index; PRE, paramagnetic relaxation enhancement; TCEP, Tris(2-CarboxyEthyl)Phosphine ; EDTA, EthyleneDiamineTetraAcetic acid; PPII, poly-proline helices; FM, Flexible Meccano; AVP, arginine vasopressin.

REFERENCES

Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, Sonenberg N, Kay LE & Forman-Kay JD (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature 519: 106–109

Bax A, Kontaxis G & Tjandra N (2001) Dipolar Couplings in Macromolecular Structure Determination. In Methods in Enzymology pp 127–174. Elsevier

Bernadó P, Bertoncini CW, Griesinger C, Zweckstetter M & Blackledge M (2005a) Defining Long-Range Order and Local Disorder in Native α-Synuclein Using Residual Dipolar Couplings. J Am Chem Soc 127: 17968–17969

Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH & Blackledge M (2005b) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. PNAS 102: 17002–17007

Bizien T, Durand D, Roblina P, Thureau A, Vachette P & Pérez J (2016) A Brief Survey of State-of-the-Art BioSAXS. Protein Pept Lett 23: 217–231

Blanchet CE & Svergun DI (2013) Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. Annu Rev Phys Chem 64: 37–54

Bockaert J & Philippe Pin J (1999) Molecular tinkering of G protein-coupled receptors: an evolutionary success. The EMBO Journal 18: 1723–1729

Bouzo-Lorenzo M, Santo-Zas I, Lodeiro M, Nogueiras R, Casanueva FF, Castro M, Pazos Y, Tobin AB, Butcher AJ & Camiña JP (2016) Distinct phosphorylation sites on the ghrelin receptor, GHSR1a, establish a code that determines the functions of ß-arrestins. Scientific Reports 6: 22495

Buchan DWA & Jones DT (2019) The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Research 47: W402–W407

Chemes LB, Alonso LG, Noval MG & de Prat-Gay G (2012) Circular Dichroism Techniques for the Analysis of Intrinsically Disordered Proteins and Domains. In Intrinsically Disordered Protein

Analysis: Volume 1, Methods and Experimental Tools, Uversky VN & Dunker AK (eds) pp 387–404. Totowa, NJ: Humana Press
Cheng J, Sweredoski MJ & Baldi P (2005) Accurate Prediction of Protein Disordered Regions by Mining Protein Structure Data. Data Min Knowl Disc 11: 213–222
Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017a) Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. Adv Exp Med Biol 1009: 107–129
Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017b) Small-angle scattering studies of intrinsically disordered proteins and their complexes. Curr Opin Struct Biol 42: 15–23
Delaforge E, Cordeiro TN, Bernadó P & Sibille N (2018) Conformational Characterization of Intrinsically Disordered Proteins and Its Biological Significance. In Modern Magnetic Resonance, Webb GA (ed) pp 381–399. Cham: Springer International Publishing
Drozdetskiy A, Cole C, Procter J & Barton GJ (2015) JPred4: a protein secondary structure prediction server. Nucleic Acids Res 43: W389–W394
Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, et al (2001) Intrinsically disordered protein. J Mol Graph Model 19: 26–59
Fredriksson R (2003) The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. Molecular Pharmacology 63: 1256–1272
Fuxreiter M, Simon I, Friedrich P & Tompa P (2004) Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins. Journal of Molecular Biology 338: 1015–1026
Geourjon C & Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics 11: 681–684
Gurevich VV & Gurevich EV (2019a) GPCR Signaling Regulation: The Role of GRKs and Arrestins. Front Pharmacol 10
Gurevich VV & Gurevich EV (2019b) The structural basis of the arrestin binding to GPCRs. Molecular and Cellular Endocrinology 484: 34–41
Hansen MR, Mueller L & Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 5: 1065–1074
Hanson J, Paliwal K, Litfin T, Yang Y & Zhou Y (2019a) Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks. Bioinformatics 35: 2403–2410
Hanson J, Paliwal K & Zhou Y (2018) Accurate Single-Sequence Prediction of Protein Intrinsic Disorder by an Ensemble of Deep Recurrent and Convolutional Architectures. J Chem Inf Model 58: 2369–2376
Hanson K, Paliwal K, Litfin T & Zhou Y (2019b) SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning - ScienceDirect.
Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB & Gloriam DE (2017) Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Discov 16: 829–842
Heffernan R, Yang Y, Paliwal K & Zhou Y (2017) Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. Bioinformatics 33: 2842–2849
Ishida T & Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. Nucleic Acids Research 35: W460–W464
Jaakola V-P, Prilusky J, Sussman JL & Goldman A (2005) G protein-coupled receptors show unusual patterns of intrinsic unfolding. Protein Engineering, Design and Selection 18: 103–110
Jones DT & Cozzetto D (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31: 857–863
Kang Y, Zhou XE, Gao X, He Y, Liu W, Ishchenko A, Barty A, White TA, Yefanov O, Han GW, et al (2015) Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. Nature 523: 561–567
Lagerström MC & Schiöth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. Nature Reviews Drug Discovery 7: 339–357
Lee W, Tonelli M & Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31: 1325–1327

Lee Y, Warne T, Nehmé R, Pandey S, Dwivedi-Agnihotri H, Chaturvedi M, Edwards PC, García-Nafría J, Leslie AGW, Shukla AK, et al (2020) Molecular basis of β-arrestin coupling to formoterol-bound β 1 -adrenoceptor. Nature 583: 862–866

Nguyen AH, Thomsen ARB, Cahill TJ, Huang R, Huang L-Y, Marcink T, Clarke OB, Heissel S, Masoudi A, Ben-Hail D, et al (2019) Structure of an endosomal signaling GPCR–G protein–β-arrestin megacomplex. Nat Struct Mol Biol: 1–9

Nielsen JT & Mulder FAA (2018) POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. J Biomol NMR 70: 141–165

Nobles KN, Guan Z, Xiao K, Oas TG & Lefkowitz RJ (2007) The active conformation of beta-arrestin1: direct evidence for the phosphate sensor in the N-domain and conformational differences in the active states of beta-arrestins1 and -2. J Biol Chem 282: 21370–21381

Nobles KN, Xiao K, Ahn S, Shukla AK, Lam CM, Rajagopal S, Strachan RT, Huang T-Y, Bressler EA, Hara MR, et al (2011) Distinct Phosphorylation Sites on the 2-Adrenergic Receptor Establish a Barcode That Encodes Differential Functions of -Arrestin. Science Signaling 4: ra51–ra51

Oakley RH, Laporte SA, Holt JA, Caron MG & Barak LS (2000) Differential Affinities of Visual Arrestin, βArrestin1, and βArrestin2 for G Protein-coupled Receptors Delineate Two Major Classes of Receptors. J Biol Chem 275: 17201–17210

Ottiger M, Delaglio F & Bax A (1998) Measurement ofJand Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. Journal of Magnetic Resonance 131: 373–378

Ozenne V, Bauer F, Salmon L, Huang J -r., Jensen MR, Segard S, Bernado P, Charavay C & Blackledge M (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28: 1463–1470

Ren X-R, Reiter E, Ahn S, Kim J, Chen W & Lefkowitz RJ (2005) Different G protein-coupled receptor kinases govern G protein and beta-arrestin-mediated signaling of V2 vasopressin receptor. Proc Natl Acad Sci USA 102: 1448–1453

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ & Dunker AK (2001) Sequence complexity of disordered protein. Proteins 42: 38–48

Rückert M & Otting G (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. Journal of the American Chemical Society 122: 7793–7797

Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M & Blackledge M (2010) NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. J Am Chem Soc 132: 8407–8418

Schöneberg T, Schulz A, Biebermann H, Hermsdorf T, Römpler H & Sangkuhl K (2004) Mutant G-protein-coupled receptors as a cause of human diseases. Pharmacology & Therapeutics 104: 173–206

Senicourt L, Maire A le, Allemand F, Carvalho JE, Guee L, Germain P, Schubert M, Bernadó P, Bourguet W & Sibille N (2020) Structural insights into the cooperative interaction of the intrinsically disordered co-activator TIF2 with retinoic acid receptor heterodimer (RXR/RAR). bioRxiv: 2020.10.28.359091

Shen Y, Roche J, Grishaev A & Bax A (2018) Prediction of nearest neighbor effects on backbone torsion angles and NMR scalar coupling constants in disordered proteins. Protein Sci 27: 146–158

Shukla AK, Manglik A, Kruse AC, Xiao K, Reis RI, Tseng W-C, Staus DP, Hilger D, Uysal S, Huang L-Y, et al (2013) Structure of active β-arrestin1 bound to a G protein-coupled receptor phosphopeptide. Nature 497: 137–141

Sibille N & Bernadó P (2012) Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochem Soc Trans 40: 955–962

Staus DP, Hu H, Robertson MJ, Kleinhenz ALW, Wingler LM, Capel WD, Latorraca NR, Lefkowitz RJ & Skiniotis G (2020) Structure of the M2 muscarinic receptor–β-arrestin complex in a lipid nanodisc. Nature 579: 297–302

Tamiola K, Acar B & Mulder FAA (2010) Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. J Am Chem Soc 132: 18000–18003

Teilum K, Kragelund BB & Poulsen FM (2002) Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. J Mol Biol 324: 349–357

Tompa P, Schad E, Tantos A & Kalmar L (2015) Intrinsically disordered proteins: emerging interaction specialists. Curr Opin Struct Biol 35: 49–59

Tovo-Rodrigues L, Roux A, Hutz MH, Rohde LA & Woods AS (2014) Functional characterization of G-protein-coupled receptors: A bioinformatics approach. Neuroscience 277: 764–779
Uversky VN (2002) What does it mean to be natively unfolded? European Journal of Biochemistry 269: 2–12
Uversky VN (2012) Size-Exclusion Chromatography in Structural Analysis of Intrinsically Disordered Proteins. In Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools, Uversky VN & Dunker AK (eds) pp 179–194. New York, NY: Springer
Uversky VN, Gillespie JR & Fink AL (2000) Why are 'natively unfolded' proteins unstructured under physiologic conditions? Proteins 41: 415–427
Venkatakrishnan A, Flock T, Prado DE, Oates ME, Gough J & Madan Babu M (2014) Structured and disordered facets of the GPCR fold. Current Opinion in Structural Biology 27: 129–137
Vuister GW & Bax A (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins. Journal of the American Chemical Society 115: 7772–7777
Walsh I, Martin AJM, Di Domenico T & Tosatto SCE (2012) ESpritz: accurate and fast prediction of protein disorder. Bioinformatics 28: 503–509
Wishart DS & Sykes BD (1994) [12] Chemical shifts as a tool for structure determination. In Methods in Enzymology pp 363–392. Elsevier
Woody RW (1996) Theory of Circular Dichroism of Proteins. In Circular Dichroism and the Conformational Analysis of Biomolecules, Fasman GD (ed) pp 25–67. Boston, MA: Springer US
Woody RW (2010) Circular Dichroism of Intrinsically Disordered Proteins. In Instrumental Analysis of Intrinsically Disordered Proteins pp 303–321. John Wiley & Sons, Ltd
Wright PE & Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16: 18–29
Xue B, Dunbrack RL, Williams RW, Dunker AK & Uversky VN (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1804: 996–1010
Yan R, Xu D, Yang J, Walker S & Zhang Y (2013) A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. Sci Rep 3: 2619
Zhou XE, He Y, de Waal PW, Gao X, Kang Y, Van Eps N, Yin Y, Pal K, Goswami D, White TA, et al (2017) Identification of Phosphorylation Codes for Arrestin Recruitment by G protein-Coupled Receptors. Cell 170: 457-469.e13

SUPPLEMENTARY DATA

Structural insights into the intrinsically disordered GPCR C-terminal region, major actor in arrestin-GPCR interaction

Myriam Guillien[a], Assia Mouhand[a], Aurélie Fournet[a], Amandine Gontier[a,b], Aurélien Thureau[c], Frédéric Allemand[a], Jean-Louis Banères[d], Pau Bernadó[a], Nathalie Sibille[a,*,1]

[a] Centre de Biochimie Structurale (CBS), CNRS, INSERM, Univ Montpellier, Montpellier, France

[b] Sorbonne University, École Normale Supérieure, PSL University, CNRS, Laboratoire des biomolécules (LBM), Paris, France

[c] Beamline SWING, Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91190 Gif-sur-Yvette, France

[d] IBMM, UMR5247 CNRS, Pôle Chimie Balard Recherche, 1919 route de Mende, 34293 Montpellier cedex 5, France

* Corresponding author. Email addresses: nathalie.sibille@cbs.cnrs.fr

[1] Lead contact. Tel: +33 4 67 41 77 05; Fax: +33 4 67 41 79 13; Email: nathalie.sibille@cbs.cnrs.fr

Disordered predictor	Attributes		Meta-predictor	Webserver
	Sequence alignment	Structural information		
Pr Dos	Psi-Blast	No	Yes	http://prdos.hgc.jp/cgi-bin/top.cgi
PONDR FIT	No	No	Yes	http://original.disprot.org/metapredictor.php
DISOPRED3	Psi-Blast	Psi-Pred	Yes	http://bioinf.cs.ucl.ac.uk/psipred/
SPOT-disorder 2	Psi-Blast + Hhblits	SPOT-1D	No	https://sparks-lab.org/server/spot-disorder2/
SPOT-disorder single	No	No	No	https://sparks-lab.org/server/spot-disorder-single/
Espritz NMR	No	No	No	http://old.protein.bio.unipd.it/espritz/
DISpro	UniqueProt	Sspro + ACCpro	No	http://scratch.proteomics.ics.uci.edu/

Table S1: Related to Figure 2. Comparison of the disorder predictors used in this study. For each predictor (first column), disordered is predicted based on amino-acid sequence and can be completed with sequence alignment and/or secondary structure information computed with tools listed in the second and third column. Some of these servers use multiple predictors (meta-predictor) as indicated in the fourth column. Webservers are listed in the last column of the table.

Predictor	Sequence alignment	Predicted secondary structure	Webserver
SOPMA	BLAST + Clustal	Helix, Strand, Turn, Coil	https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html
PSIPRED	PSI-BLAST	Helix, Strand, Coil	http://bioinf.cs.ucl.ac.uk/psipred/
JPRED4	PSI-BLAST + HMMER	Helix, Strand, Coil	https://www.compbio.dundee.ac.uk/jpred/
PSSpred	PSI-BLAST	Helix, Strand, Coil	https://zhanggroup.org/PSSpred/
SPOT1D	PSI-BLAST + HHblists	Helix, Strand, Coil	https://sparks-lab.org/server/spot-1d/
SPIDER3	PSI-BLAST + HHblists	Helix, Strand, Coil	https://sparks-lab.org/server/spider3/

Table S2: Related to Figure 2. Comparison of the secondary structure predictors used in this study. Predictors (first column) use sequence alignments computed with tools listed in the second column, predicted secondary structures are listed in the third column. Webservers are reported in the last column of the table.

Protein	3D	$^{3}J_{HNHA}$	relaxation	RDC alcohol	RDC Pf1	PRE
V2R	170	170	373	310	139	NT
GHSR	312	314	124	390	285	NT
β2AR	255	487	487	487	220	50

Table S3: Concentration used for the NMR characterization of phosphomimetic GPCR C-termini of V2R-Cter (blue), GHSR-Cter (green) and β2AR-Cter (purple). Concentration used in each experiment is indicated in µM. The letters NT mean that these experiments have not been carried out.

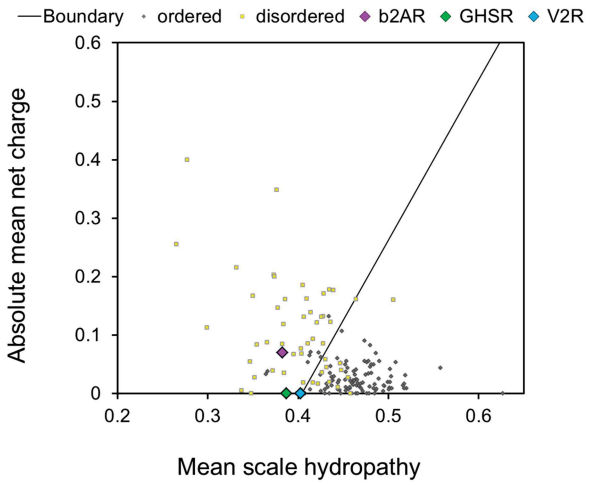


Figure S1: **Charge-Hydropathy plot** (Uversky *et al*, 2000). 54 disordered proteins are shown in yellow circles while 105 ordered proteins are shown in grey circles. The boundary between order/disorder (dark line) was calculated with the equation: $\overline{Hydropathy} = (\overline{net\ charge} + 1{,}151)/2{,}785$. V2R-Cter (blue), GHSR-Cter (green) and β2AR-Cter (pink) are represented with diamonds. This graph was generated using PONDR (www.pondr.com).

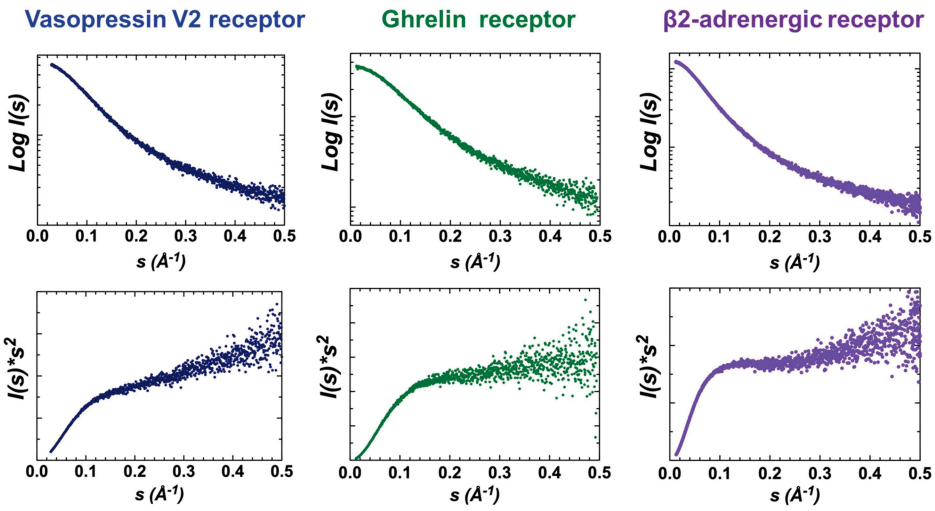


Figure S2: Related to Figure 1. SAXS data of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (purple, right). Semi-logarithmic representation of the experimental SAXS curves versus scattering angle (upper panel) and Kratky plots (bottom panel). The Kratky plots of GPCR-Cters display a typical profile of IDP with no clear maximum and a monotone increase along the momentum transfer range (Sibille & Bernadó, 2012; Cordeiro *et al*, 2017b, 2017a).

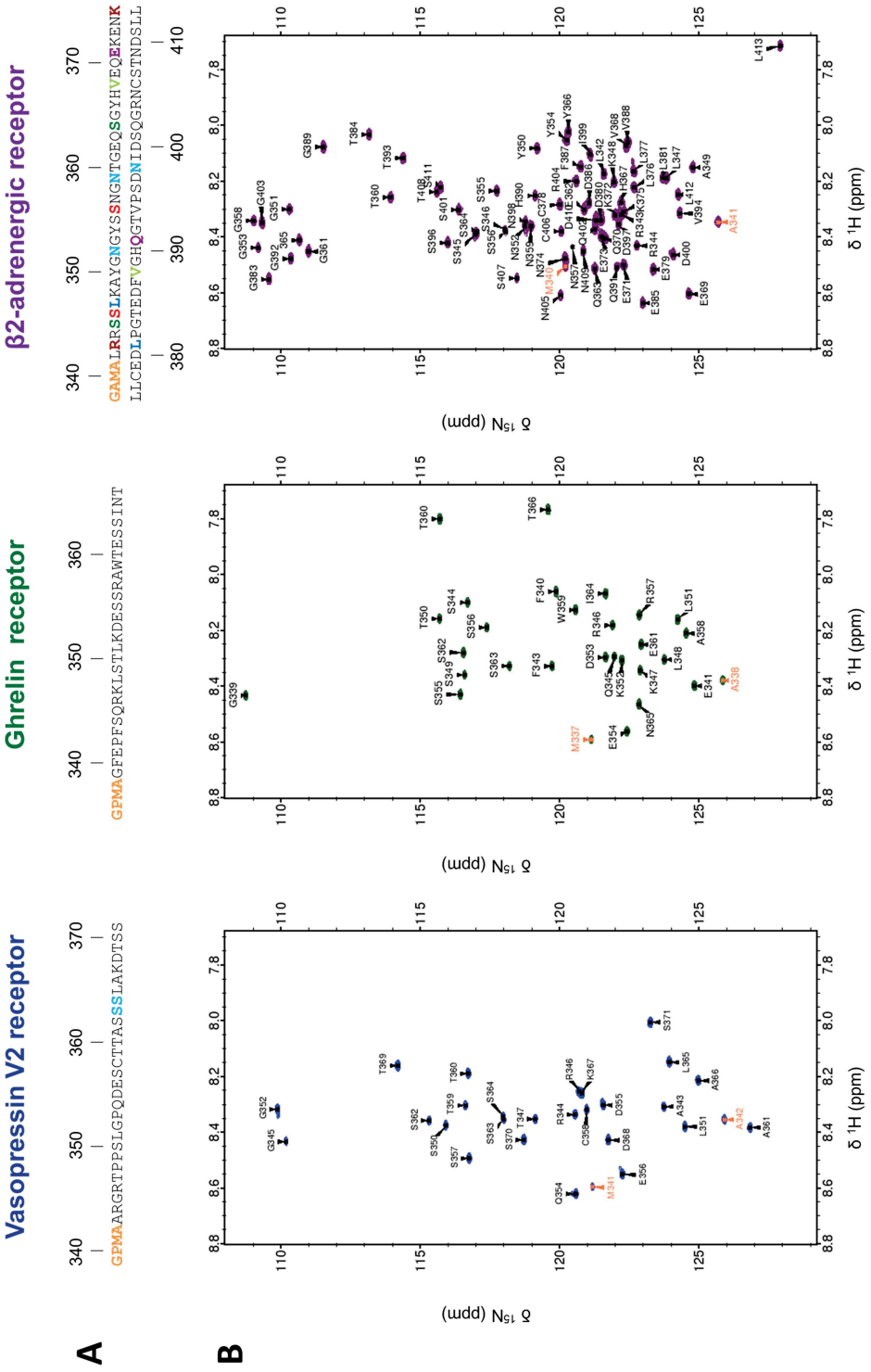


Figure S3: Related to Figure 2. Assignments of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (purple, right). A) Sequence of the GPCRs C-termini. Assigned residues are indicated in black while assignments of overlapping cross-peak are noticed in color. Amino acids coming from the protease cleavage site are shown in orange and set of overlapping residues in ^{15}N-HSQC are in the same color. **B)** ^{15}N-HSQCs were recorded on 300 µM samples at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer. Residue-specific assignments are shown on spectra. Amino acids coming from the protease cleavage site are shown in orange.

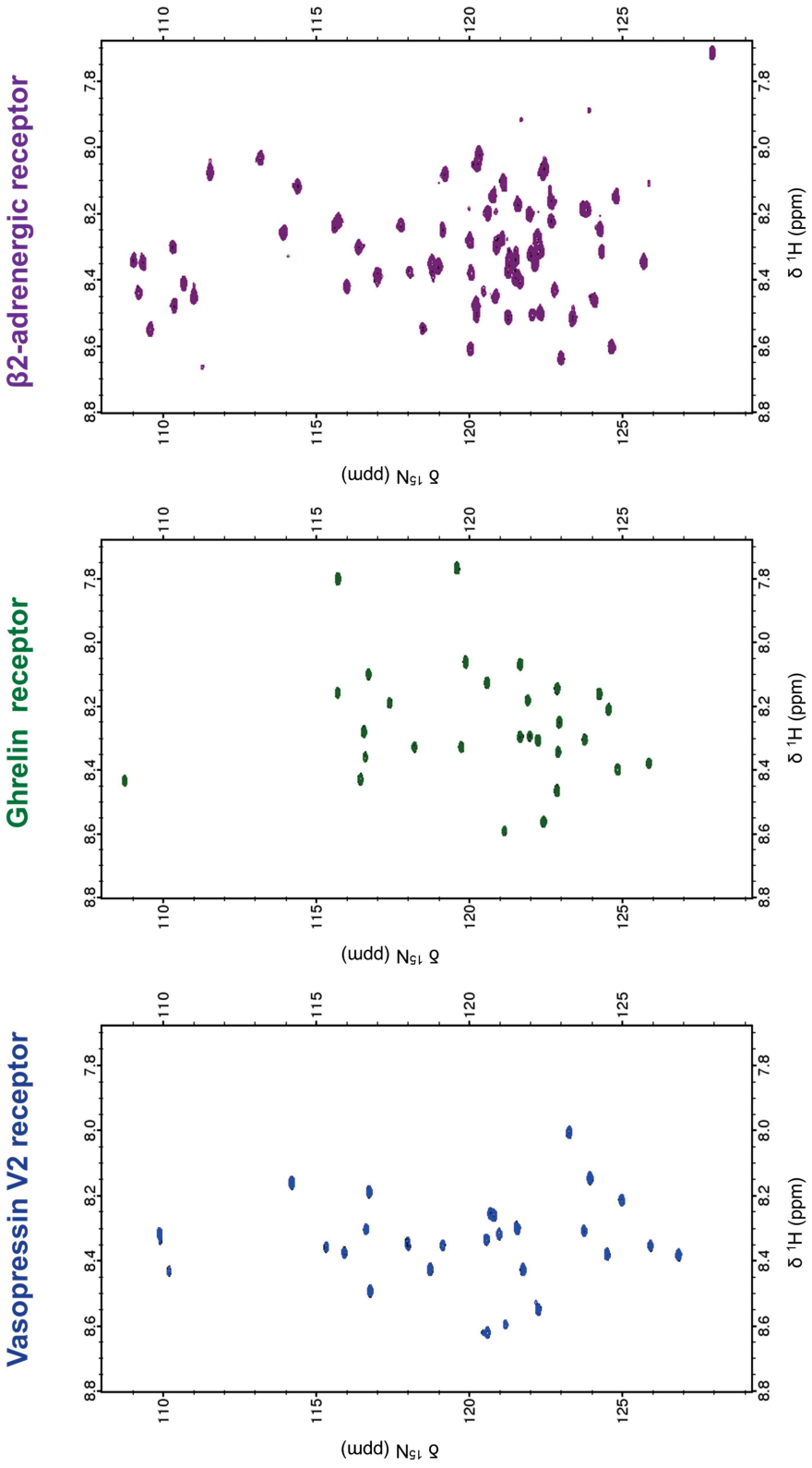


Figure S4: Related to Figure 2. The concentration of GPCR-Cter does not affect their conformations. Overlay of ^{15}N-HSQCs of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (purple, right) recorded on samples at 300 µM (black) or 100 µM (colored) at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl bufer. No chemical shift changes were observed in NMR spectra at different concentrations.

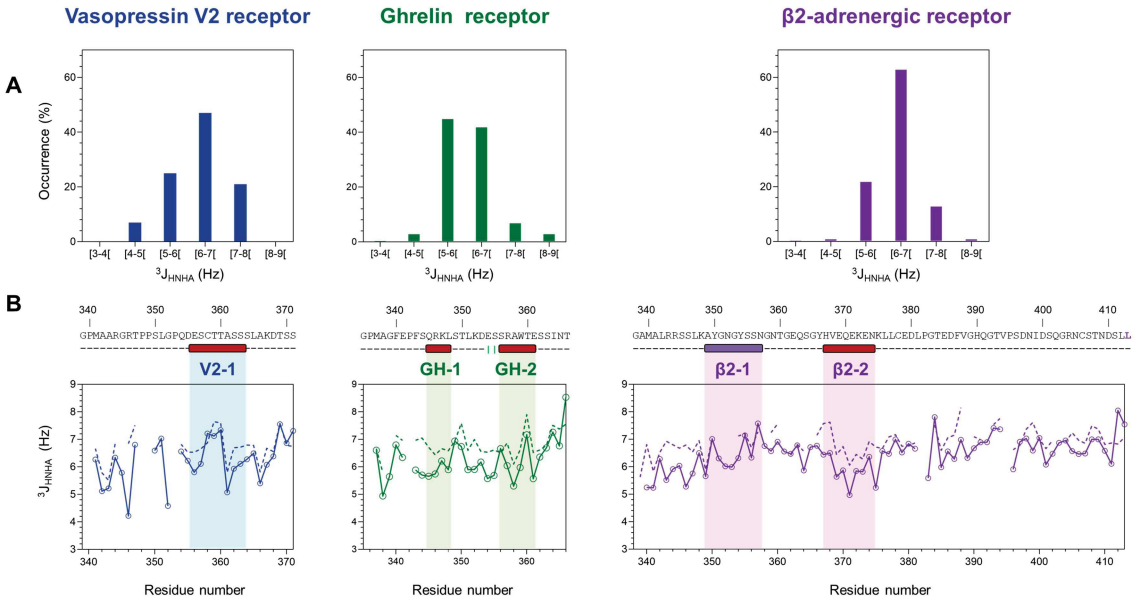


Figure S5: Related to Figure 2. $^{3}J_{HNHA}$ scalar coupling measured for V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (purple, right). A) Histogram of the $^{3}J_{HNHA}$ in Hz. The couplings are consistent with an overall unstructured protein (values from 6 to 8 Hz) with a low content of helical and extended conformation (values below 6 Hz and above 8 Hz respectively). **B)** On top, sequence and NMR consensus of GPCR C-termini are shown according to figure 2. Below, the coupling values (colored line) along the sequence are compared to predicted (RC_3JHNHa server) couplings for a fully disorder chain (dash colored line). $^{3}J_{HNHA}$ were recorded at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

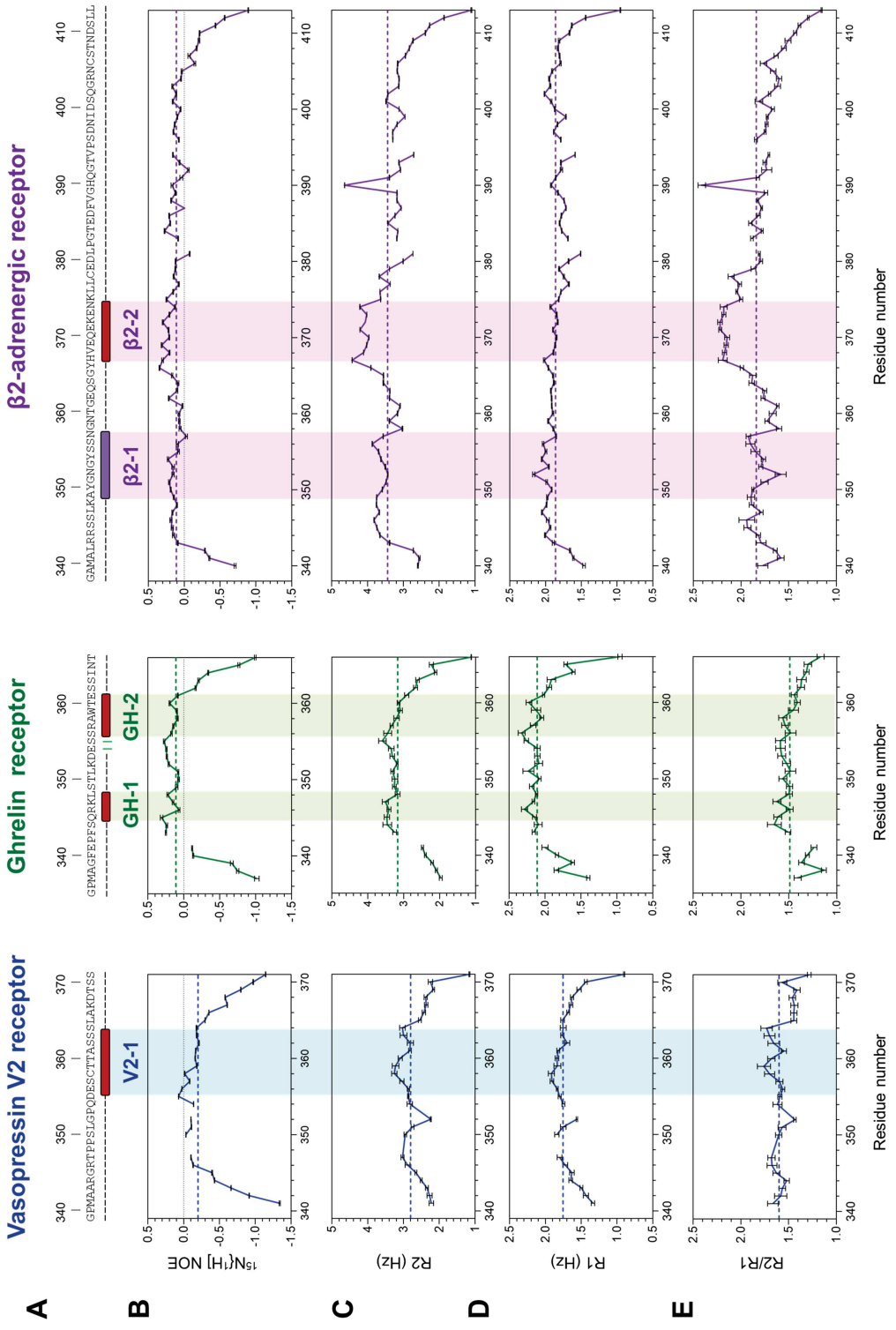


Figure S6: Related to Figure 2E. Relaxation data of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (purple, right). A) Sequence and NMR secondary structure consensus are indicated according to figure 2. **B)** $^{15}N\{^{1}H\}$ heteonuclear NOE, **C)** transversal relaxation rates R2, **D)** longitudinal relaxation rates R1 and **E)** R2/R1 ratio are shown in colored lines. Average values are indicted by dashed colored line. Relaxation data were recorded at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

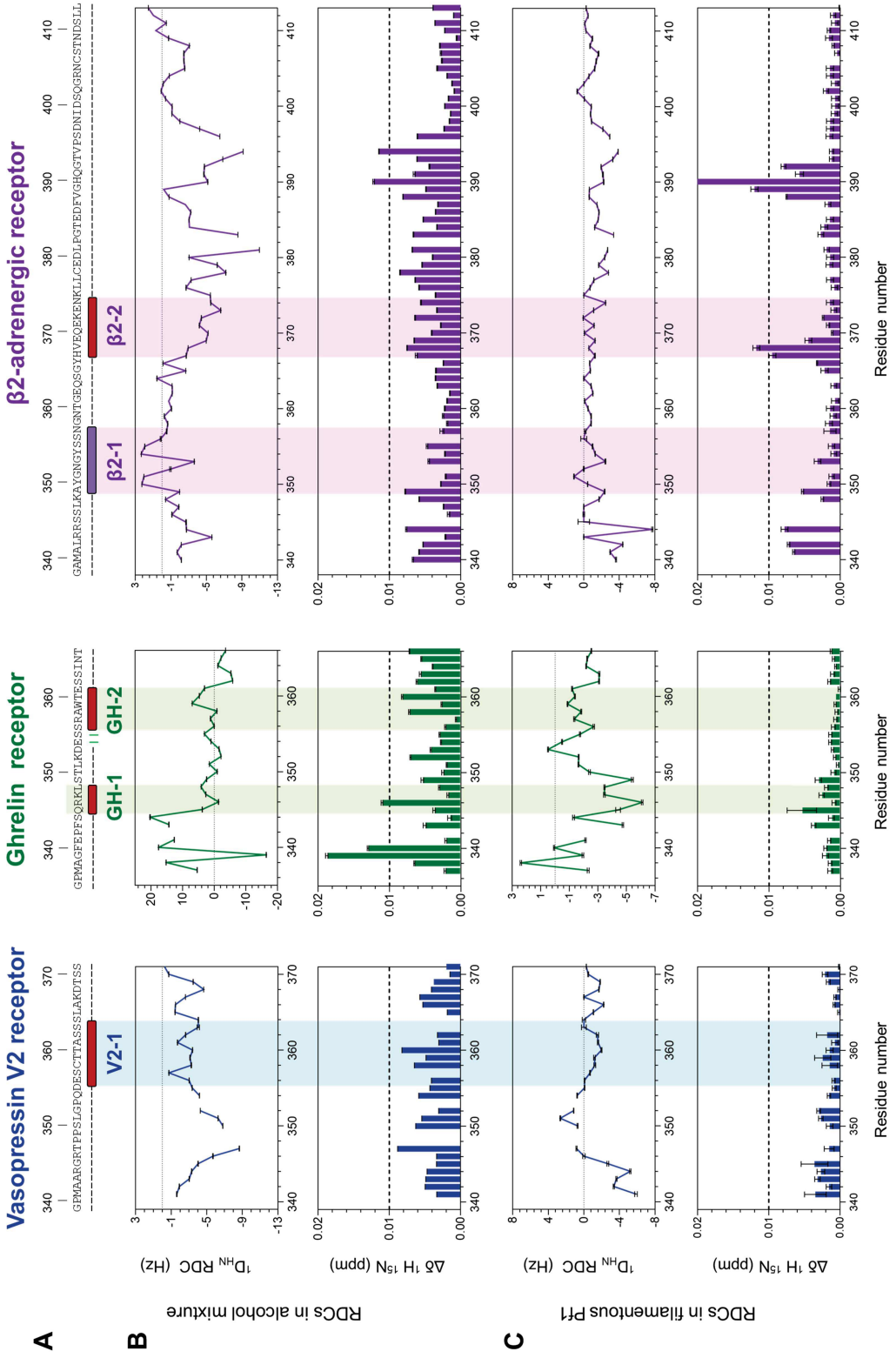


Figure S7: Related to Figure 2F. $^{1}D_{NH}$ RDCs values of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (purple, right). A) Sequence and NMR secondary structure consensus are indicated according to figure 2. RDCs data measured in **B)** alcohol mixture or in **C)** filamentous phage Pf1. In each media, $^{1}D_{NH}$ RDCs profile (top), chemical shift perturbations (bottom) are shown. RDCs were measured at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

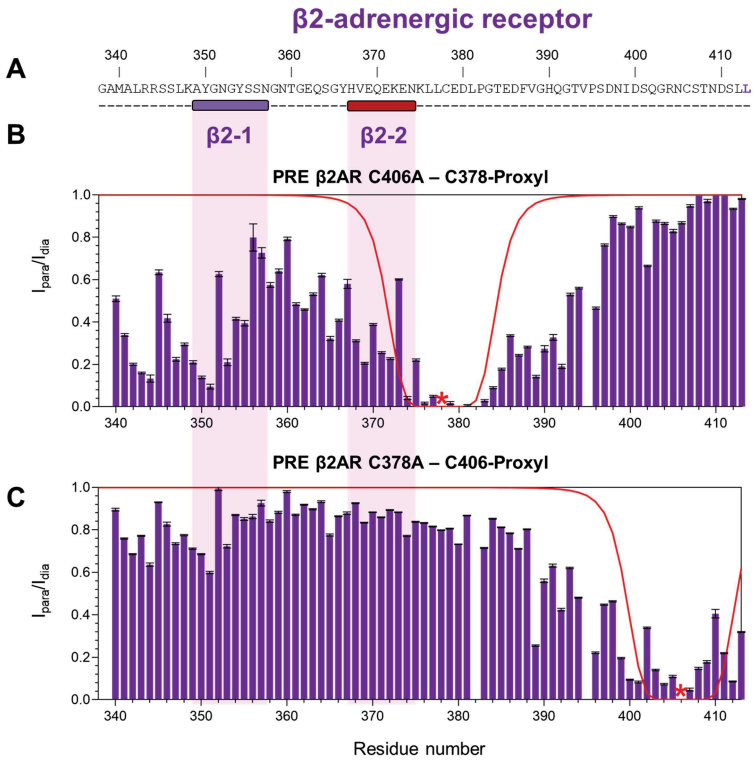


Figure S8: Related to Figure 2. PRE data of β2AR-Cter variants. **A)** Sequence and NMR secondary structure consensus are indicated according to Figure 2. **B)** PRE data of C406A mutant labeled on C378 (red star). The C378 labeled with the paramagnetic probe, which just lies after β2-2 region, induced a reduction of the intensity ratio in β2-1 region, suggesting long-range interaction **C)** PRE data of C378A mutant labeled on C406 (red star). The C406 labeled with the paramagnetic probe, does not lead to reduction of intensity ratio. This region does not form long-range interaction with other regions. PREs were calculated from the ratio of intensity (I_{para}/I_{dia}) of ^{15}N-HSQC spectra measured for the paramagnetic (I_{para}) sample and for the diamagnetic sample (I_{dia}). PRE were recorded at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer on samples at 50 µM. Theoretical PRE curves for a completely disordered protein are shown in red.

REFERENCES

Cordeiro, T.N., Herranz-Trillo, F., Urbanek, A., Estaña, A., Cortés, J., Sibille, N., and Bernadó, P. (2017a). Small-angle scattering studies of intrinsically disordered proteins and their complexes. Curr Opin Struct Biol 42, 15–23.

Cordeiro, T.N., Herranz-Trillo, F., Urbanek, A., Estaña, A., Cortés, J., Sibille, N., and Bernadó, P. (2017b). Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. Adv Exp Med Biol 1009, 107–129.

Sibille, N., and Bernadó, P. (2012). Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochem Soc Trans 40, 955–962.

Uversky, V.N., Gillespie, J.R., and Fink, A.L. (2000). Why are “natively unfolded” proteins unstructured under physiologic conditions? Proteins 41, 415–427.

4.2 ARTICLE 2 : PROBLEMATIQUE ET RESUME

Caractérisation structurale et dynamique des variants phosphomimétiques des domaines C-terminaux des RCPGs

Dans l'article précédent, nous avons montré que les domaines C-terminaux des récepteurs V2R, GHSR et β2AR sont des régions intrinsèquement désordonnées (IDRs) qui contiennent de courtes structures secondaires transitoires, qui pourraient être impliquées dans l'interaction avec l'arrestine. Dans le modèle actuel d'interaction, le modèle du phospho-barcode, le domaine C-terminal des RCPGs est phosphorylé sur différents sites selon la GRK associée, ce qui affecterait la conformation de l'arrestine liée et ses fonctions cellulaires (Nobles et al, 2011). Cependant, les mécanismes moléculaires associés à cette interaction sont encore peu connus à ce jour et ce modèle manque de démonstration expérimentale. Notamment, l'impact de la phosphorylation et/ou de la liaison de l'arrestine sur la conformation des domaines C-terminaux ne sont toujours pas décrits. Pour combler ce gap, nous avons entrepris l'analyse structurale et fonctionnelle des domaines C-terminaux de nos trois RCPGs (RCPG-Cter) sous une forme mimant la phosphorylation. Pour ce faire, nous avons créé des variants phosphomimétiques (pm) dont les sérines et thréonines, connues pour être phosphorylées par les GRKs, ont été remplacées par des acides glutamiques. Ces variants ont été caractérisés avec la même stratégie que celle utilisée pour la caractérisation des domaines C-terminaux non-phosphorylés du RCPG (sauvage, wt) dans l'article 1 (p. 129).

L'analyse RMN a révélé la présence d'un brin bêta chez les variants phosphomimétiques qui est absent chez les RCPG-Cters sauvages. Cela indique un changement conformationnel du domaine C-terminal, dû aux mimes de la phosphorylation. De façon intéressante, l'interaction entre les variants phosphomimétiques des RCPGs et l'arrestine-2 montre que leur région d'interaction correspond à celle ayant subi la transition structurale en brin bêta. De plus, nous avons montré que la région d'interaction du variant phosphomimétique de V2R reste en brin bêta lorsqu'il est lié à l'arrestine-2. Ainsi, nos résultats suggèrent un mode d'interaction par sélection conformationnelle (pour revue sur la sélection conformationnelle voir (Di Cera, 2020)). Lors du processus d'interaction, l'arrestine

existerait dans de multiples conformations en équilibre. La conformation adoptée par le domaine C-terminal des RCPGs sélectionnerait une conformation de l'arrestine, qui aurait un positionnement optimale de ses poches conservées pour se lier.

Par ailleurs, nous n'observons pas d'interaction entre les RCPG-Cters sauvages et l'arrestine-2, ce qui confirme l'importance de la phosphorylation et du repliement des domaines C-terminaux pour la reconnaissance à l'arrestine.

Ainsi, il est possible que la phosphorylation par les GRKs dicte le repliement des domaines C-terminaux des RCPGs. La structuration adoptée par les RCPG-Cters pourrait alors sélectionner une conformation particulière de l'arrestine (sélection conformationnelle), et donnerait lieu à des fonctions spécifiques de cette dernière.

En résumé :

Les variants phosphomimétiques sont capables d'interagir avec l'arrestine, ils sont donc fonctionnels.

Les mimes de la phosphorylation induisent un changement de conformation des régions C-terminales des RCPGs, dont une structuration en brin bêta.

Ce brin bêta est impliqué dans l'interaction avec l'arrestine.

L'interaction avec l'arrestine n'affecte pas la conformation des variants phosphomimétiques, ce qui suggère un mode d'interaction par sélection conformationnelle.

Dans cet article, j'ai contribué à la production des variants phosphomimétiques et j'ai effectué leur caractérisation structurale (CD, MALS, SAXS, RMN). J'ai également contribué à leur caractérisation fonctionnelle (étude d'interaction).

La partie 'matériel et méthode' de l'article reprend les protocoles présentés précédement dans le chapitre 3, matériel et méthode, il n'est donc pas nécessaire de le lire.

Phospho-code dictates GPCR C-terminal domain conformation and arrestin interaction

Myriam Guillien[a,†], Assia Mouhand[a,†], Aurélie Fournet[a], Aurélien Thureau[b], (Marjorie/Michela)[c], Pau Bernadó[a], Jean-Louis Banères[c], Nathalie Sibille[a,*,1]

[a] Centre de Biochimie Structurale (CBS), CNRS, INSERM, Univ Montpellier, Montpellier, France
[b] Beamline SWING, Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91190 Gif-sur-Yvette, France
[c] IBMM, UMR5247 CNRS, Pôle Chimie Balard Recherche, 1919 route de Mende, 34293 Montpellier cedex 5, France
† These authors contributed equally to this work
* Corresponding author. Email addresses: nathalie.sibille@cbs.cnrs.fr
[1] Lead contact. Tel: +33 4 67 41 77 05; Fax: +33 4 67 41 79 13; Email: nathalie.sibille@cbs.cnrs.fr

ABSTRACT

Arrestin dependent G protein-coupled receptor (GPCR) signaling pathway is regulated by the phosphorylation state of GPCR C-terminal domain. G protein-coupled receptor kinases (GRKs) generate different pattern of phosphorylation according to their type and subsequently lead to the activation of distinct arrestin-dependent cellular functions. The so-called phospho-barcode model states that this phosphorylation impacts on the way arrestin interacts with these regions, affecting its conformation and consequently the intracellular signal arrestin triggers. In this study, we investigated the receptor C-terminal region conformation of three class A GPCRs: the vasopressin V2 Receptor (V2R), the Growth Hormone Secretagogue or ghrelin Receptor type 1a (GHSR) and the β2-Adernergic Receptor (β2AR). We have shown, using GRKs phosphomimetic variants, the impact of phosphorylation on their C-terminal region conformations. We compared their basal conformation (non-phosphorylated) to a conformation mimicking the phosphorylated state (GRK2/3 and GRK5/6 variants). We showed that the mimicked phospho-sites induced a conformational transition that engendered arrestin-2 binding. These pre-formed secondary structure elements, or short linear motif (SLiMs), can be defined as the functional motifs that arrange charged residues to promote arrestin recognition. Thus, we can hypothesis that the first step for the cellular signaling specificity of GPCRs could be encoded in the C-terminal regions structuration which will consecutively modulate arrestin conformations and GPCR:arrestin signaling outcomes.

Keywords

GPCR; arrestin; Intrinsically Disordered Proteins or Regions (IDPs/IDRs); NMR; interaction

INTRODUCTION

Arrestins are key regulators of G Protein-Coupled Receptor (GPCR) signaling and they are involved in three main functions: desensitization, internalization and signaling linked to other scaffolding proteins. How these distinct functions are specifically triggered upon GPCR activation remains an open

question. A decade ago, it was proposed that the outcome of arrestin signaling depends on the phosphorylation state of active receptors, leading to the so-called phospho-barecode model (Nobles *et al*, 2011). This model states that, upon activation by their ligand, GPCRs are phosphorylated at different sites on their C-terminal domain, impacting their interaction with arrestin, in turn affecting arrestin conformation and, ultimately, the intracellular signal triggered by arrestine. Later, atomic-resolution structures of GPCR:arrestin complexes revealed different organization of arrestins, supporting the phospho-barcode model (Zhou *et al*, 2017; Staus *et al*, 2020; Huang *et al*, 2020; He *et al*, 2021). Additionally, based on the crystal structure of the rhodopsin:arrestin complex, Zhou and co-workers proposed a set of phosphorylation motifs required for the high-affinity binding to arrestin (Zhou *et al*, 2017). These phosphorylation codes correspond to a long p[S/T]XXp[S/T]XXp[S/T]/E/D motif or a short p[S/T]X p[S/T]XXp[S/T]/E/D motif, where p[S/T] represents a phospho-serine or phospho-threonine, and X any amino acid residue, except proline in the second XX occurrence. These motifs seem to be conserved across the major GPCR subfamilies despite the very low sequence conservation of GPCR C-terminal domains. Among GPCR:arrestin complexes, few of them revealed that negatively charged residues and phosphates of GPCR-Cter make a network of electrostatic interactions with conserved basic residues of arrestin pockets (Zhou *et al*, 2017; Min *et al*, 2020; He *et al*, 2021). Consequently, it was proposed that these complementary interactions induce conformational changes in arrestin and determine its specific function (He *et al*, 2021). This study also supported the existence of a functional phosphorylation barcode system. However, there is still a lack of experimental data explaining how GRK phosphorylation affects the C-terminal domain of GPCRs and leads to different interactions with arrestin.

Interestingly, the GPCR C-terminal domain presents all the structural, dynamic and functional characteristics of intrinsically disorder region (IDR). Intrinsically disorder proteins (IDPs) or IDRs are highly flexible protein with a low content of secondary structure, whose specific role in fundamental signaling and regulation processes started to be revealed two decades ago (Wright & Dyson, 2015). We have previously described the transient secondary structure contents of three truncated GPCR C-terminal domains by Nuclear Magnetic Resonance (NMR): the vasopressin V2 Receptor (V2R) (from residue 343 to 371), the Growth Hormone Secretagogue or ghrelin Receptor type 1a (GHSR) (from residue 339 to 366) and the β2-Adernergic Receptor (β2AR) (from residue 342 to 413) (Guillien *et al*, in preparation). Here, we compared these basal conformations to these of the variants mimicking GRK phosphorylation pattern. The specific sites phosphorylated by each GRKs not being described, we designed phosphomimetic variants of these truncated C-terminal regions by substituting all the potential GRK2/3 and GRK5/6 phosphorylation sites (Nobles *et al*, 2007, 2011; Bouzo-Lorenzo *et al*, 2016) by glutamic acids (Figure 1A). We showed important conformational changes between the basal and the phosphomimetic forms. Interestingly, the extended conformation appearing in the phosphomimetic variant was found to form specific interactions with arrestin. Moreover, these binding regions of the three truncated GPCR C-terminal domains studied here contained motifs described as putative phosphorylation motif, in line with the above mentioned models of interaction. Altogether, our results suggest that GRK phosphorylation might induce the formation of a β-strand in the C-terminal

domains of GPCRs that may serve as a general mechanism by which arrestins recognize the phosphorylated carboxy-terminal domains of receptors.

RESULTS

Structural characterization of the phosphomimetic forms of the three GPCR C-terminal domains. Conformational changes between their basal (wt) and phosphomimetic (pm) forms

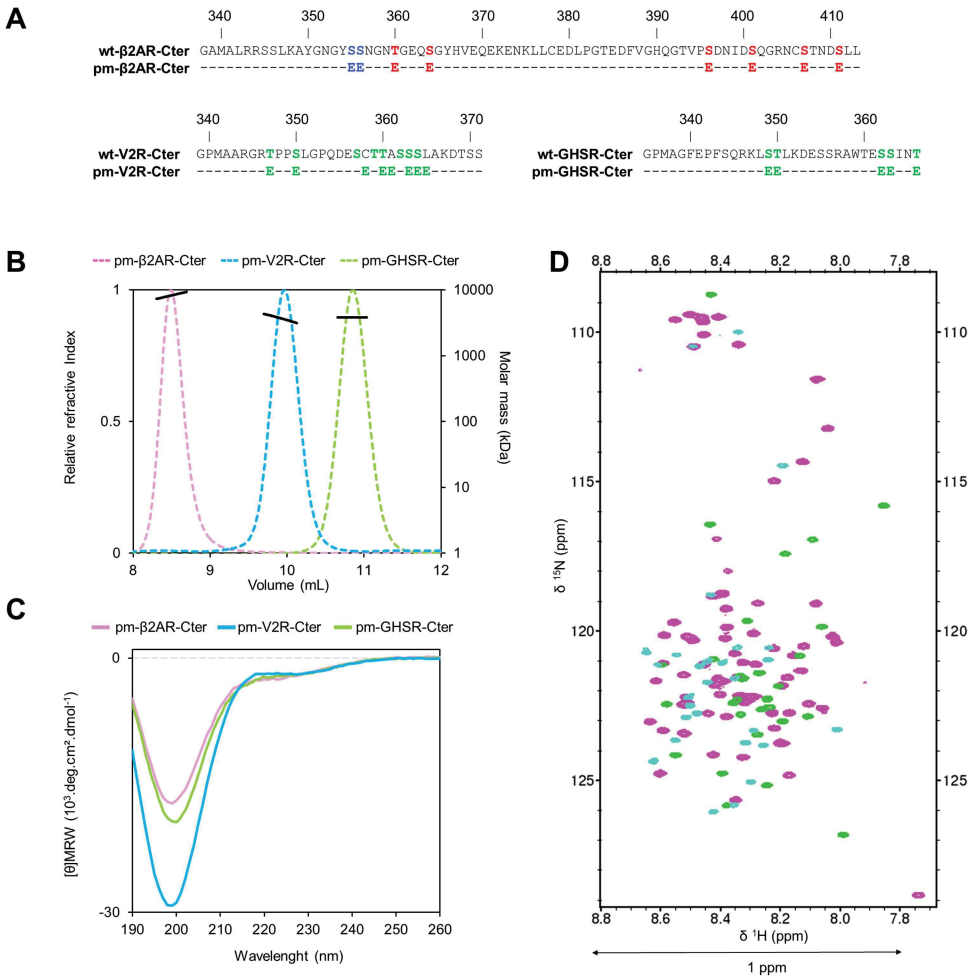


Figure 1: V2R-Cter (blue), GHSR-Cter (green) and β2AR-Cter (pink) phosphomimetic (pm) variants are monomeric, disordered with transient secondary structures. This color code is maintained for all panels. **A)** Sequence of GPCR C-terminal domains in their basal (wt) and phosphomimetic (pm) form. GRKs (green), GRK2 (red) and GRK6 (blue) phosphorylated sites according to (Nobles *et al*, 2011) for β2AR, (Bouzo-Lorenzo *et al*, 2016) for GHSR, and (Nobles *et al*, 2007) for V2R are indicated in the wt sequence. These sites were mutated to glutamic acid (E) in the pm variants. **B)** SEC-MALS curves of pm variants (colored dashed lines) revealed a single elution pic for each GPCR-Cter at a volume that corresponds to the volume of standard proteins with a molecular mass greater than 13 kDa. However, the masses derived from MALS analysis (full black line) are 3.7 (± 2.4 %) for V2R-Cter, 3.8 (± 1.8 %) for GHSR-Cter and 8.3 (± 0.8 %) for β2AR-Cter. They are in agreement with their monomeric molecular weight of 3.6, 3.7 and 8.5 kDa, respectively and were confirmed by mass spectrometry (MS, data not show). This behavior on a SEC column is typical of disordered protein (Uversky, 2012). **C)** Far-UV Circular Dichroism (CD) spectra (colored lines) of pm variants present a minimum around 200 nm and a shoulder at 220 nm characteristic of disorder proteins with transient secondary structure contents (Woody, 1996). **D)** ^{15}N-HSQC spectra of pm-V2R-Cter, pm-GHSR-Cter and pm-β2AR-Cter showed a low proton spectral dispersion, typical of IDPs (Sibille & Bernadó, 2012). HSQCs were recorded on 300 µM samples at 700 MHz, 20 °C, 50 mM Bis-Tris pH 6.7, 150 mM NaCl.

In order to study the effect of GRK phosphorylation on the conformation of the C-terminal domains of the three GPCRs, we produced phosphomimetic variants of these regions (pm-GPCR-Cter) (Figure 1). These variants were designed by replacing the phosphorylated serines and threonines by GRK2/3 and 5/6 reported in the literature (Nobles *et al*, 2007, 2011; Bouzo-Lorenzo *et al*, 2016) by glutamic acids (Figure 1A). Biophysical characterization of these variants revealed that they were monomeric and disordered proteins (Figure 1B) with transient secondary structures, Circular Dichroism (CD) (Figure 1C) and Nuclear Magnetic Resonance (NMR) (Figure 1D), similarly to their non-phosphorylated forms (Guillien *et al*, in preparation). These observations were also confirmed by Small Angle X-ray Scattering (SAXS) where their Kratky plots displayed the typical profile of an IDP with no clear maximum and a monotone increase along the momentum transfer range (Figure S1) (Sibille & Bernadó, 2012; Cordeiro *et al*, 2017b, 2017a). We worked at different concentrations of pm-GPCR-Cters along the study, but we checked by NMR that it did not affect their conformations (Table S1, Figure S2).

To localize and identify secondary structures, we performed an analogous NMR study as for the non-phosphorylated C-terminal domains (Guillien *et al*, in preparation) and described as well in (Senicourt *et al*, 2020). Before the structural investigation, the backbone assignments were performed for the three phosphomimetic variants (pm) of the studied C-terminal domains of GPCRs (Figure S3) (CODE BMRB, in progress). Below, we describe the NMR study of the C-terminal domains in their phosphomimetic form and we compare the structural features of the non-phosphorylated (Guillien *et al*, in preparation) and the phosphomimetic forms of the three studied GPCR C-terminal domains.

Phosphomimetic Vasopressin V2 receptor C-terminal domain (pm-V2R-Cter)

V2R-Cter phosphomimetic variant presented a large stretch of ^{13}C Chemical Shift Index (CSI) negative values (Figure 2B), from residues 356 to 364 (called V2-1), suggesting an extended conformation in this region. Interestingly, we showed previously that the non-phosphorylated wt form (wt-V2R-Cter) contained a central helical conformation in this very specific V2-1 region (Guillien *et al*, in preparation). This loss of helical content between the wild type and the variant was also observed in the decrease of percentage of $^{3}J_{HNHA}$ scalar coupling bellow 6 Hz from 32% to 29%. However the variant remained mostly disordered with the majority of scalar couplings between 6 and 8 Hz (Figure S4A). Another difference in the CSI profile appeared for residues 346-347 where the pm variant presented highly positive values compared to the wt, certainly related to the presence of a turn. In terms of dynamics, in V2-1 region, relaxation rates R_1 and R_2 displayed lower values for the variant compared to the wild type (Figure S5C and -D). This could be explained by the fact that the extended conformation in pm-V2R-Cter is structurally less constrained than the α-helice in wt-V2R-Cter. But this change in dynamics is not observed in heteronuclear $^{15}N\{^{1}H\}$-NOE (Figure 2C and S5B) suggesting small variation of dynamics. Residual Dipolar Couplings (RDC) values also indicated the presence of an extended conformation in V2-1 region of the variant. We measured RDCs in alcohol mixture or filamentous bacteriophage Pf1 alignment media (quadrupolar splitting of 33 Hz and 20 Hz, respectively). In the alcohol alignment mixture, small chemical shift changes were observed in V2-1 (from 0.01 to 0.02 ppm) suggesting a slight interaction between the medium and the variant, which did not appear for Pf1 medium. As the RDC profiles in both media were similar but reduced in the Pf1 medium, we analyzed

the data in the alcohol mixture. It showed highly negative values in the V2-1 region of the variant compared to the wild type that showed less negative values (Figure 2D and S6), confirming an extended conformation in this V2-1 region, which is not present in the wild type. Interestingly, the RDC profile in Pf1 also showed a high positive value for residues 346-347, which was in good agreement with the potential turn observed in the CSI profile. (Figure 2 and S6C).

Phosphomimetic growth hormone secretagogue 1a receptor C-terminal domain (pm-GHSR-Cter)

The CSI profile of GHSR-Cter phosphomimetic (pm) variant revealed the presence of a α-helix (from residues 345 to 348, called GH-1) and an extended conformation (from 359 to 366, called GH-3) with positive and negative CSI values, respectively (Figure 2B). Previously, we identified two helixes for the wild type form: GH-1 (from residues 345 to 348) and GH-2 (from residues 356 to 361) (Figure 2A) (Guillien *et al*, in preparation). CSI values of the pm variant in GH-1 region were higher than in the wt, suggesting a stronger helical structuration. Interestingly, GH-2 helix found in the wt was shifted downstream and converted in an extended conformation (GH-3) (Figure 2B). The transition from a helical to an extended conformation was confirmed by the decreased from 46% to 29% (lower than 6 Hz) and a small increased from 3% to 4% (upper than 8 Hz) of $^{3}J_{HNHA}$ scalar couplings between the wt and the pm forms (Figure S4A). Also, scalar couplings in GH-3 are higher for the pm variant than for the wt, which is in line with a tendency to form an extended conformation (Figure S4B). Dynamical parameters showed an increase of rigidity in GH-1 and GH-3 regions for the pm variant, with higher averaged heteronuclear NOE (0.18 ± 0.01) and transversal relaxation R_2 rate (3.33 ± 0.04 Hz) compared to the wt (0.11 ± 0.01, 3.17 ± 0.06 Hz) (Figure 2C and S5). Conformational changes between wt and pm were also highlighted by RDC measurements. RDCs measured in alcohol mixture showed a small interaction with the variant (chemical shift change > 0.01 ppm; Figure S6), thus we used the RDC data measured in Pf1 medium with a quadrupolar splitting of 15.2 Hz. In agreement with CSI data, the comparison of RDC profiles showed an increase of RDCs of GH-1 and a slight decrease of these of GH-3 (Figure 2D), suggesting a stronger helical structuration in GH-1 and a switch from a helical (wt) to an extended conformation (pm) from GH-2 to GH-3.

Phosphomimetic β2 adrenergic receptor C-terminal domain (pm- β2AR-Cter)

The CSI profile of the β2AR-Cter phosphomimetic variant displayed three regions with negative and positive values indicating transient secondary structures: an extended conformation from residues 349 to 357 (called β2-1), a helix from 368 to 376 (called β2-2) and an extended conformation from 407 to 413 (called β2-3) (Figure 2, S4, S5 and S6). The wild type NMR characterization gave similar results for the two first regions (Guillien *et al*, in preparation) while no secondary structure was observed in the third identified region (Figure 2B). The appearance of a β-strand in β2-3 was observed in the increase of $^{3}J_{HNHA}$ values between 7 and 8 Hz (from 13 to 28 %), which was localized in β2-3 region (Figure S4). In β2-3 region, heteronuclear NOE and transversal relaxation R_2 data showed a slight increase of their values in the pm variant compared to the basal wt form, while the longitudinal relaxation R_1 remained unchanged (Figures 2C and S5). These results suggested a local structuration and reduced conformational motion in agreement with the presence of a transient secondary structure in this β2-3 region. This additional β2-3 region was also seen in our RDC analysis. RDCs were measured in alcohol mixture and in filamentous Pf1 with a quadrupolar splitting of 36 Hz and 16.5 Hz,

respectively. Weak interactions were observed between the alcohol mixture and the pm variant, and between the Pf1 medium and the wild type in the very same regions (chemical shift > 0.01 ppm at the beginning of β2-2 and from residues 387 to 394 that is out of any delineated regions, Figure S7). To be able to compare pm and wt RDC profiles, we analyzed pm data in alcohol mixture. RDCs adopted more negative values in β2-3 for the pm variant compared to the wild type, suggesting the presence of an extended conformation (Figure 2D, S6). Altogether, comparison of all β2AR-Cter NMR data in their non-phosphorylated (basal wt) and phosphomimetic (pm) forms confirmed a conformational transition of the β2-3 region from a random coil (wt) to a more extended one (pm) (Figure 2A). Finally, no changes were observed in the long-range interaction previously identified in wt-β2AR-Cter (Figure S7).

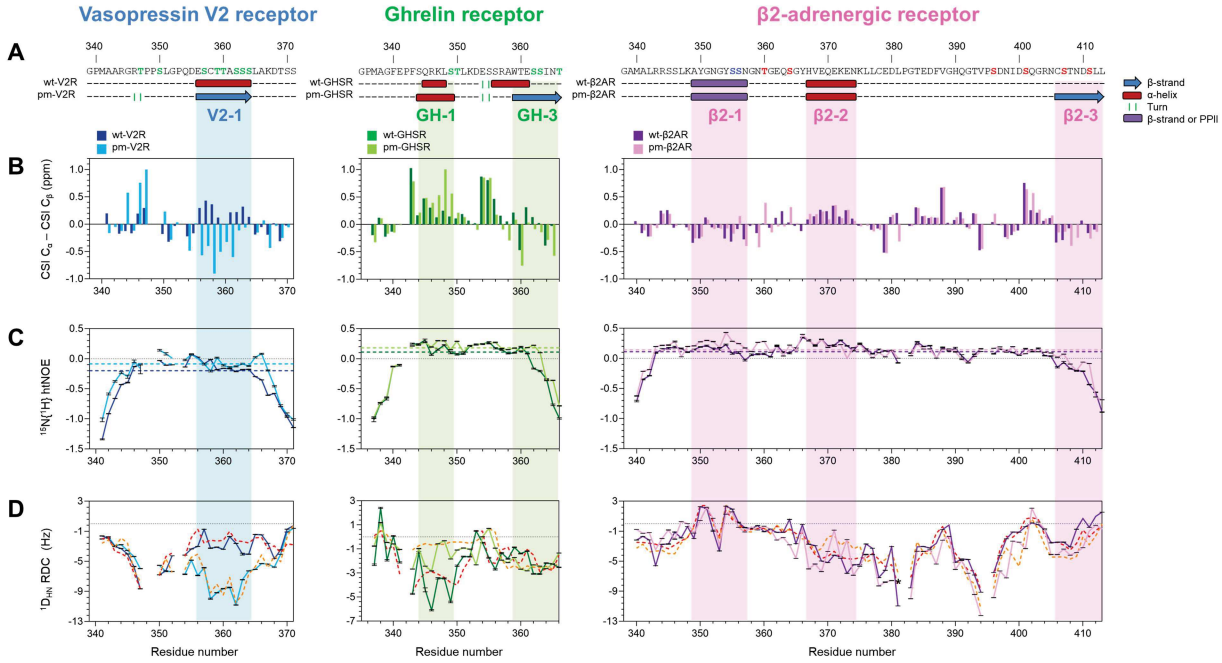


Figure 2: Comparison of the structural features of GPCR C-terminal domains in their basal wild type (dark color, data from Guillien *et al*, in preparation)) and phosphomimetic variant forms (light color) of V2R-Cter (blue, left); GHSR-Cter (green, middle) and β2AR-Cter (purple/pink, right). This color code is conserved in all panels. **A)** Schematic secondary structure consensus of all NMR data along the primary sequences. Known phosphorylation sites mutated to glutamic acid (E) in the variants (pm) are indicated in green, red and blue on the sequence (according to Figure 1). Helix, β-strand, extended conformation (β-strand or polyproline helix 2, PPII) and turns are represented as red cylinder, blue arrow, purple box and green bars, respectively. **B)** Computed secondary structure CSI Cα – CSI Cβ. **C)** $^{15}N\{^{1}H\}$ heteronuclear NOEs profiles from the ratio of intensities measured in saturated (I) and unsaturated (I_0) spectra. Averages are in dash colored line. **D)** Experimental and back-calculated $^{1}D_{NH}$ RDC. Comparison of experimental RDCs measured in bacteriophage Pf1 for GHSR-Cter, alcohol mixtures for V2R-Cter and β2AR-Cter. Back-calculated RDCs from biased ensembles generated with FM (dashed lines) for wt (red) and pm (orange). Outliers due to overlapping residues were removed (dark star). NMR experiments were recorded at 700 MHz and 20 °C in a 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

Conformational ensemble of GPCR-Cters phosphomimetic forms.

To go further in the description of pm-GPCR-Cter conformations, we generated random-coil ensembles of 200 000 conformers for each phosphomimetic variants (pm) using Flexible Meccano (FM) (Bernadó *et al*, 2005b; Ozenne *et al*, 2012). The quality of these ensembles was evaluated by comparing their back-calculated RDCs to the experimental ones recorded in alcohol mixture for V2R-Cter and β2AR-Cter or in Pf1 medium for GHSR. Back-calculates RDCs from random-coil ensemble

show discrepancy with experimental RDCs for each pm variants, especially in beta-stranded regions (V2-1, GH-3 and β2-3, Figure S8). Thus, we include conformational preferences defined by NMR data in biased ensemble of 50 000 conformers. We particularly focused on the comparison between the best ensembles obtained for the GPCR-Cters in their wild type (Guillien *et al*, in preparation) and phosphomimetic forms (Figures 2D, S8). The best agreement for pm-V2R-Cter ensemble (χ^2 = 1.23) was found by imposing: 10 %, 50 % and 25 % of β-strands for residues 352-357, 359-362 (V2-1) and 364-368 (Figure S8A), while wt-V2R-Cter ensemble revealed a 5 % helix content in this specific V2-1 region. Thus, ensemble description of V2R-Cter clearly highlighted a conformational change between pm- and wt-V2R-Cter and a strong propensity of pm variant to adopt a β-strand conformation in V2-1 region. In addition, two poly-proline helixes (PPII) already present in the wt ensemble were also added in position 344-345 (20 %) and 350-351 (30 %). In the best ensemble of pm-GHSR-Cter (χ^2 = 0.77), the major difference between the variant and wild type ensembles was the introduction of 60 % of β-strand in GH-3 (Figure S8B) instead of 8 % of helix in GH-2, respectively. It confirmed the structural transition from a helix (wt) to a β-strand (pm) in GHSR-Cter. In addition, a stronger helical structuration was observed in GH-1 variant (4 % of helix from 349 to 351) (Figure S8B), compared to wild type ensemble where no conformational bias was required in this region. For the β2AR-Cter variant, the best ensemble (χ^2 = 4.21) was obtained by imposing 5 % of β-strand near β2-3 region (Figure S8C), while no constraints were added in this region for the wild type. This result showed that the conformational change to a β-strand also occurred in β2AR-Cter variant but to a lower extent in the two other C-termini. Another difference appeared in β2-2 region where 5 % helix was added instead of a PPII in the wild type. Variant ensemble was also constrained by a 10 % β-turn I in residues 401-402 (Figure S8C). Altogether, wild type ensembles show a conformational transition occurring when introducing phosphomimetic residues involving larger populations of extended conformations.

Arrestin binding sites contain putative phospho-codes

To study the impact of the previously describe conformational changes on the function, we performed interaction experiments of the three GPCR-Cter with their biological partner, arrestin. Indeed, NMR is a suitable tool to study transient interaction (Schumann *et al*, 2007; Williamson, 2013). Location and strength of the interaction can be mapped by following the changes in chemical shift and/or intensity of a labeled protein with an increasing amount of the partner.

With this aim, we recorded a set of ^{15}N-HSQC of ^{15}N GPCR-Cters incubated with different ratios of unlabeled arrestin-2 to perform a Chemical Shift Perturbation (CSP) assay. For the three phosphomimetic variants, changes in chemical shifts were observed when comparing the spectra of the protein in the presence (1:10) and absence of arrestin-2 (Figure S9). For the pm-V2R-Cter, chemical shift changes and an intensity decrease proportional to the amount of added arrestin-2 were observed from residues 356 to 366 (Figure 3 and Figure S10A). Similar results were obtained for pm-GHSR-Cter from residues 358 to 366 (Figure 3 and S10B). Concerning pm-β2AR-Cter, small chemical shift changes were observed for residues 408-413 at a 1:10 ratio (Figure 3), these changes decreased when decreasing the amount of arrestin-2 and only residue 412 showed a reduction of its intensity (Figure S11).

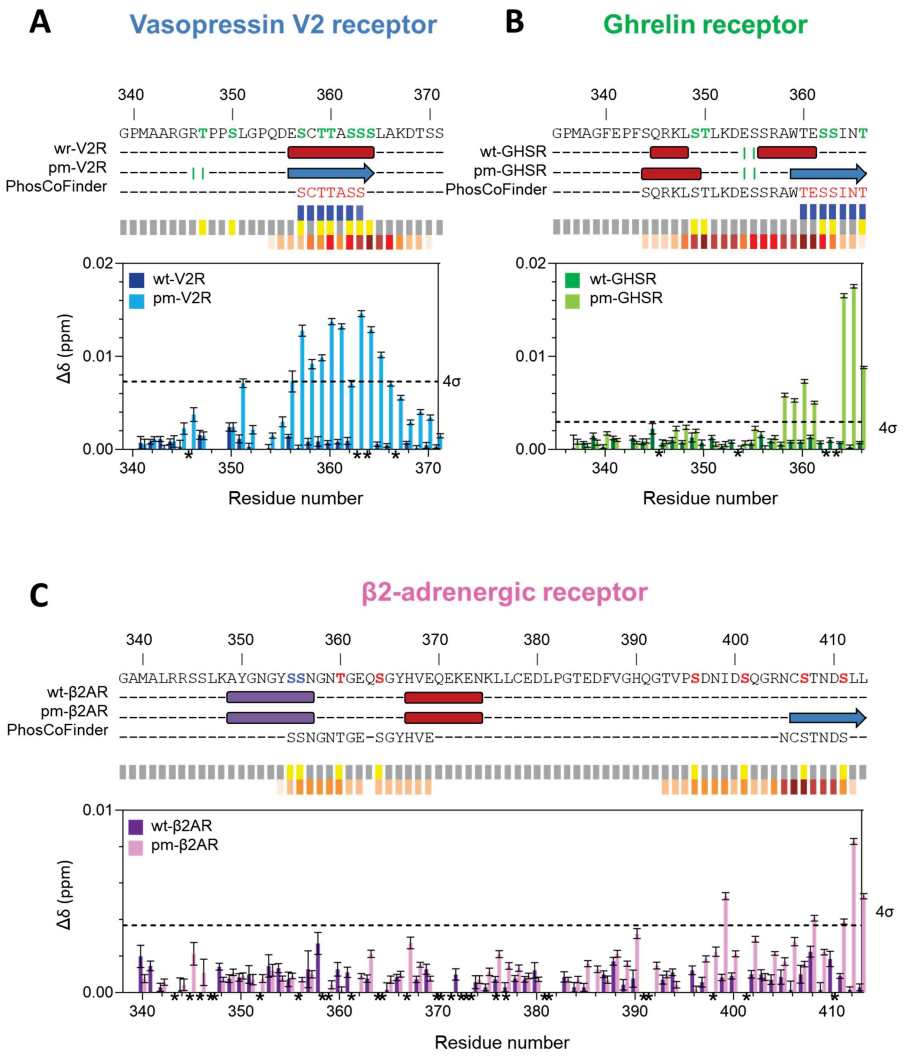


Figure 3: Interaction between arrestin-2 and the C-terminal domains of V2R (A), GHSR (B) and β2AR (C) in their wild type (dark colors) and phosphomimetic (light colors) forms. For each receptor, on top, sequence of C-terminal domain is indicated according to Figure 1. The secondary structure contents of wild type and phosphomimetic variants, and predicted phospho-codes (PhosCoFinder, (Zhou *et al*, 2017)) are shown below. Putative phospho-codes (p[S/T](X)Xp[S/T]XXp[S/T]/E/D motifs) are displayed on three different tracks: complete codes (motif containing three negatively charged residues) are in blue on the top track, the sequence is in grey in the middle, and partial codes (motif containing two negatively charges residues instead of three) are on the bottom. For partial codes, darker colors indicate residues involved in more codes. Known phosphorylation sites from Uniprot and PhosphoSitePlus database (Hornbeck *et al*, 2015) are indicated in yellow in the grey sequence. $^{1}H/^{15}N$ Chemical Shift Pertubation (CSP) observed at ratio 1:10 (10 µM GPCR-Cter: 100 µM arrestin) are under the putative phospho-codes. CSP were recorded on sample at 10 µM with 0 or 100 µM of arrestin-2 at 800 MHz and 20 °C in a 50 mM Bis-Tris pH 6.7, 50 mM NaCl buffer. Cut-offs are in black dashed lines. Outliers due to overlapping residues were removed (dark star, near the residue number).

Importantly, no interaction was observed between the three non-phosphorylated C-terminal domains (wt-GPCR-Cter) and arrestin-2, highlighting the importance of phosphorylation for GPCR recognition of arrestin (Figure 3, S10 and S11).

Noteworthy, the identified binding regions overlapped with the regions undergoing a conformational transition between the wild types and the phosphomimetic variants (Figure 3). Moreover, these regions were also predicted to contain phospho-code (PhoCoFinder (Zhou *et al*, 2017)). These codes were proposed as motifs required for high-affinity arrestin binding. For pm-V2R-Cter and pm-GHSR-

Cter, the binding region encompassed a so-called complete phosphorylation motif (three negatively charges residus), while pm-β2AR-Cter contained solely a partial motif (two negatively charges residues instead of three) (Figure 3). This was in line with a shorter binding region with a weaker chemical shift perturbation (0.009 ppm ± 0.001), suggesting a lower affinity for arrestin-2. On the opposite, the phosphomimetic variant of V2R-Cter experienced the largest chemical shift perturbation with a Δδ of 0.015 ppm (± 0.001) and the largest binding region with a dozen of impacted residues, which should be related to a stronger interaction with arrestin-2. These observations provide the structural bases of the previously reported higher affinity for arrestin of V2R with respect to GHSR1a and β2AR (Oakley *et al*, 2000).

According to microscale thermophoresis (MST) and fluorescence anisotropy (Figure S12), affinities of these transient complexes were very weak. For instance, pm-GHSR-Cter, which showed a clear interaction with arrestin-2 by CSP (Figure 3 and S10B), could not be saturated in MST or fluorescence anisotropy experiments (Figure S12), and we therefore could only estimate that the complexes underwent a very low interaction affinity with a K_d in the ~ 200 µM range.

GRKs phosphorylation acts as a switch of GPCR C-termini conformation?

Conformation changes upon binding to their targets or after post-translational modifications (PTMs) is a quite common mechanism for IDPs (Wright & Dyson, 2015). Here, we showed that the GPCR C-terminal domains conformational state changed between the non-phosphorylated (basal wild type, wt) and the phosphomimetic (variant, pm) form. Moreover, the regions that exhibited conformational transition from helical or random coil to β-extended were involved in the interaction with arrestin. In order to discriminate between a recognition with arrestin that took place due to a pre-configuration of the C-terminal domain of GPCRs thanks to GRK phosphorylation (or phospho-mimicry) or an interaction induced by the folding of the C-terminal domain of GPCRs upon binding to arrestin, we checked if there was a change in conformation before and after binding of arrestin. For the phosphomimetic variant of V2R-Cter, secondary CSI profiles in presence and in absence of arrestin were similar (Figure 4), suggesting that the interaction did not induce a folding but a conformational selection took place. Thus, a pre-configuration with well-disposed charges along the C-terminal domain of GPCRs was necessary for arrestin recognition.

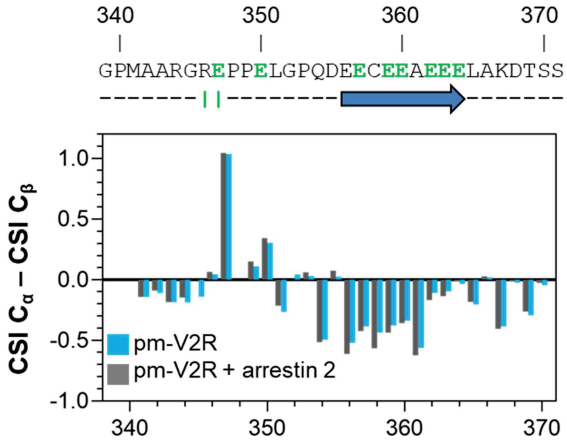


Figure 4: Conformational selection of phosphomimetic V2R-Cter by arrestin-2. ^{13}Cα/Cβ chemical shift index (CSI) computed with POTENCI (Tamiola *et al*, 2010; Nielsen & Mulder, 2018) of pm-V2R-Cter in absence (blue) or in presence of arrestin (grey). 3D spectra were recorded on ^{13}C/^{15}N pm-V2R-Cter (50 µM) in absence or in presence of arrestin-2 (250 µM) at a ratio of 1:5 at 800 MHz and 20 °C in a 50 mM Bis-Tris pH 6.7, 50 mM NaCl buffer.

DISCUSSION AND CONCLUSION

Understanding how GRK phosphorylation of the C-terminal domain of GPCRs (GPCR-Cter) dictates the recruitment of arrestin is conceptually and clinically important. The phosphorylation-barcode model was proposed a decade ego, stating that the phosphorylation pattern on receptor C-terminal domain direct arrestin binding and subsequent functions (Nobles *et al*, 2011). However, how phosphorylation of GPCR-Cter at specific sites lead to distinct arrestin conformation and signaling outcomes remains an open question.

Here, we studied how GRK phosphorylation affected the conformation of GPCR C-terminal domains and their interaction with arrestin-2 using a phosphomimetic strategy. This study was carried out for three receptors, *i.e.* the vasopressin V2 receptor (V2R), the β2-adrenergic receptor (β2AR) and the ghrelin receptor (GHSR). Besides being important pharmacological targets, these receptors belong to the two different classes of arrestin:GPCR complexes (Oakley *et al*, 2000), *i.e.* class A (receptors interacting transiently with arrestins, *e.g.* β2AR and GHSR) and class B (receptors interacting tightly with arrestins, *e.g.* V2R), and therefore they represent models for these two families of GPCRs. We produced phosphomimetic variant of these receptor C-terminal domains (pm-V2R-Cter, pm-GHSR-Cter and pm-β2AR-Cter) by substituting potential GRKs phosphorylation sites by glutamic acids (Figure 1A).

We compared the conformational properties of these GPCR C-termini in their non-phosphorylated, basal, wild type, form (Guillien *et al*, in preparation) with these in the phosphomimetic ones. As their counterpart, variants behaved as intrinsically disordered regions (IDRs) in solution and contained transient secondary structure elements (Figure 1). Localization and identification of these secondary structure elements was possible thanks to the combination of several NMR parameters (Figure 2), SAXS data and computational tools. In the phosphomimetic variants, we identified beta extended conformations that were not present in the basal wild type C-termini. For V2R-Cter, the V2-1 region

underwent a conformational change from a helix (wt) to a β-extended conformation (pm). In GHSR-Cter, the GH-1 helix was conserved and stabilized in the variant, while GH-2 helix was shifted and substituted by a β-strand (GH-3). In β2AR-Cter, a disordered region at the end of the domain folded into a β-strand conformation (β2-3). These results suggested that mimicking GRK phosphorylation causes the formation of extended conformations in localized segments of the GPCR C-termini. Altogether, these results suggested that the GPCR C-terminus functional folding is regulated by GRK phosphorylation. Folding associated to phosphorylation was also observed for the eukaryotic translation factor 4E-binding protein 2 (4E-BP2) (Bah *et al*, 2015). The authors showed that phosphorylation led to a folding of disordered residues into a β-sheet that prevented its association to the eukaryotic translation initiation factor 4E (eIF4E). Thus, phosphorylation was involved in the folding of 4E-BP2 and in its ability to interact with its partner.

Is there a mechanism in the arrestin dependent signalization of GPCR where phosphorylation would dictate GPCR C-terminal domain folding and, thus GPCR:arrestin interaction? To answer this question, we studied the interaction of GPCR-Cters with arrestin-2 by NMR by performing chemical shift perturbation assay (CSP). We showed that phosphomimetic C-terminal domain variants were able to interact with arrestin-2 (Figure 3, S10 and S11). Moreover, no conformational changes appeared upon binding to arrestin-2 in the V2R-Cter variant, suggesting a conformational selection rather than an induced fit or a folding upon binding mode of interaction (Figure 4). Interestingly, no interaction between basal wild type GPCR-Cters and arrestin-2 were observed (Figure 3, S10 and S11), which is in line with the fact that phosphorylation is generally necessary for binding (for review see (Gurevich & Gurevich, 2019a)). We can conclude here that the phosphomimetic models used in this study were functional and could be used as a general tool in the future to study GPCR:arrestin complexes and their associated arrestin-dependent signaling pathways.

Noteworthy, our results showed that the β-extended conformations in the three C-terminal domains, only present in the phosphomimetic variants, corresponded to the binding region with arrestin (Figure 3). Interestingly, several atomic resolution structures, obtained by X-ray crystallography and cryo-electron microscopy, showed a β-extended conformation of the fully phosphorylated C-terminus peptide of vasopressin V2 receptor (V2pp), which formed an anti-parallel β-sheet with the N-domain of arrestin-2 (Shukla *et al*, 2013; Nguyen *et al*, 2019; Staus *et al*, 2020; He *et al*, 2021). A similar fold was also observed for the phosphorylated rhodopsin C-terminus (pRho) in interaction with visual arrestin-1 (Zhou *et al*, 2017). In these complexes, an extensive network of electrostatic interactions exists between the negatively charged residues of the C-termini and three conserved pockets of positive residues in the N-domain of arrestin: A (19, R22 and R175), B (K19, R33 and K304) and C (K18 and K114) (Figure 5; residue number corresponding to the human arrestin-1). V2Rpp and pRho β-strands are located between arrestin pockets B and C. However, in a recent X-ray atomic structure showing a synthetic fully phosphorylated peptide of CXCR7 receptor (C7pp) in interaction with arrestin-3 (Min *et al*, 2020), the C-terminal domain of CXCR7 adopted an elongated loop that, in addition to pockets A and B, interacts with a new pocket P (K154 and K170, Figure 5). Morevover, Yu and its collaborators showed that the interaction of phosphate of V2R C-terminus with specific pocket of arrestin affect the interaction surfaces of arrestin with its cytosolic

partners (He et al, 2021). Thus, it is possible that depending on the conformation of the GPCR C-terminal domain, phosphates do not interact with the same pockets of arrestin, leading to specific arrestin function. This folding could also be used to select a particular arrestin subtype.

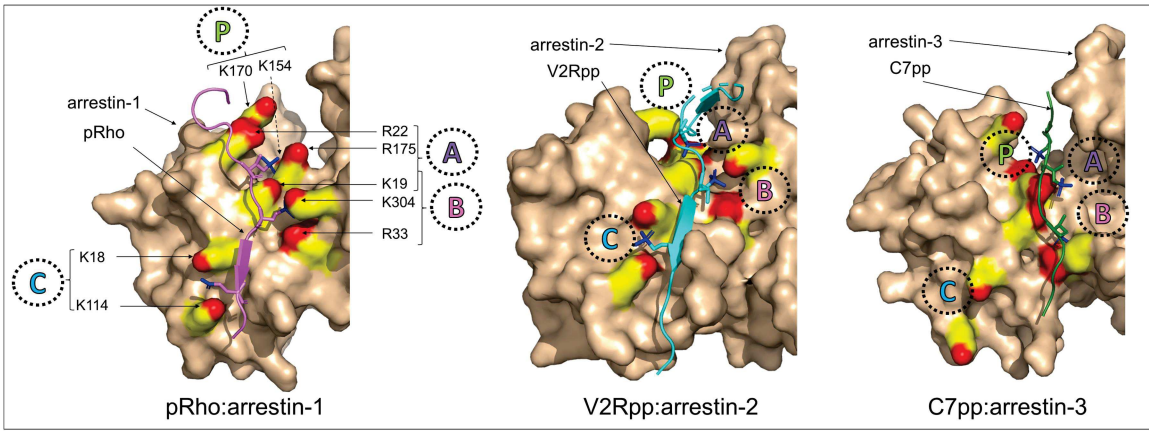


Figure 5: Atomic structures of GPCR:arrestin complex. For the rhodopsin:arrestin-1 complex (Zhou *et al*, 2017), the C-terminal domain was phosphorylated during its production in insect cells (pRho); while in the two other complexes, V2R and CXCR7, the C-terminal domain correspond to synthetic fully phosphorylated peptides (respectively V2Rpp and C7pp) (Shukla *et al*, 2013; Min *et al*, 2020). V2pp and pRho interact with pockets A (K19, R22 and R175), B (K19, R33 and K304) and C (K18 and K114) while C7pp interacts with pockets P (K154 and K170), A and B. These pockets are conserved among arrestins and residue numbers correspond to the sequence of human arrestin-1. K154 is localized in a crest behind R175 and is not visible in the figure (dashed arrow).

Binding regions identified in our study contained predicted phosphorylation motif (or phospho-code) specific to GPCRs (PhosCoFinder (Zhou *et al*, 2017) and Figure 3). These phosphorylation motifs correspond to a long p[S/T]XXp[S/T]XXp[S/T]/E/D motif or a short p[S/T]Xp[S/T]XXp[S/T]/E/D motif, where p[S/T] represents a phospho-serine or phospho-threonine, and X any amino acid residue, except proline in the second XX occurrence. It was proposed that these motifs are involved in a high-affinity arrestin binding thanks to the presence of three negatively charged residues (full motif). Conversely, the absence of one of the three charges (partial motif) reduces the interaction. The vasopressin V2 receptor and the ghrelin receptor encompassed full motif, while the β2-adrenergic receptor displayed two clusters of partial motifs (Figure 3). This was consistent with our NMR observations where V2R-Cter encompassed the largest chemical shift perturbation and binding region, whereas β2AR-Cter presented the lowest arrestin binding affinity and the shortest binding region. Low affinities were previously observed between phospho-peptides of V2R-Cter and arrestin (He *et al*, 2021). They used an arrestin-2 variant containing an unnatural fluorescent amino acid named Cou (for L-(7-hydroxycoumarin-4-yl) ethylglycine) and a fully synthetic phospho-peptide of V2R (V2Rpp) to monitor interaction by FRET (He *et al*, 2021). They extracted a dissociation constant of approximately 1 µM in the presence of Fab30 antibody, and they showed that the lack of one phosphate could reduce 17 times the affinity. Thus, in our study, the absence of antibody combined with the use of phosphomimetic variants might explain the ~ 200 µM affinity observed between pm-GPCR-Cters and arrestin-2. Moreover, the entire β2AR receptor phosphorylated by GRK2, which is known to have a lower affinity than V2R toward arrestin (Oakley *et al*, 2000), showed a dissociation constant of 0.14 nM

towards arrestin-2 using radioactivity assay (Gurevich *et al*, 1995). This exemplified the lower affinity of arrestins for peptides than receptors.

It has been already shown in IDPs that multisite phosphorylation was also proposed to modulate their affinity toward partners (Van Roey *et al*, 2012). For example, successive phosphorylation events of the p53 transcriptional factor increase its recruitment toward the CREB-binding protein (Lee *et al*, 2010). Thus, it has been proposed that multisite phosphorylation acts as a rheostat to enhance binding of IDRs/IDPs to their partner. A similar binding mode could occur between GPCR-Cters and arrestin where multivalency of phosphates would enhance the affinity of GPCR to arrestin and stabilized the interaction. This would be in agreement with the p[S/T](X)Xp[S/T]X(X)p[S/T]/E/D motif where it has been proposed that three phosphates are required to have high affinity binding to arrestin (Zhou *et al*, 2017). This would also be in agreement with the work of Yu and coll. showing that the depletion of only one phosphate, in the synthetic phosphorylation peptide of V2R (V2Rpp), reduces the affinity of V2R to arrestin (He *et al*, 2021).

In order to describe the general principles of GPCR:arrestin interaction, we worked on two different classes of arrestin:GPCR complexes. We designed phosphomimetic variants of three C-terminal domains by substituting the potential phosphorylation sites of GRK2/3 and GRK5/6 by glutamic acids. Our study highlights a conformational transition between the basal non-phosphorylated state and the phosphomimetic state from a helical or random coil to a β-extended conformation. We showed that this major conformational change occurred in the binding region with arrestin-2. Moreover, no additional conformational changes were observed in the bound state of the variant of V2R-Cter, compared to its free state, suggesting a mode of interaction by conformational selection. Based on this evidence, we propose that GRK phosphorylation may dictate the final configuration of GPCR C-terminal domains to trigger a conformational selection of arrestin, by placing the charged residues in an optimal position for the recognition. The selected arrestin conformation results from the interaction of GPCR phosphate with specific pockets of arrestin. It will expose specific interaction surfaces recognized by differents cytosolic partners, and thus, modulate arrestin function.

The same approach with different phosphorylation pattern (only GRK2/3 or GRK5/6 phosphorylation sites) would be very important in order to understand the specificity of the interaction between the different configurational states of GPCR C-termini and arrestin.

MATERIAL AND METHODS

Expressions and purifications

Recombinant human GPCR-Cter. Phosphomimetic GPCR C-termini genes were purchased from GeneArt® gene synthesis (life technologies). They were subcloned into pETM33 vector for GHSR-Cter and V2R-Cter and PET1a vector for β2AR-Cter. The constructs were fused with a $(his)_6$-GST N-terminal tag. This tag is followed by the HRV 3C protease recognition site (for GHSR-Cter and V2R-Cter) or by the TEV protease recognition site (for β2AR-Cter). The plasmids were transformed in *E. coli* BL21 (DE3). Cells were grown overnight at 25 °C in N5052 auto-induced medium supplemented with 50 µg/mL kanamycin for ^{15}N labelling or in M9 auto-induced medium supplemented with 50 µg/mL

kanamycin for $^{15}N/^{13}C$ labelling. Cells were harvested by centrifugation at 6 000 g at 10 °C during 20 min. The pellet was resuspended in Tris-HCl 20 mM pH 7.5, NaCl 300 mM and 2 mM Dithiothreitol (DTT) (buffer A) supplemented with a *Complete*® tablet (Roche). Cells were lysed by sonication, insoluble proteins and cell debris were harvested by centrifugation at 47 000 g at 10 °C during 30 min. Supernatant was clarified with 0.45 µm filter and supplemented with 10 mM imidazole and loaded onto 5 mL HisTrap FF column (GE Healthcare) equilibrated with buffer A. The column was washed with 30 volume columns with buffer A. Proteins were eluted by using gradient from 0 to 100 % of buffer B (buffer A supplemented with 500 mM imidazole). The peak fractions were analyzed by SDS-PAGE (Sodium Dodecyl Sulfate-PolyAcrylamide Gel Electroporesis). Fractions containing tagged proteins were pooled and dialyzed overnight at 4 °C against buffer A and incubated with 3C protease (ratio protease:protein 1:100 (w/w)) or with TEV protease (ratio protease:protein 1:50 (w/w)). The cleaved proteins were loaded again on HisTrap FF column, washed with buffer A and the flow through was collected. The latter was pooled, concentrated, and loaded into Superdex 75 (HiLoad 16/60, GE Healthcare) pre-equilibrated into buffer C (50 mM Bis-Tris pH 6.7, 50 mM NaCl and 2 mM DTT). The eluted protein was pooled, concentrated, flash-freezed and kept at - 80 °C in 50 mM Bis-Tris pH 6.7, 50 mM NaCl, 1 mM TCEP (Tris(2-CarboxyEthyl)Phosphine) and 0.5 mM EDTA (EthyleneDiamineTetraAcetic acid)). Concentration was determined using absorbance at 280 nm for GHSR-Cter and β2AR-Cter or using refractometry for V2R-Cter.

Recombinant rat arrestin-2. Wild type rat arrestin-2 sequence was subcloned into a pETM11 vector with a histidine-tag and a TEV cleavage site at its N-termini. This construction was confirmed by DNA sequencing and transformed into *E. Coli* strain BL21(DE3) (Fisher). Briefly, cells grown at 37 °C in Loria Broth media supplemented with 50 µg/mL of kanamycin. After 2 - 3 h, overexpression was induced with 0.25 mM IPTG at 20 °C over-night. The cells were resuspended in ice, in a lysis buffer containing 50 mM Tris pH 8, 200 mM NaCl, 0.5 mM PMSF (PhenylMethylSulfonyl Fluoride), 2 mM DTT, 1 mg/mL lysosyme and a *Complete*® tablet (Roche). The cells were sonicated and then centrifuged at 47 000 g for 30 min at 10 °C. The supernatant was clarified with 0.45 µm filter and injected onto 5 mL HisTrap HP column (GE Healthcare). The column was washed with buffer A (50 mM Tris pH 8, 200 mM NaCl and 2 mM DTT). Using step gradient, the protein was eluted at 25 % of buffer A supplemented with 500 mM imidazole. To cleave the tag, fractions containing the protein were analysed by SDS-PAGE, pooled, dialyzed against buffer A and incubated with TEV protease (ratio protease:protein 1:50 (w/w)) over-night and at 4 °C. To recover the protein cleaved, the pool was loaded in HisTrap HP column, washed with buffer A and the flow through was collected. The latter were pooled and supplemented with 20 mM Tris pH 8 and 2 mM DTT to dilute the sample at 50 mM NaCl, then it was injected into HiTrap Q HP (GE Healthcare). The protein was eluted using a gradient of 5 to 50 % buffer B (20 mM Tris pH 8, 1 M NaCl and 2 mM DTT). The fractions containing arrestin-2 were pooled, concentrated, and loaded into Superdex 200 (HiLoad 16/60, GE Healthcare) equilibrated in buffer C (50 mM Bis-Tris pH 6.7, 50 mM NaCl and 2 mM DTT). The eluted protein was pooled, concentrated, flash-freezed and kept at - 80 °C. Concentration was determined using absorbance at 280 nm.

Biophysical characterization

Size Exclusion Chromatography – Multi-Angle Light Scattering (SEC-MALS). The experiments were performed at 25 °C using a Superdex 75 10/300 GL column (GE HealthCare) connected to a miniDAWN-TREOS light scattering detector and an Optilab T-rEX differential refractive index detector (Wyatt Technology, Santa Barbara, CA). The column was equilibrated in 50 mM BisTris pH 6.7, 50 mM NaCl, 1 mM TCEP and 0.5 mM EDTA buffer filtered at 0.1 µM, and the SEC-MALS system was calibrated with a sample of Bovine Serum Albumin (BSA) at 1 mg/mL. Samples of phosphomimetic variants were prepared at 1.16 mM, 1.30 mM and 0.35 mM for V2R-Cter, GHSR-Cter and β2AR-Cter, respectively. For each GPCR-Cter, 40 µL of sample were injected at 0.5 mL/min. Data acquisition and analyses were performed using the ASTRA software (Wyatt).

Circular Dichroism (CD) spectroscopy. Far UV-spectra of the three pm-GPCR-Cters were recorded in a quartz cuvette (path length 0.1 cm) at 0.08 mg/mL in H_2O at 20 °C using a Chirascan. The ellipticity was scanned from 190 to 260 nm with an increment of 0.5 nm, an integration time of 3 s, and a constant band-pass of 1 nm. Data were treated using Chirascan and, after substraction of the buffer signal, were converted to mean residue ellipticity ($[\theta]_{MRW}$, mdeg.cm².dmole^{-1}) using equation (1) (Chemes *et al*, 2012):

$$(1) \quad [\theta]_{MWR} = [(\theta * Mw)/(L * C * 10)]/(n-1)$$

where θ is the ellipticity (mdeg), Mw is the molecular weight (gmol), L is the cuve lengh (cm), C is the protein concentration (mg/mL) and n is the number of peptid bond.

Small-angle X-ray scattering (SAXS) measurement and analysis. Synchrotron radiation small-angle X-ray scattering (SAXS) data were acquired for pm-GPCR-Cters at the SWING beamline at the SOLEIL synchrotron (Saint-Aubin, France) (Bizien *et al*, 2016) using an X-ray wavelength of 1.03 Å and a sample-to-detector distance of 1.99 m. Samples were measured at 15 °C and at two concentrations, 5 mg/mL and 10 mg/mL, for all GPCR-Cters, in 50 mM BisTris pH 6.7, 50 mM NaCl and 2 mM DTT buffer. Before exposure to X-rays, 45 µl of sample were injected into 3 mL Superdex 75 5/150 GL column (GE HealthCare) at 0.2 mL/min, pre-equilibrated into the same buffer as samples. The intensity was measured as function of the magnitude of the scattering vector, s, using equation (2) (Blanchet & Svergun, 2013):

$$(2) \quad s = 4\pi.\sin(\theta)/\lambda$$

where θ the scattering angle and λ the X-ray wavelength.

The scattering patterns of the buffer were recorded before the void volume of the column (1 mL). The scattering profiles measured covered a momentum transfer range of $0.002 < s < 0.5$ Å^{-1}. Data processing (buffer subtraction) was done automatically with the software available at the beamline.

Nuclear Magnetic Resonance (NMR) spectroscopy

All NMR experiments were performed on Bruker Avance III 700 MHz or 800 MHz spectrometers, equipped with a cryogenic triple-resonance (^{1}H, ^{15}N, ^{13}C) probe and shielded z-gradients. The sample temperature was set to 20 °C and the buffer for all samples (named NMR buffer) is composed of 50 mM Bis-Tris pH 6.7, 150 mM NaCl, 1 mM EDTA, 0.5 mM TCEP, 5 % D2O (Eurisotop) and 5 mM DSS-

d6 (2,2-dimethyl-2-silapentane-5-sulfonate, Sigma) as internal reference (Wishart & Sykes, 1994b). All experiments used the pulse sequences provided by Bruker Topspin 3.2. Squared cosine apodization was used in indirect dimensions, prior to zero-filling and Fourier transformation using TOPSPIN (version 4.0.6, Bruker) and data processing was performed using NMRFAM-SPARKY (version 1.414, (Lee *et al*, 2015)). For each NMR experiments, except CSP, concentrations of pm-GPCR-Cters were indicated in Table S1. For all NMR experiments, data were measured for all residues of C-terminus regions excepted proline residues, the residue A339 of β2AR-Cter and the first N-terminal residue.

Backbone assignment. For the sequential backbone assignment of the $^{13}C/^{15}N$ labeled pm-V2R-Cter, pm-GHSR-Cter and pm-Cβ2AR-Cter, HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCA(CO)NH and HNCACB triple resonance 3D experiments were recorded at 700 MHz and 20 °C in NMR buffer with 150 mM NaCl. H_N, N, CO, C_α and C_β nuclei of all residues were assigned, expected the first glycine N-terminal residue, the residue A339 of β2AR-Cter and proline residues.

Chemical Shift Index (CSI). $^{13}C_\alpha$ and $^{13}C_\beta$ chemical shifts were used to calculate CSI by subtraction of experimental chemical shifts (from the 3D experiment) from random-coil chemical shift calculated by POTENCI database (Tamiola *et al*, 2010; Nielsen & Mulder, 2018). In addition to proline residues, the first N-terminal residue and the residue A339 of β2AR-Cter, CSI were not computed for the last residue. For CSI perturbation upon interaction, HNCACB triple resonance 3D experiments were recorded on a sample of 50 µM $^{13}C/^{15}N$ labeled pm-V2R-Cter free and in presence of 250 µM of arrestin-2, at 800 MHz, 20 °C, and in a NMR buffer with 50 mM NaCl.

$^3J_{HNHA}$ scalar coupling. $^3J_{HNHA}$ scalar coupling measurements were obtain accordingly to Vuister and Bax (Vuister & Bax, 1993). Briefly, HNHA experiments were recorded at 800 MHz in NMR buffer with 150 mM NaCl on ^{15}N-labelled GPCR C-termini. Intensity of the cross-peak (S_{cross}) and intensity of the corresponding diagonal peak (S_{diag}) were extracted using Sparky. They were used to calculate the $^3J_{HNHA}$ scalar coupling of each amino acid using the equation (3):

$$(3) \quad S_{cross}/S_{diag} = -tan^2\,(2\pi^3\, J_HNHA\, \xi)$$

Where 2ξ is the total evolution time for the homonuclear $^3J_{HNHA}$ coupling, which has been set to 26.1 ms.

^{15}N NMR relaxation experiments. Experiments were performed at 700 MHz in NMR buffer containing 150 mM NaCl on ^{15}N-labelled GPCR C-termini. Heteronuclear $^{15}N\{^1H\}$-NOE values were determined from two experiments with on- (saturated spectrum) and off-resonance ^1H saturation (unsaturated spectrum) which were recorded in an interleaved manner. The saturation time by 120 ° pulses (~ 10 kHz) was set to 6 s and the recycle delay to 6 s. NOEs values were obtained from the ratio of intensities measured in the saturated (I) and unsaturated (I_0) spectra. Longitudinal (R_1) and transversal (R_2) relaxation rate were measured through acquisition of ^{15}N-HSQC spectra with different relaxation delays: 10, 50, 100, 200, 400, 600, 800, 1 000 ms for R_1, and 16, 32, 64, 96, 160, 240, 480, 640 ms for R_2. For each pick, intensity was fitted to a single exponential decay using Sparky (Lee *et al*, 2015) to obtain the relaxation parameters. For all relaxation parameters, three residues at N- and C-extremity were discarded from the calculation of average values due to their inherent higher flexibility.

^1H-^{15}N Residual Dipolar Coupling (RDC). Residual dipolar couplings were recorded at 700 MHz in NMR buffer containing 150 mM NaCl on ^{15}N-labelled GPCR C-termini. The samples were dissolved in

a liquid-crystalline medium composed by 5 % (w/v) mixture of 0.85 molar ratio of polyoxyethylene 5-lauryl ether (PEG/$C_{12}E_5$) (Sigma) and 1-hexanol (Sigma) (Rückert & Otting, 2000) or by adding ~ 20 mg/mL of filamentous phage Pf1 (Asla biotech) (Hansen *et al*, 1998). The $^1D_{NH}$ coupling constants were obtained from the relative peak positions in the ^{15}N dimension measured in the *sum* and *diff* subspectra (upfield and downfield components of the doublet). Flexible Meccano (FM) (Ozenne et al, 2012) was used to build biased ensemble of 50 000 conformations or random-coil ensemble of 200 000 conformations from which RDCs were back-calculated.

Paramagnetic Relaxation Enhancement (PRE). The C378A or C406A variants of ^{15}N-pm-β2AR-Cter were labeled on the remaining cysteine using 3-(2-Iodoacetamodi)-proxyl (Merck). Paramagnetic samples were recorded at 700 MHz in NMR buffer containing 150 mM NaCl with a recycling delay of 2 s. Reference diamagnetic samples were recorded in the same condition after addition of 5 mM fresh ascorbic acid, pH 6.7, in the NMR tube. PRE were analyzed by measuring the peak intensity ratios (I_{para}/I_{dia}) between two ^{15}N-HSQC spectra of paramagnetic and diamagnetic samples. The theoretical profile as in a strictly random coil polymer was calculated according to (Teilum *et al*, 2002).

Chemical Shift Perturbation (CSP). The interaction between ^{15}N-labelled GPCR C-termini and the unlabelled arrestin-2 was measured in NMR buffer containing 50 mM NaCl. ^{15}N-HSQC spectra were recorded on the 800 MHz spectrometer at 1:0, 1:1, 1:2, 1:5 and 1:10 ^{15}N-GPCR-Cter:arrestin-2 ratio. At each point of the titration and before to record the spectrum, the complex is incubated 30 min at room temperature. For each peak in ^{15}N-HSQC spectra, the intensity ratio (I/I_{REF}) is calculated between the bound (I) and free states (I_{REF}). Means of intensity ratio subtracted with the mean of uncertainty value was chosen as cut-off values, by excluding residues in interaction. Chemical shift perturbation Δδ were calculated using equation (4) (Grzesiek *et al*, 1996; Williamson, 2013):

$$(4) \quad \Delta\delta = \{0.5[(\Delta H)^2 + (\Delta N/5)^2]\}^{1/2}$$

Where Δδ is the chemical shift perturbation (in ppm), ΔH is the subtraction of ^{1}H chemical shift in the bound and free state and ΔN is the subtraction of ^{15}N chemical shift in the bound and free state. The overlapped residues in HSQC spectra were removed for the chemical shift perturbation calculation. Standard deviation multiplied by 4 was chosen as cut-off values (Williamson, 2013). Residues in interaction were removed from standard deviation calculation.

All interactions were performed at 50 mM NaCl instead of 150 mM in order to limit the screening of potential electrostatic interactions. As the C-terminal domains were characterized at 150 mM NaCl, we checked that the salt did not affect the conformation of GPCR-Cters alone. The effect of salt was estimated by measuring chemical shift perturbation using equation (4) between two ^{15}N-HSQC at 150 and 50 mM NaCl (Figure S13A). Spectra of 100 µM of ^{15}N-GPCR-Cter in NMR buffer with 150 and 50 mM NaCl were recorded on the 700 MHz spectrometer. 3D NMR experiments on pm-V2R-Cter at 150 and 50 mM of NaCl were used to highlight the salt effect on $^{13}C_\alpha/C_\beta$ chemical shift (Figure S13B) by computing CSI.

Dissociation constant estimation

Microscale thermophoresis (MST). Arrestin-2 was labeled with Alexa Fluor® 488 Maleimide C5 (ThermoFisher) (ARR*) to monitor the interaction. The sample (1 mg/mL) was buffer-exchanged by using PD MiniTrap^{TM} G-10 (GE Healthcare), in labeling buffer composed of 50 mM Bis-Tris pH 7.5, 50 mM NaCl and 1 mM TCEP. This step allows to remove dithiothreitol (DTT) and exchanges the pH. Then 10 molar equivalents of dye stock dissolved in dimethylsulfoxide (DMSO) were added into the reaction mixture. After an incubation of 1 h on ice protected from the light, the reaction was quenched by adding 100 mM of L-cysteine and incubates during 15 min at 20 °C. To remove excess of dye, the sample was buffer-exchanged using PD MiniTrap^{TM} G-10 (GE Healthcare) into 50 mM HEPES pH 7.5, 50 mM NaCl. The sample concentration was determined by absorbance at 280 nm for ARR* and at 495 nm for the dye alone. An extinction coefficient of 71 000 $M^{-1}.cm^{-1}$ and the correction factor (contribution of the dye to the absorbance at 280 nm) of 0.1470 were assumed for calculating the concentration. Before each measurement, all samples were centrifuged at 12 000 g during 15 min. Measurement of fluorescence was performed using the Monolith NT.115 MST instrument with standard MST capillaries (NanoTemper Technologies). To monitor the interaction, a range excitation wavelength between 460 and 480 nm and a range emission wavelength measured between 515 and 530 nm were used. The fluorescence was measured after the incubation of 5 nM of ARR* with varying amounts of phosphomimetic GHSR-Cter (from 0.01 to 450 µM) at 20 °C for 30 min. Dose-response curve was plotted using the MO.Affinity Analysis software (Nanotemper) and outliers values were removed. One-site binding model was used to extract the dissociation constant.

Fluorescence anisotropy. To monitor the interaction, pm-GHSR-Cter was labeled with Alexa Fluor® 488 succinimidyl ester (ThermoFisher) specifically on the first amine group (pm-GHSR-Cter*). Sample (100 µL at 50 µM) was buffer-exchanged by using PD MiniTrap^{TM} G-10 (GE Healthcare), in labeling buffer composed of 50 mM Bis-Tris pH 7, 50 mM NaCl and 1 mM TCEP. This step allows to remove DTT and to exchange the pH. Then two molar equivalents of dye stock dissolved in DMSO were added into reaction mixture. After an incubation of 1 h at 20 °C protected from the light and under agitation, the reaction was stopped by adding 10 µL of 1 M Tris and incubated for 15 min. To remove excess of dye, the sample was buffer-exchanged using PD MiniTrap^{TM} G-10 (GE Healthcare) into 50 mM Bis-Tris pH 6.7, 50 mM NaCl and 1 mM TCEP. The sample concentration was determined by absorbance at 280 nm for pm-GHSR-Cter* and at 495 nm for the dye alone. An extinction coefficient of 71 000 $M^{-1}.cm^{-1}$ and a correction factor of 0.11 were assumed for calculating the concentration and the degree of labeling. Before each measurement, all sample were centrifuged at 12 000 g during 15 min. Measurement of fluorescence anisotropy was performed using a Tecan Safire II micro plate reader fluorescence spectrometer and a Corning 384 Low Flange Black Flat Bottom plate. An excitation wavelength of 470 nm was used and emission was measured at 530 nm. 1 nM of pm-GHSR-Cter* were incubated with varying amounts of arrestin-2 in an buffer containing 50 mM Bis-Tris pH 6.7, 50 mM NaCl and 1 mM TCEP at 20 °C for 30 minutes. The fluorescence anisotropy was measured for varying amounts of arrestin-2 (0 to 670 µM). Affinitie (k_D) value was calculated by nonlinear regression analysis using a sigmoid curve fit (GraphPad Prism, GraphPad Software).

ACCESSION NUMBERS

The ^{1}H, ^{15}N and ^{13}C assignment of variants of V2R-Cter, GHSR-Cter and β2AR-Cter are available in the BioMagResBank (http://www.bmrb.wisc.edu) under the accession number BMRB (in progress).

SUPPLEMENTARY DATA

CREDIT AUTHOR STATEMENT

Myriam Guillien: Investigation, Methodology, Writing - Original Draft. **XXX** Investigation, Writing - Review & Editing. **XXX**: Investigation, Resources. **XXX**: Software, Formal Analysis. **XXX**: Investigation, Resources. **XXX**: Conceptualization, Resources. **XXX**: Software, Formal Analysis. **CCC**: Investigation, Writing - Review & Editing. **XXX**: Conceptualization, Writing - Review & Editing. **Nathalie Sibille**: Conceptualization, Investigation, Supervision, Writing - Original Draft, Review & Editing.

FUNDING

This work has benefited from support by the Labex EpiGenMed, an « Investissements d'avenir » program, reference ANR-10-LABX-12-01awarded to MG. N.S. and J.L.B. acknowledges the support of the ANR GPCteR (ANR-17- CE11-0022-01). The CBS is a member of the French Infrastructure for Integrated Structural Biology (FRISBI) supported by the French National Research Agency (ANR- 10-INSB-05).

ACKNOWLEDMENTS

We thank the synchrotron facility SOLEIL (St. Aubin) for allocating regular beam time (Block Allocation Group, proposal No. 20201085) and its dedicated staff for technical help with the beamline SWING. The authors are grateful to David Lemaire for MS support.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

IDP, intrinsically disordered protein; IDR, intrinsically disordered region; PTM, post-translational

modifications; GPCR, G protein-coupled receptor; GRK, G protein receptor kinase; V2R, vasopressin V2 receptor; GHSR, ghrelin receptor 1a; β2AR, β2-adreneric receptor; wt-V2R-Cter, wild type V2R C-terminus; pm-V2R-Cter, phosphomimetic V2R C-terminus; wt-GHSR-Cter, wild type GHSR C-terminus; pm-GHSR-Cter, phosphomimetic GHSR C-terminus; wt-β2AR-Cter, wild type β2AR C-terminus; pm-β2AR-Cter, phosphomimetic β2AR C-terminus; V2Rpp, synthetic fully phosphorylated peptide of V2R C-terminus; C7pp, synthetic fully phosphorylated peptide of CXCR7 C-terminus; pRho-Cter, phosphorylated C-terminus of rhodopsin receptor; PPII, polyproline helix 2; CD, circular dichroism; MS, mass spectrometry; MST, microscale thermophoresis; FRET, fluorescence resonance energy transfert; SAXS, small angle X-ray scattering; SDS-PAGE, sodium dodecyl sulphate-polyacrylamide gel electrophoresis; SEC-MALS, size exclusion chromatography – multi angle light scattering; NMR, nuclear magnetic resonance; CSI, chemical shift index; CSP chemical shift perturbation; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser effect; PRE, paramagnetic relaxation enhancement; RDC, residual dipolar coupling; FM, flexible meccano; EDTA, EthyleneDiamineTetraAcetic acid; DTT, dithiothreitol; Cou, L-(7-hydroxycoumarin-4-yl ethylglycine; PMSF, phenylmethylsulfonyl fluoride; TCEP, Tris(2-CarboxyEthyl)Phosphine.

REFERENCES

Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, Sonenberg N, Kay LE & Forman-Kay JD (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature 519: 106–109

Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH & Blackledge M (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. PNAS 102: 17002–17007

Bizien T, Durand D, Roblina P, Thureau A, Vachette P & Pérez J (2016) A Brief Survey of State-of-the-Art BioSAXS. Protein Pept Lett 23: 217–231

Blanchet CE & Svergun DI (2013) Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. Annu Rev Phys Chem 64: 37–54

Bouzo-Lorenzo M, Santo-Zas I, Lodeiro M, Nogueiras R, Casanueva FF, Castro M, Pazos Y, Tobin AB, Butcher AJ & Camiña JP (2016) Distinct phosphorylation sites on the ghrelin receptor, GHSR1a, establish a code that determines the functions of ß-arrestins. Scientific Reports 6: 22495

Chemes LB, Alonso LG, Noval MG & de Prat-Gay G (2012) Circular Dichroism Techniques for the Analysis of Intrinsically Disordered Proteins and Domains. In Intrinsically Disordered Protein Analysis: Volume 1, Methods and Experimental Tools, Uversky VN & Dunker AK (eds) pp 387–404. Totowa, NJ: Humana Press

Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017a) Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. Adv Exp Med Biol 1009: 107–129

Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017b) Small-angle scattering studies of intrinsically disordered proteins and their complexes. Curr Opin Struct Biol 42: 15–23

Grzesiek S, Bax A, Clore GM, Gronenborn AM, Hu J-S, Kaufman J, Palmer I, Stahl SJ & Wingfield PT (1996) The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. Nat Struct Mol Biol 3: 340–345

Gurevich VV, Dion SB, Onorato JJ, Ptasienski J, Kim CM, Sterne-Marr R, Hosey MM & Benovic JL (1995) Arrestin interactions with G protein-coupled receptors. Direct binding studies of wild type and mutant arrestins with rhodopsin, beta 2-adrenergic, and m2 muscarinic cholinergic receptors. J Biol Chem 270: 720–731

Gurevich VV & Gurevich EV (2019) GPCR Signaling Regulation: The Role of GRKs and Arrestins. Front Pharmacol 10

Hansen MR, Mueller L & Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. Nat Struct Biol 5: 1065–1074

He Q-T, Xiao P, Huang S-M, Jia Y-L, Zhu Z-L, Lin J-Y, Yang F, Tao X-N, Zhao R-J, Gao F-Y, et al (2021) Structural studies of phosphorylation-dependent interactions between the V2R receptor and arrestin-2. Nat Commun 12: 2396

Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V & Skrzypek E (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res 43: D512-520

Huang W, Masureel M, Qu Q, Janetzko J, Inoue A, Kato HE, Robertson MJ, Nguyen KC, Glenn JS, Skiniotis G, et al (2020) Structure of the neurotensin receptor 1 in complex with β-arrestin 1. Nature 579: 303–308

Lee CW, Ferreon JC, Ferreon ACM, Arai M & Wright PE (2010) Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. Proc Natl Acad Sci U S A 107: 19290–19295

Lee W, Tonelli M & Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. Bioinformatics 31: 1325–1327

Min K, Yoon H-J, Park JY, Baidya M, Dwivedi-Agnihotri H, Maharana J, Chaturvedi M, Chung KY, Shukla AK & Lee HH (2020) Crystal Structure of β-Arrestin 2 in Complex with CXCR7 Phosphopeptide. Structure

Nguyen AH, Thomsen ARB, Cahill TJ, Huang R, Huang L-Y, Marcink T, Clarke OB, Heissel S, Masoudi A, Ben-Hail D, et al (2019) Structure of an endosomal signaling GPCR–G protein–β-arrestin megacomplex. Nat Struct Mol Biol: 1–9

Nielsen JT & Mulder FAA (2018) POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. J Biomol NMR 70: 141–165

Nobles KN, Guan Z, Xiao K, Oas TG & Lefkowitz RJ (2007) The active conformation of beta-arrestin1: direct evidence for the phosphate sensor in the N-domain and conformational differences in the active states of beta-arrestins1 and -2. J Biol Chem 282: 21370–21381

Nobles KN, Xiao K, Ahn S, Shukla AK, Lam CM, Rajagopal S, Strachan RT, Huang T-Y, Bressler EA, Hara MR, et al (2011) Distinct Phosphorylation Sites on the 2-Adrenergic Receptor Establish a Barcode That Encodes Differential Functions of -Arrestin. Science Signaling 4: ra51–ra51

Oakley RH, Laporte SA, Holt JA, Caron MG & Barak LS (2000) Differential Affinities of Visual Arrestin, βArrestin1, and βArrestin2 for G Protein-coupled Receptors Delineate Two Major Classes of Receptors. J Biol Chem 275: 17201–17210

Ozenne V, Bauer F, Salmon L, Huang J -r., Jensen MR, Segard S, Bernado P, Charavay C & Blackledge M (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28: 1463–1470

Rückert M & Otting G (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. Journal of the American Chemical Society 122: 7793–7797

Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig K-P & Kalbitzer HR (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. J Biomol NMR 39: 275–289

Senicourt L, Maire A le, Allemand F, Carvalho JE, Guee L, Germain P, Schubert M, Bernadó P, Bourguet W & Sibille N (2020) Structural insights into the cooperative interaction of the intrinsically disordered co-activator TIF2 with retinoic acid receptor heterodimer (RXR/RAR). bioRxiv: 2020.10.28.359091

Shukla AK, Manglik A, Kruse AC, Xiao K, Reis RI, Tseng W-C, Staus DP, Hilger D, Uysal S, Huang L-Y, et al (2013) Structure of active β-arrestin1 bound to a G protein-coupled receptor phosphopeptide. Nature 497: 137–141

Sibille N & Bernadó P (2012) Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochem Soc Trans 40: 955–962

Staus DP, Hu H, Robertson MJ, Kleinhenz ALW, Wingler LM, Capel WD, Latorraca NR, Lefkowitz RJ & Skiniotis G (2020) Structure of the M2 muscarinic receptor–β-arrestin complex in a lipid nanodisc. Nature 579: 297–302

Tamiola K, Acar B & Mulder FAA (2010) Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. J Am Chem Soc 132: 18000–18003

Teilum K, Kragelund BB & Poulsen FM (2002) Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. J Mol Biol 324: 349–357

Uversky VN (2012) Size-Exclusion Chromatography in Structural Analysis of Intrinsically Disordered Proteins. In Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools, Uversky VN & Dunker AK (eds) pp 179–194. New York, NY: Springer

Van Roey K, Gibson TJ & Davey NE (2012) Motif switches: decision-making in cell regulation. Curr Opin Struct Biol 22: 378–385

Vuister GW & Bax A (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins. Journal of the American Chemical Society 115: 7772–7777

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. Progress in Nuclear Magnetic Resonance Spectroscopy 73: 1–16

Wishart DS & Sykes BD (1994) [12] Chemical shifts as a tool for structure determination. In Methods in Enzymology pp 363–392. Elsevier

Woody RW (1996) Theory of Circular Dichroism of Proteins. In Circular Dichroism and the Conformational Analysis of Biomolecules, Fasman GD (ed) pp 25–67. Boston, MA: Springer US

Wright PE & Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16: 18–29

Zhou XE, He Y, de Waal PW, Gao X, Kang Y, Van Eps N, Yin Y, Pal K, Goswami D, White TA, et al (2017) Identification of Phosphorylation Codes for Arrestin Recruitment by G protein-Coupled Receptors. Cell 170: 457-469.e13

SUPPLEMENTARY DATA

Phospho-code dictates GPCR C-terminal domain conformation and arrestin interaction

Myriam Guillien[a,†], Assia Mouhand[a,†], Aurélie Fournet[a], Aurélien Thureau[b], (Marjorie/Michela)[c], Pau Bernadó[a], Jean-Louis Banères[c], Nathalie Sibille[a,*,1]

[a] Centre de Biochimie Structurale (CBS), CNRS, INSERM, Univ Montpellier, Montpellier, France
[b] Beamline SWING, Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91190 Gif-sur-Yvette, France
[c] IBMM, UMR5247 CNRS, Pôle Chimie Balard Recherche, 1919 route de Mende, 34293 Montpellier cedex 5, France
† These authors contributed equally to this work
* Corresponding author. Email addresses: nathalie.sibille@cbs.cnrs.fr
[1] Lead contact. Tel: +33 4 67 41 77 05; Fax: +33 4 67 41 79 13; Email: nathalie.sibille@cbs.cnrs.fr

Protein	3D	$^{3}J_{HNHA}$	Relaxation	RDC alcohol	RDC Pf1	PRE
pm-V2R	50	93	93	93	138	NT
pm-GHSR	377	263	263	263	250	NT
pm-β2AR	206	476	476	476	212	100

Table S1: Concentration used for the NMR characterization of phosphomimetic GPCR C-termini of V2R-Cter (blue), GHSR-Cter (green) and β2AR-Cter (pink). Concentration used in each experiment is indicated in µM. The letters NT mean that these experiments have not been carried out.

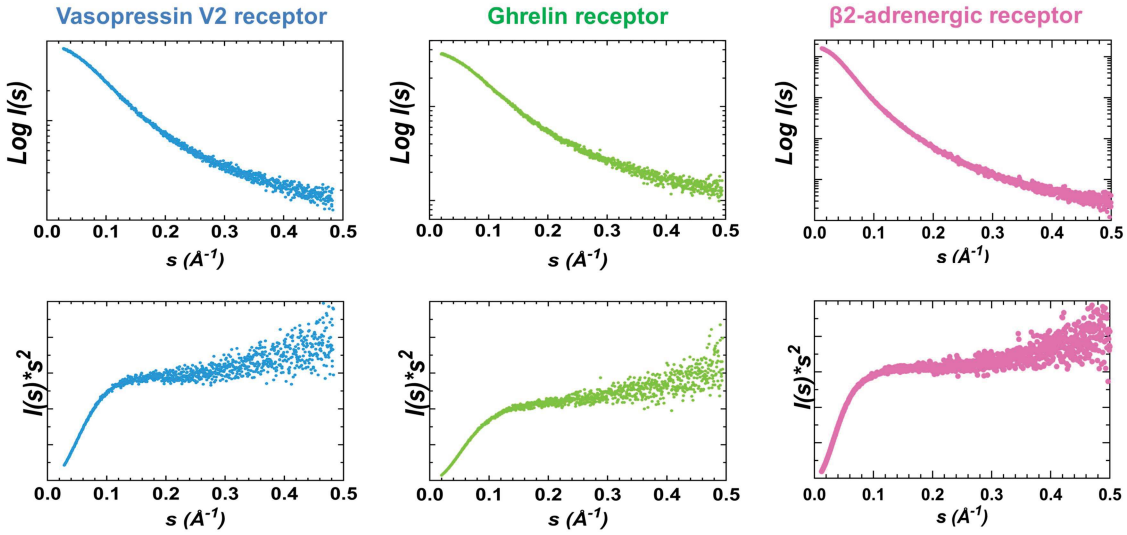


Figure S1: Related to Figure 1. SAXS data of phosphomimetic variants of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (pink, right). Semi-logarithmic representation of the experimental SAXS curves versus scattering angle (upper panel) and Kratky plots (lower panel). The Kratky plot of variants display a typical profile of IDP with no clear maximum and a monotone increase along the momentum transfer range (Sibille & Bernadó, 2012; Cordeiro *et al*, 2017a, 2017b).

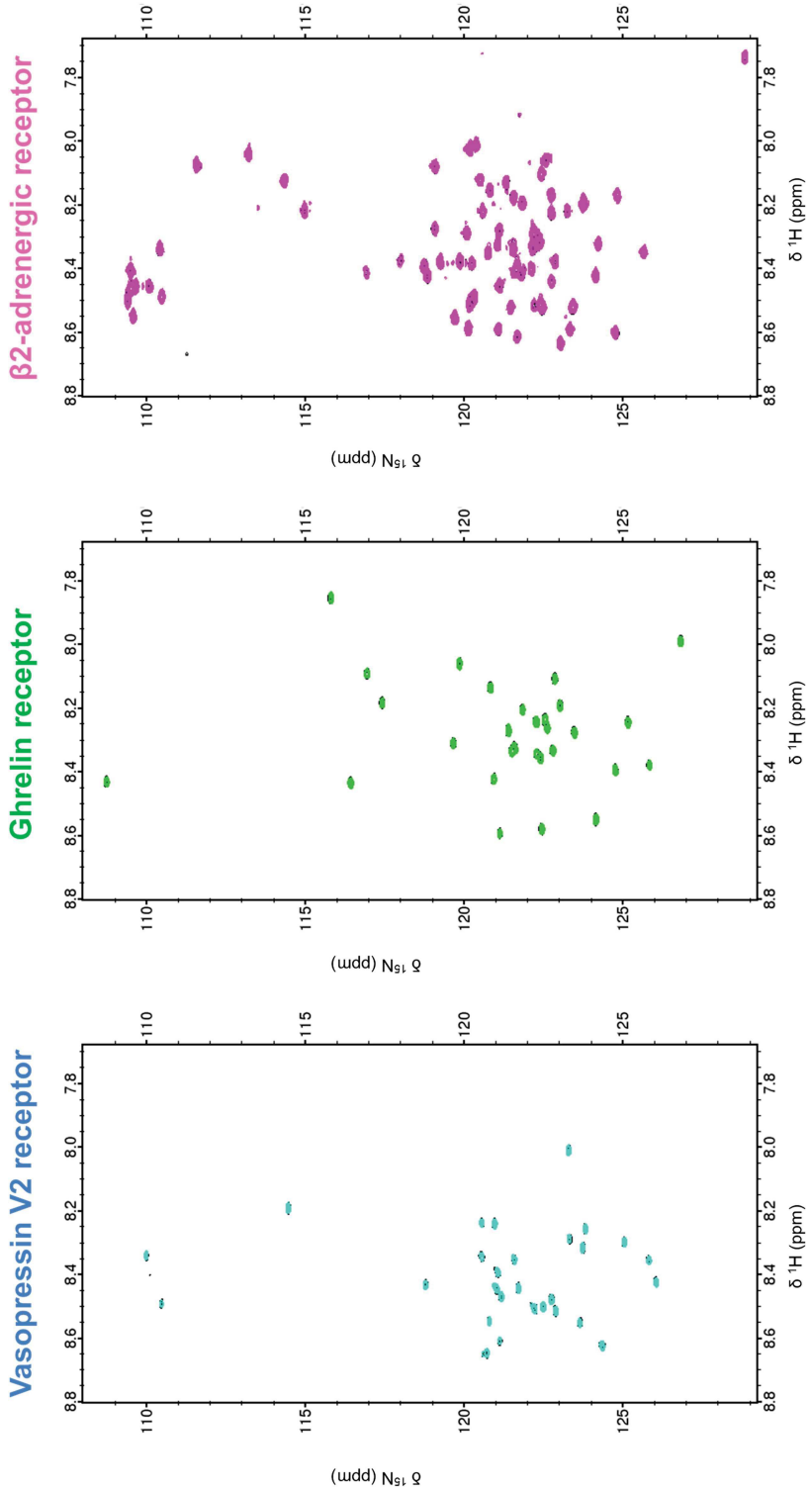


Figure S2: Related to Figure 2. The concentration of pm-GPCR-Cters doesn't affect their conformations. Overlay of ^{15}N-HSQCs recorded on samples at 300 µM (black) or 100 µM (colored) at 700 MHz, 20 °C, 50 mM Bis-Tris pH 6.7, 150 mM NaCl of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (pink, right). No chemical shift changes were observed in NMR spectra at different concentrations.

.

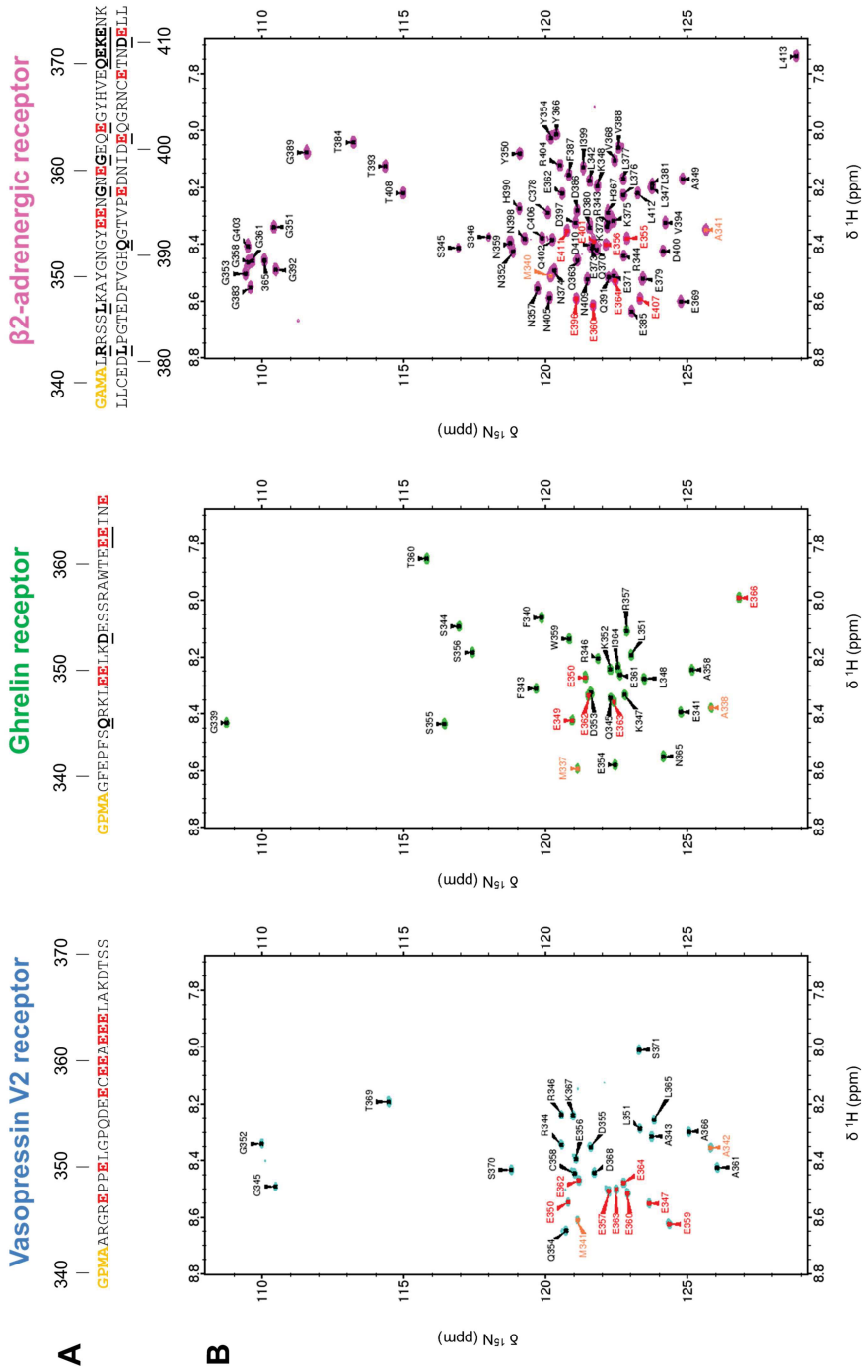


Figure S3: Related to Figure 2. Assignments of phosphomimetic variants of V2R-Cter (blue, left), GHSR-Cter (green, middle) and β2AR-Cter (pink, right). A) Sequence of the phosphomimetic C-terminal domains with known phosphorylation sites mutated by glutamic acid (E) indicated in red. Amino acids coming from the protease cleavage site are shown in orange and overlapping residues are underlined in black. **B)** ^{15}N-HSQC of the C-terminal domains were recorded on 300 µM samples at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer. Residue-specific assignments are shown on spectra and are colored according to A).

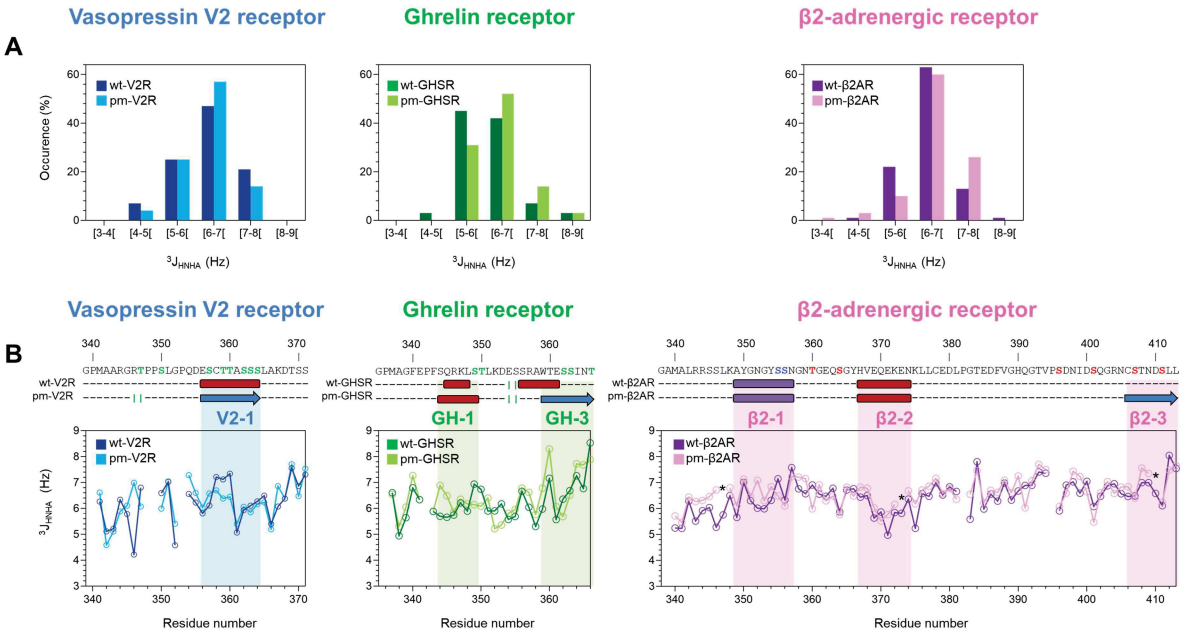


Figure S4: Related to Figure 2. Overlay of 3JHNHA scalar coupling of wild type (dark color) and variant (light color) forms of V2R-Cter (blue, left); GHSR-Cter (green, middle) and β2AR-Cter (purple/pink, right). A) Histogram of the 3JHNHA in Hz. The couplings seem to shift to higher values in the variant suggesting a lower content of helical conformation (values below 6 Hz) and a higher content of extended conformation (values above 8 Hz). **B)** On top, sequence and NMR secondary structure consensus of GPCR C-termini are shown according to figure 2. Below, the coupling values along the sequence are represented. In the variant, higher coupling values are localized in secondary structure region. For β2AR variant, outlier values for overlapping residues in HSQCs were removed (dark stars). They were recorded at 700 MHz, 20 °C, 50 mM Bis-Tris pH 6.7, 150 mM NaCl.

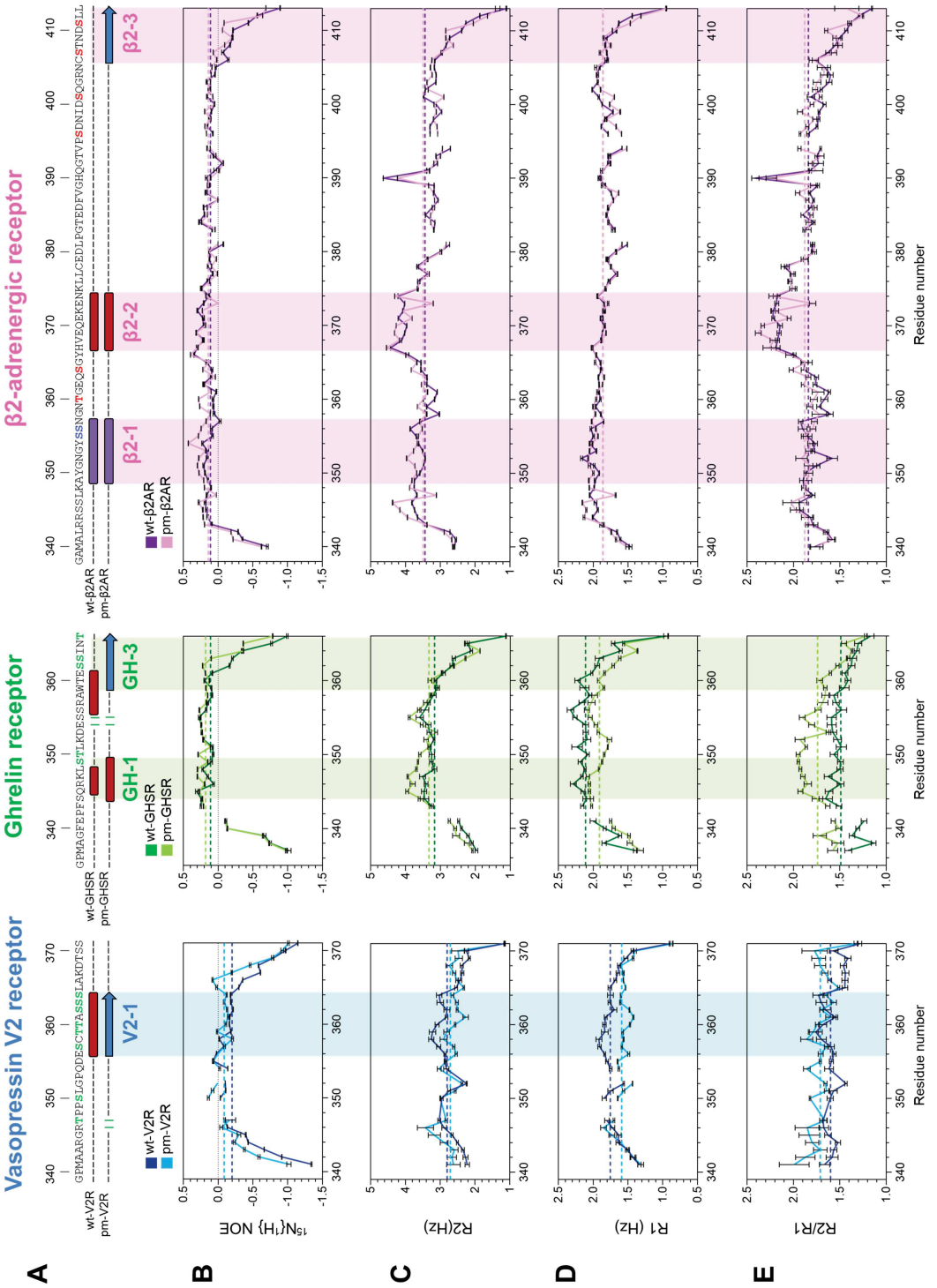


Figure S5: Related to Figure 2. Overlay of relaxation data of wild type (dark color) and variant (light color) forms of V2R-Cter (blue, left); GHSR-Cter (green, middle) and β2AR-Cter (purple/pink, right). A) Sequence and NMR secondary structure consensus of GPCR C-termini are shown according to figure 2. **B)** $^{15}N\{^{1}H\}$ heteronuclear NOE, **C)** transversal relaxation rates R2, **D)** longitudinal relaxation rates R1 and **E)** R2/R1 ratio are shown in colored line. Average values are indicted by dashed colored line. Relaxation data were recorded at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

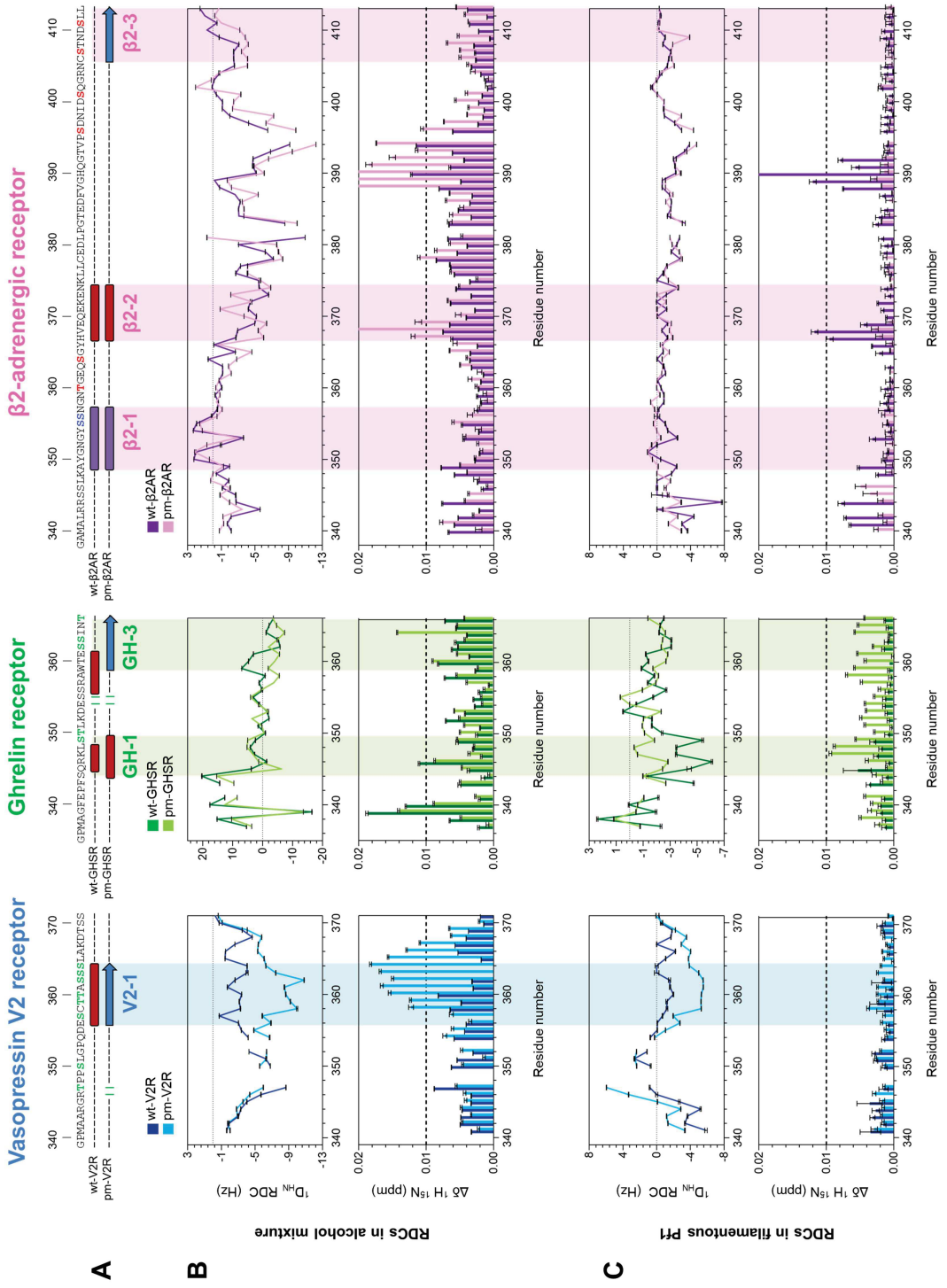


Figure S6: Related to Figure 2. Overlay of $^{1}D_{NH}$ RDCs values of wild type (dark color) and variant (light color) forms of V2R-Cter (blue, left); GHSR-Cter (green, middle) and β2AR-Cter (purple/pink, right). A) Sequence and NMR secondary structure consensus of GPCR-Cters are shown according to figure 2. RDCs data were measured in **B)** alcohol mixture or in **C)** filamentous phage Pf1. In each media, $^{1}D_{NH}$ RDCs profile (top) and chemical shift changes (bottom) are shown. RDCs were measured at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer.

.

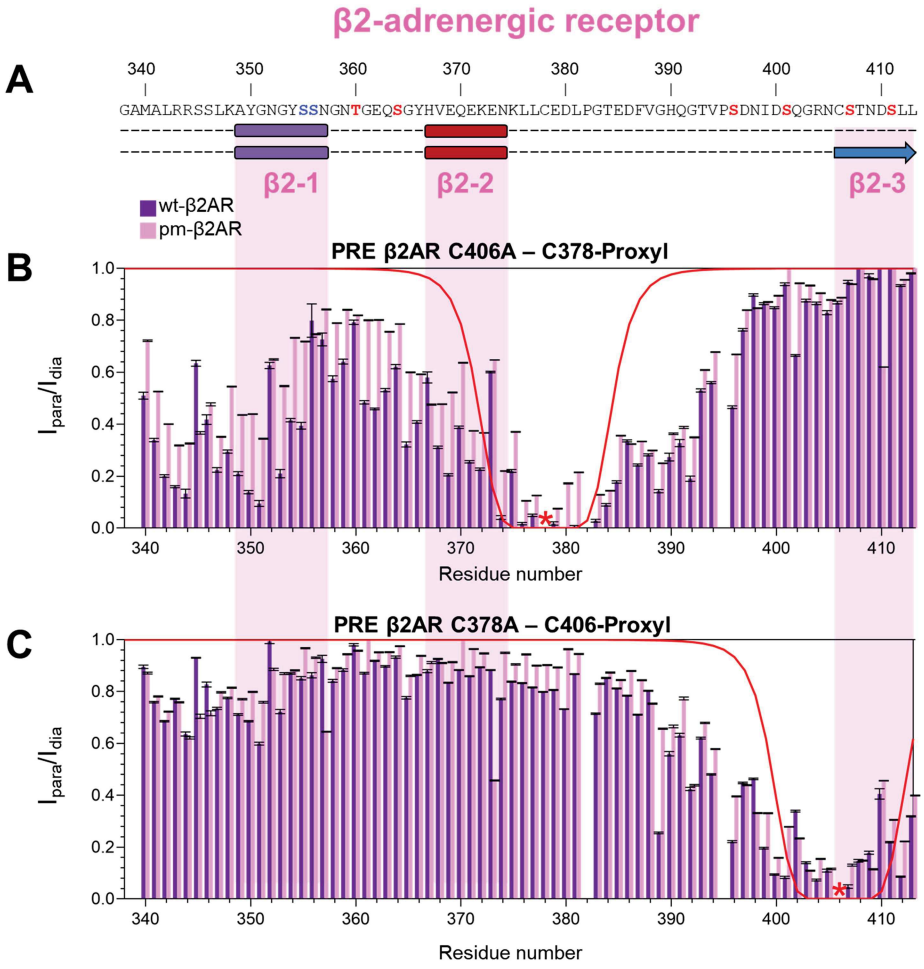


Figure S7: Related to Figure 2. Overlay of PRE data of β2AR variants in wild type and phosphomimetic forms. A) Sequence and NMR secondary structure consensus are indicated according to Figure 2. **B)** PRE data of C406A variant labeled on C378 (red star). **C)** PRE data of C378E variant labeled on C406 (red star). PREs were calculated from the ratio of intensity (Ipara/Idia) of ^{15}N-HSQC spectra measured for the paramagnetic (Ipara) sample (position of proxyl is indicated with a red star), and for the diamagnetic sample (Idia). They were recorded at 700 MHz, 20 °C, in 50 mM Bis-Tris pH 6.7, 150 mM NaCl buffer, on sample at 50 µM for β2AR wild-type and 100 µM for β2AR variant. Theoretical PRE curves for a completely disordered protein are shown in red.

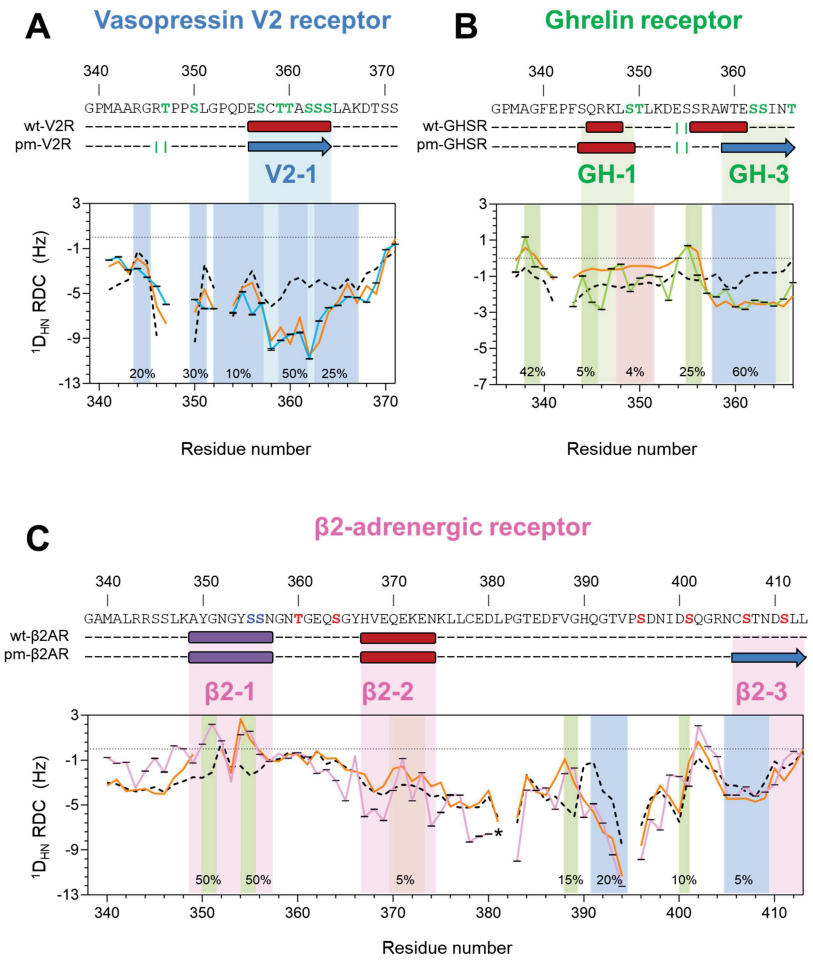


Figure S8: Related to Figure 2. Experimental and back-calculated $^1D_{NH}$ RDCs of the phosphomimetic variants of A) V2R-Cter (blue), B) GHSR-Cter (green) and C) β2AR-Cter (pink). Top, sequence and NMR secondary structure consensus of GPCR C-termini are shown according to figure 2. Bottom, comparison of experimental RDCs values (full line) measured in bacteriophage Pf1 for GHSR-Cter, and alcohol mixture for V2R-Cter and β2AR-Cter pm with back-calculated RDC computed by FM on random-coil (black dash line), or on biased ensemble (orange line) populated as indicated in the figure (regions colored in red (helical structure), blue (extended structure) or green (turn). Outliers due to overlapping residues were removed (dark star).

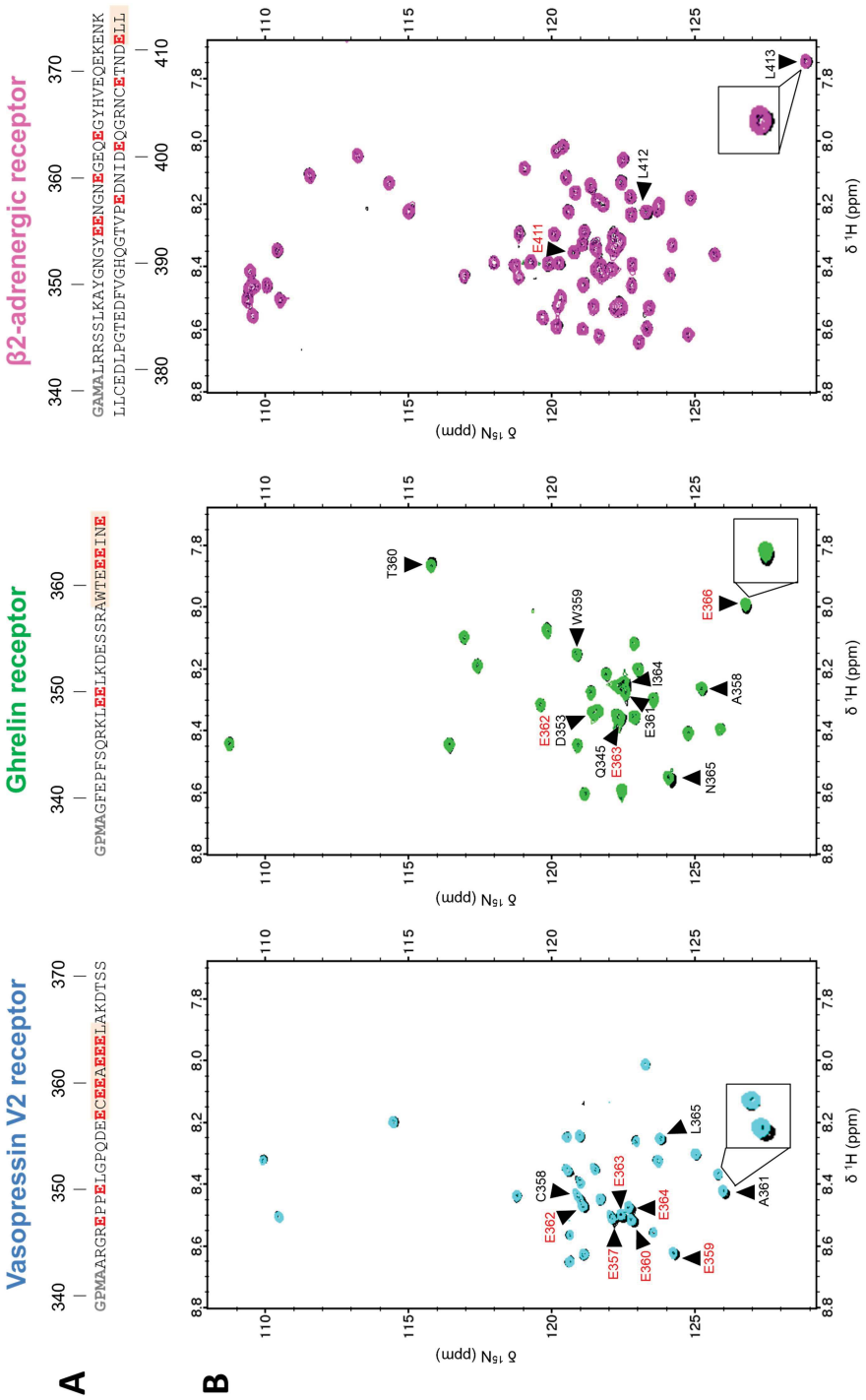


Figure S9: Related to Figure 3. Visualization of pm-GPCR-C-ter:arrestin interaction by NMR of V2R-Cter (lbue, left), GHSR-Cter (green, middle) and β2AR-Cter (pink, right) A) Sequence of the GPCR C-termini. Glutamic acids that mimic GRK phosphorylation sites are indicated in red. Binding regions are highlighted in orange. **B)** Overlay of ^{15}N-HSQC spectra without arrestin (in black) and with arrestin at ratio 1:10 (in color). Residues involved in the interaction are noted in black or red according to (A). Spectra were recorded on sample at 10 µM C-termini with 0 or 100 µM of arrestin at 800 MHz, 20 °C, 50 mM Bis-Tris pH 6.7, 50 mM NaCl.

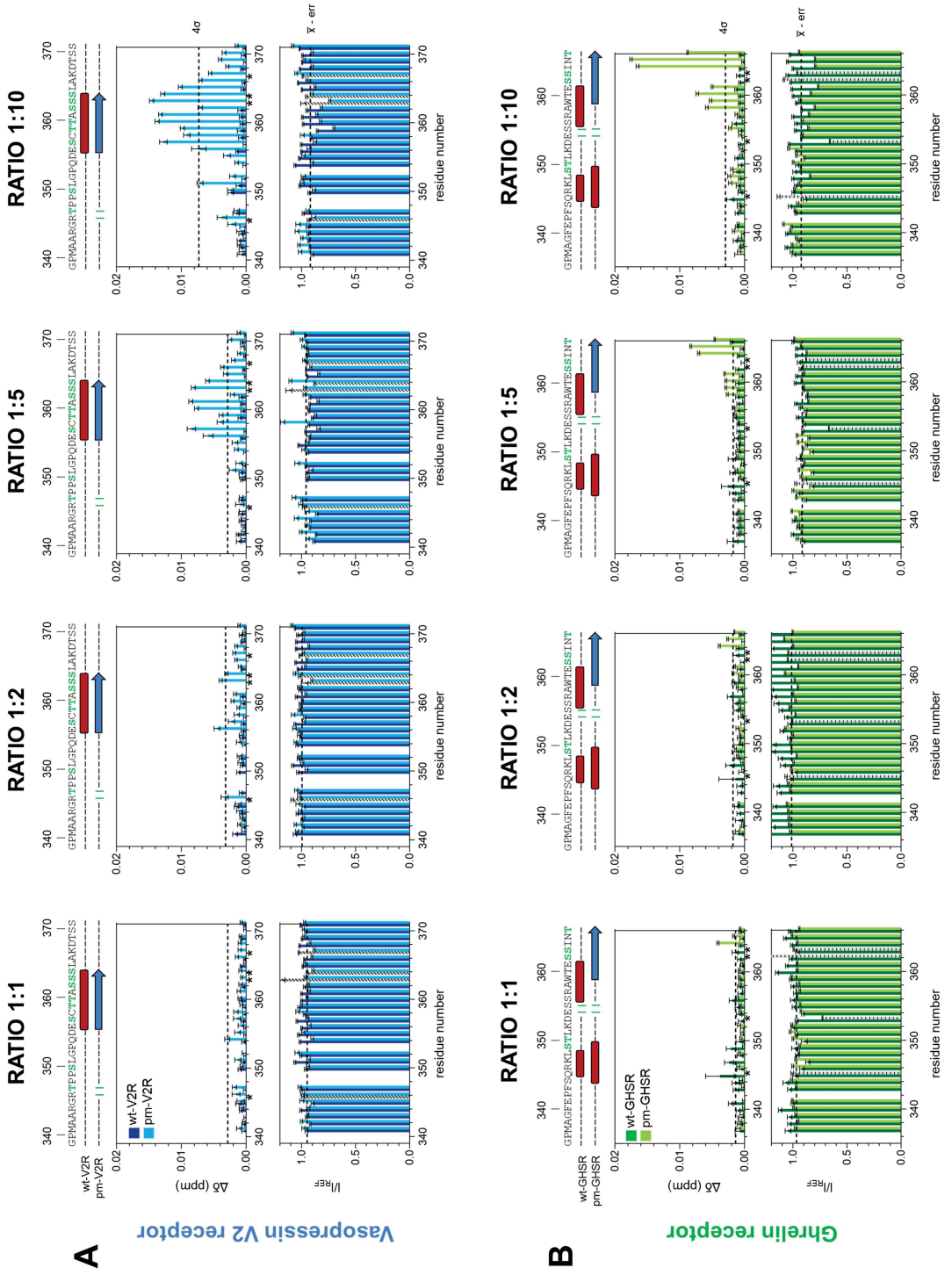


Figure S10: Related to Figure 3. Interaction by NMR between arrestin-2 and A) wt-V2R-Cter (dark blue) or pm-V2R-Cter (light blue) and B) wt-GHSR-Cter (dark green) or pm-GHSR-Cter (light green) at different ratios. For each C-terminus, the sequences and NMR secondary structure consensus are indicated according to Figure 2. Below, histogram of $^{1}H/^{15}N$ chemical shift changes and intensity ratios (I/I_{REF}) are shown. Cut-offs are indicated in black, dashed line and on the right. In the histogram of chemical shit changes, overlapping residues in HSQCs were removed and are shown in black stars near residues numbers. In the histogram of intensity ratio, they are represented in dashed column. Spectra were recorded at 800 MHz, 20 °C, 50 mM Bis-Tris pH 6.7, 50 mM NaCl, on sample at 10 µM C-termini with 0, 10, 20, 50 or 100 µM of arrestin

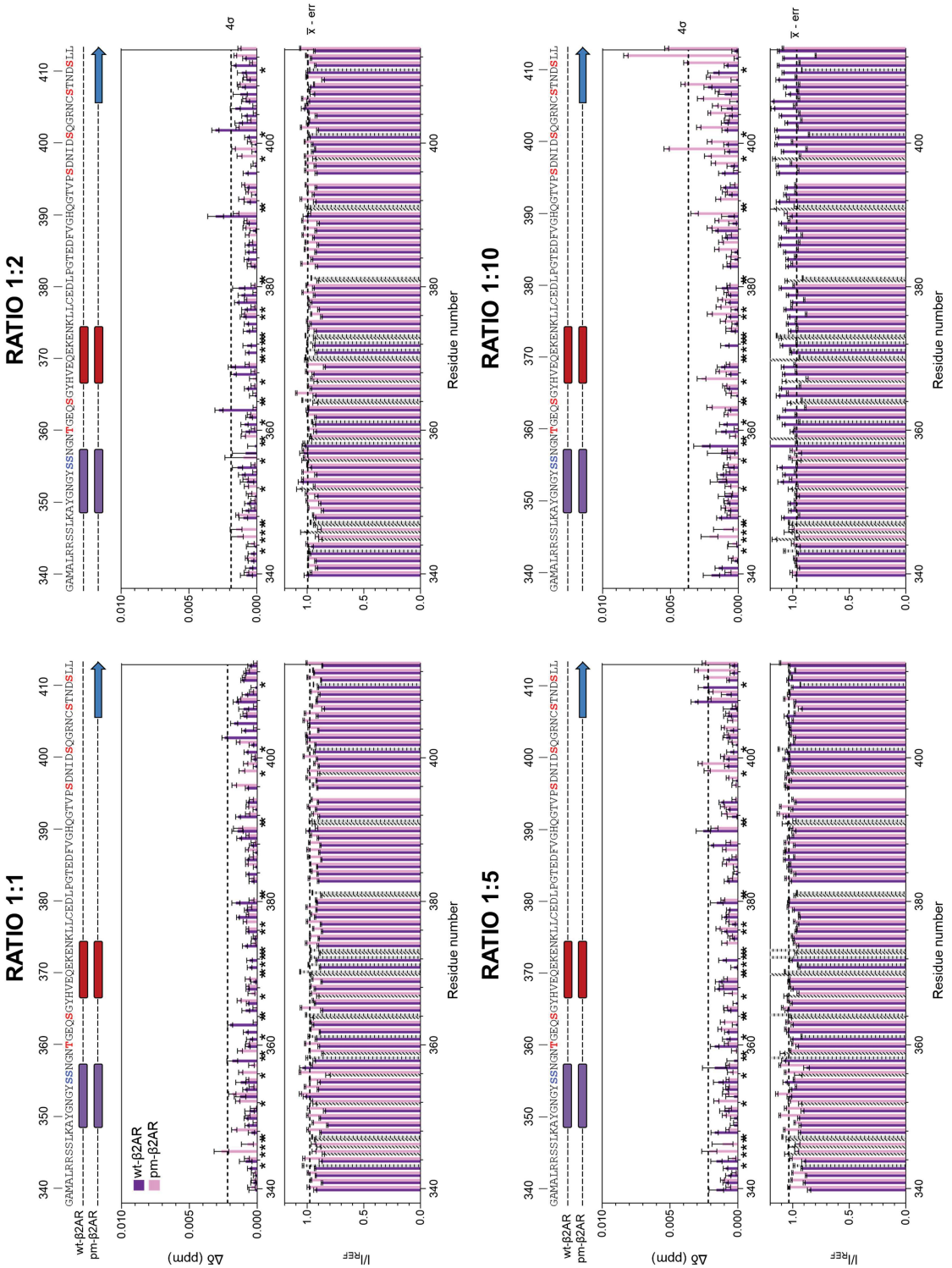


Figure S11: Related to Figure 3. Interaction by NMR between arrestin-2 and wt-β2AR-Cter (purple) or pm-β2AR-Cter (pink) at different ratios. For each ratio, the sequences and NMR secondary structure consensus are indicated according to Figure 2. Below, histogram of $^{1}H/^{15}N$ chemical shift changes and intensity ratios (I/I_{REF}) are shown. Cut-offs are indicated in black, dashed line. . In the histogram of chemical shit changes, overlapping residues in HSQCs were removed and are shown in black stars near residues numbers. In the histogram of intensity ratio, they are represented in dashed column. Spectra were recorded at 800 MHz, 20 °C, 50 mM Bis-Tris pH 6.7, 50 mM NaCl, on sample at 10 µM C-termini with 0, 10, 20, 50 or 100 µM of arrestin.

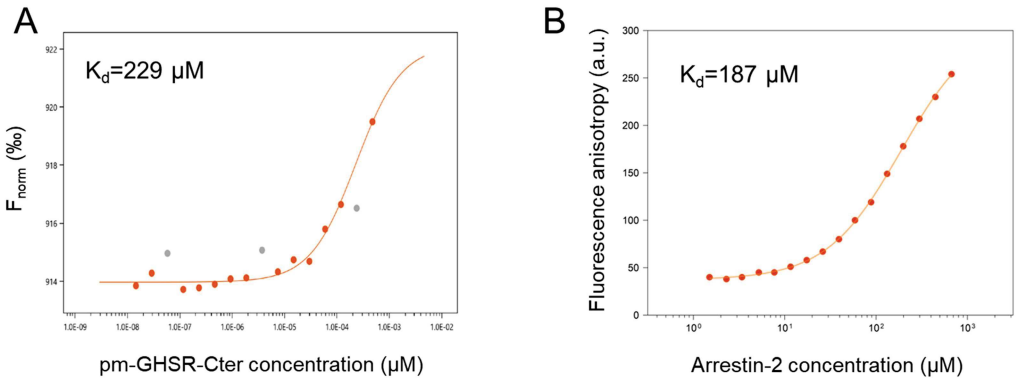


Figure S12: Low affinity interaction between arrestin 2 and the C-terminus of GHSR phosphomimetic. A) Dose-response curve of the Micro-Scale Thermophoresis (MST) titration. Labeled arrestin 2 at 5 nM was incubated with GHSR peptide (from 0.01 to 450 µM), at pH 6.7, 20 °C, 50 mM NaCl. An MST on-time of 5 sec was used for analysis and a dissociation constant (k_D) of 229 µM was obtained. Three data points were removed from the calculation (grey dots). **B)** Fluorescence anisotropy curve. Labeled GHSR was incubated with varying amounts of arrestin 2: from 0 to 670 µM, at pH 6.7, 20 °C, 50 mM NaCl. A 187 µM dissociation constant (k_D) was determined.

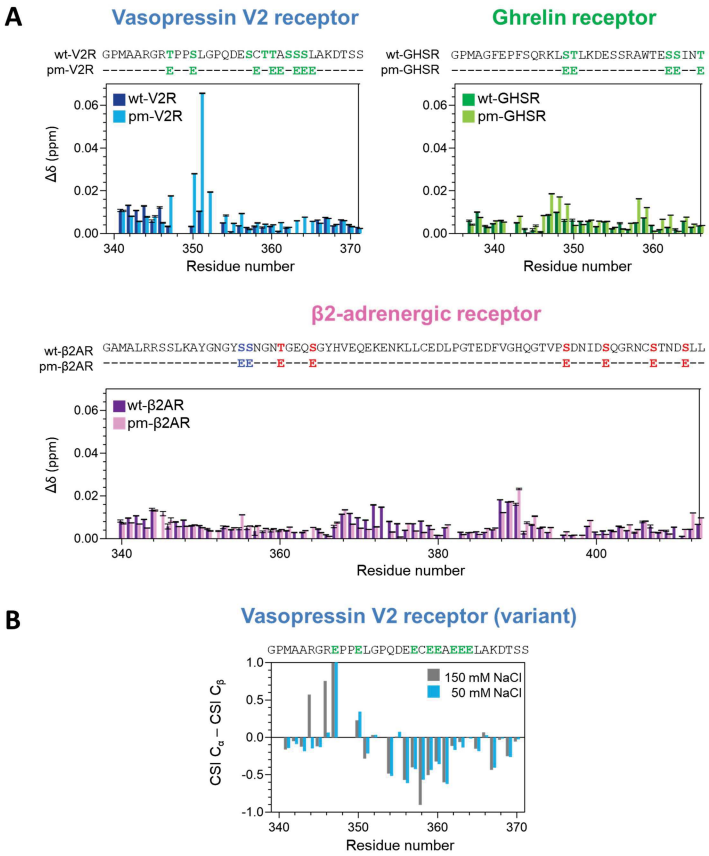


Figure S13: Effect of salt on the conformation of wild type (dark color) and variant (light color) forms of V2R-Cter (blue); GHSR-Cter (green) and β2AR-Cter (purple/pink). A) Top, sequence of wild type (wt) or phosphomimetic (pm) GPCR-Cters. Bottom, $^{1}H/^{15}N$ chemical shift change between HSQC recorded on samples at 100 µM at 700 MHz, 20°C, 50 mM Bis-Tris pH 6,7 with 150 or 50 mM NaCl. **B)** Impact of NaCl on the chemical shift of pm-V2R-Cter $^{13}C_{\alpha}/C_{\beta}$ nuclei. Computed $^{13}C_{\alpha}/C_{\beta}$ CSI at 150 mM (grey) and 50 mM (blue) NaCl are represented.

REFERENCES

Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017a) Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. Adv Exp Med Biol 1009: 107–129

Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017b) Small-angle scattering studies of intrinsically disordered proteins and their complexes. Curr Opin Struct Biol 42: 15–23

Sibille N & Bernadó P (2012) Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochem Soc Trans 40: 955–962

4.3 PARTIE 3 : PROBLEMATIQUE ET RESUME

La phosphorylation des RCPG-Cters par les GRKs module leur région d'interaction avec l'arrestine

Dans les chapitres précédents, nous avons comparé les conformations des régions C-terminales de nos trois RCPGs modèles dans leur état basal versus phospho-mimétique GRK2/3 et GRK5/6, et étudié leurs interactions respectives avec l'arrestine-2. Ainsi, nous avons montré que les mimes de la phosphorylation ont pour effet de modifier la conformation des régions C-terminales des RCPGs et permettent de déclencher l'interaction avec le partenaire arrestine. Cependant, plusieurs points restent à élucider : **1**) Le même phénomène de repliement aurait-il lieu avec les régions C-terminales phosphorylées par les GRKs ? **2**) Est-ce que l'interaction avec le partenaire arrestine se déroulerait de la même manière ? **3**) Quels seraient les sites spécifiquement phosphorylés par chaque kinase ?

Pour répondre à ces questions, nous avons mis en place, la phosphorylation des régions C-terminales des trois modèles de récepteur par GRK2 et/ou GRK6. Tout d'abord, nous avons mis en place la production des kinases GRK2 et GRK6 en système baculovirus (Chapitre 3, p. 90). La production de GRK2 présente un rendement de ~ 7 mg/L de culture, ce qui nous a permis d'effectuer des tests de phosphorylation. À l'inverse, la production actuelle de GRK6 permet d'obtenir ~ 0.3 mg/L de culture, et elle nécessite d'être optimisée. Ensuite, nous avons mis au point un protocole de phosphorylation des domaines C-terminaux des RCPGs et nous avons suivi la phosphorylation par RMN. En effet, la RMN constitue un outil adapté pour l'étude de phosphorylations multiples au sein de protéines intrinsèquement désordonnées (IDPs) (Landrieu *et al*, 2006). Elle permet l'identification des sites phosphorylés et l'obtention du taux de phosphorylation. Elle permet également de caractériser les paramètres cinétiques de la réaction de phosphorylation, pour chaque site phosphorylé, à partir de simple spectres ^{15}N-HSQCs. Ainsi, nous avons identifié les sites phosphorylés des domaines C-terminaux du récepteur β2-adrénergique (β2AR) et du récepteur à la ghréline (GHSR), par GRK2. Certains de ces sites ont déjà été décrits comme phosphorylés par les GRKs dans la littérature, alors que d'autres ont été identifiés pour la première fois dans notre étude. Les sites

phosphorylés seront confirmés par spectrométrie de masse après phosphorylation des récepteurs entiers reconstitués en nanodisque (collaboration avec Jean-Louis Banères).

Nous avons souhaité comprendre comment la phosphorylation des domaines C-terminaux des RCPGs par les GRKs affecte leur conformation et leur interaction avec l'arrestine. Pour cela, nous avons commencé par la caractérisation structurale et fonctionnelle du domaine C-terminal du récepteur GHSR phosphorylé par GRK2. Nous avons observé une transition structurale entre la forme non-phosphorylée et la forme phosphorylée par GRK2, similaire à celle des phosphomimétiques. De plus, la région d'interaction avec l'arrestine est localisée dans la région subissant cette transition structurale, comme pour celle des variants phosphomimétiques. Ainsi, un mécanisme similaire d'interaction et de repliement semble prendre place et nous pouvons conclure que les phosphomimetiques sont de très bons modèles de phosphorylation pour les régions C-terminales des RCPGs.

En résumé :

La production des kinases GRK2 et GRK6 a été mise en place au laboratoire (Chapitre 3, p. 81) et un protocole de phosphorylation a été développé.

Nous avons identifié les sites phosphorylés des domaines C-terminaux des récepteurs β2AR et GHSR par GRK2.

L'étude structurale et fonctionnelle du domaine C-terminal de GHSR phosphorylé par GRK2 confirme que la phosphorylation régule la conformation des domaines C-terminaux des RCPGs et leur interaction avec l'arrestine.

Les résultats étant similaires à ceux obtenus pour les variants phosphomimétiques, les variants sont de très bons modèles de phosphorylation pour les régions C-terminales des RCPGs.

Dans cette partie, j'ai contribué à la mise en place et à la production des kinases GRK2 et GRK6. J'ai mis au point le protocole de phosphorylation et obtenu les formes phosphorylées des récepteurs β2AR et GHSR. J'ai débuté l'analyse structurale et fonctionnelle par RMN du domaine C-terminal de GHSR phosphorylé par GRK2.

Les figures supplémentaires de cette partie sont en Annexe 2, p.278

4.3.1 Pourquoi mettre en place la production des GRKs au laboratoire ?

La difficulté majeure de cette étude correspond à la phosphorylation *in vitro* des peptides C-terminaux des RCPGs (RCPG-Cters), isolés de leur environnement d'origine, qui est le récepteur entier lié à la membrane. En effet, il a été montré que l'activité catalytique des GRKs envers les RCPG-Cters est réduite par rapport à celles des récepteurs entiers. Par exemple, GRK2 est capable de phosphoryler le peptide du domaine C-terminal du récepteur β2-adrénergique (β2AR-Cter), cependant, la cinétique de la réaction révèle que l'affinité, représentée par le K_m (constante de Michaelis), est de ~ 3 mM pour le peptide contre ~ 0.2 µM pour le récepteur activé par de l'isoprotérénol (Benovic *et al*, 1990). De plus, la vitesse de la réaction enzymatique maximale (V_{max}) est de ~ 2.5 nmol.min^{-1}.mg^{-1} pour le peptide contre 37 nmol.min^{-1}.mg^{-1} pour le récepteur activé. Ainsi, selon le ratio V_{max}/K_m, le peptide de β2AR-Cter constitue un substrat 10^6 fois moins efficace que le récepteur entier activé. Une étude similaire a montré que GRK1 a une affinité de 8 mM pour le peptide du domaine C-terminal de la rhodopsine (Rho-Cter), contre 3 µM pour le récepteur entier activé par illumination (Palczewski *et al*, 1988). Cette différence d'activité entre peptides et RCPG entier peut s'expliquer par l'absence d'interaction entre les GRKs et le domaine central du récepteur activé. En effet, dans leur environnement cellulaire, les GRKs interagissent avec le cœur du RCPG, ce qui permet une proximité naturelle des domaines C-terminaux aux GRKs, ainsi qu'une stabilisation des GRKs dans leur conformation active (Chapitre 1, p. 48, Figure 7).

Ainsi, la quantité de peptide phosphorylé obtenue est entravée par l'activité réduite des kinases dans ce système simplifié. S'ajoute à cette limitation, la difficulté du nombre de sites à phosphoryler. En effet, les peptides des domaines C-terminaux des RCPGs sont connus pour être phosphorylés sur plusieurs sites. Pour le récepteur β2AR, six sites ont été décrits comme étant phosphorylés par GRK2 (Nobles *et al*, 2011). Pour les récepteurs à la ghréline (GHSR) et à la vasopressine V2 (V2R), minimum cinq sites ont été identifiés comme phosphorylés par des GRKs (Innamorati *et al*, 1998; Bouzo-Lorenzo *et al*, 2016). La description des sites phosphorylés de chaque RCPG-Cter sera présentée plus en détail dans les paragraphes suivants.

La faible affinité des GRKs pour les peptides isolés couplée au nombre potentiel de sites à phosphoryler induit une contrainte technique, nécessitant l'ajout de la kinase en excès pour la réaction de phosphorylation. L'utilisation de kinases commerciales (GRK2 et 6), chères et en faible quantité (ratio maximal de 1 kinase:5 peptides), ne permettaient pas d'obtenir un ratio adéquat à cette étude. Ensuite, les tests effectués avec la kinase GRK2 (ratio maximal de 2 kinases pour 1 peptide) fournie par Martin J. Lohse (Université de Würzburg, Allemagne), ont confortés notre souhait de mettre en place l'expression de nos propres kinases au laboratoire en système d'expression baculovirus.

Aujourd'hui, la production de la kinase GRK2 au laboratoire donne un rendement de ~7 mg/L de culture. Cela nous permet de travailler à des ratios de 4 kinases pour 1 peptide et cela nous a permis d'obtenir les formes phosphorylées de nos trois RCPG-Cters. À l'inverse, le rendement actuel de GRK6 est de ~ 0.3 mg/L de culture et sa production nécessite des optimisations. Par conséquent, les résultats présentés dans cette partie concernent la phosphorylation des RCPG-Cters par GRK2 uniquement.

4.3.2 Suivi de la phosphorylation des IDPs par RMN

Il a été montré pour la première fois que la RMN est un outil bien adapté à l'étude de la phosphorylation multiple des protéines intrinsèquement désordonnées (IDPs) lors d'une étude de la phosphoryation de Tau (Landrieu *et al*, 2006). Elle peut permettre l'identification des sites phosphorylés, l'obtention du taux de phosphorylation et des paramètres cinétiques de la réaction pour chacun de ces sites à partir de simple spectres ^{15}N-HSQCs.

Sur un spectre ^{15}N-HSQC, les résidus phosphorylés peuvent être facilement détectés car la phosphorylation entraine des changements de leur déplacement chimique, causés par la modification de leur environnement local (Hendus-Altenburger *et al*, 2019). Ainsi, les phospho-sérines et phospho-thréonines sont déplacées vers les bas champ du spectre (~ 8.6 à 9.6 ppm), en dehors de la fenêtre spectrale normale des IDPs (~ 7.6 à 8.6 ppm) (Landrieu *et al*, 2006) (Figure 48). Par ailleurs, cet effet de la phosphorylation sur le déplacement chimique atteint généralement jusqu'à deux résidus voisins de part et d'autre du résidu phosphorylé. Des résidus plus lointains peuvent aussi être affectés si des réarrangements

structuraux sont associés à la phosphorylation (Landrieu *et al*, 2006; Hendus-Altenburger *et al*, 2019).

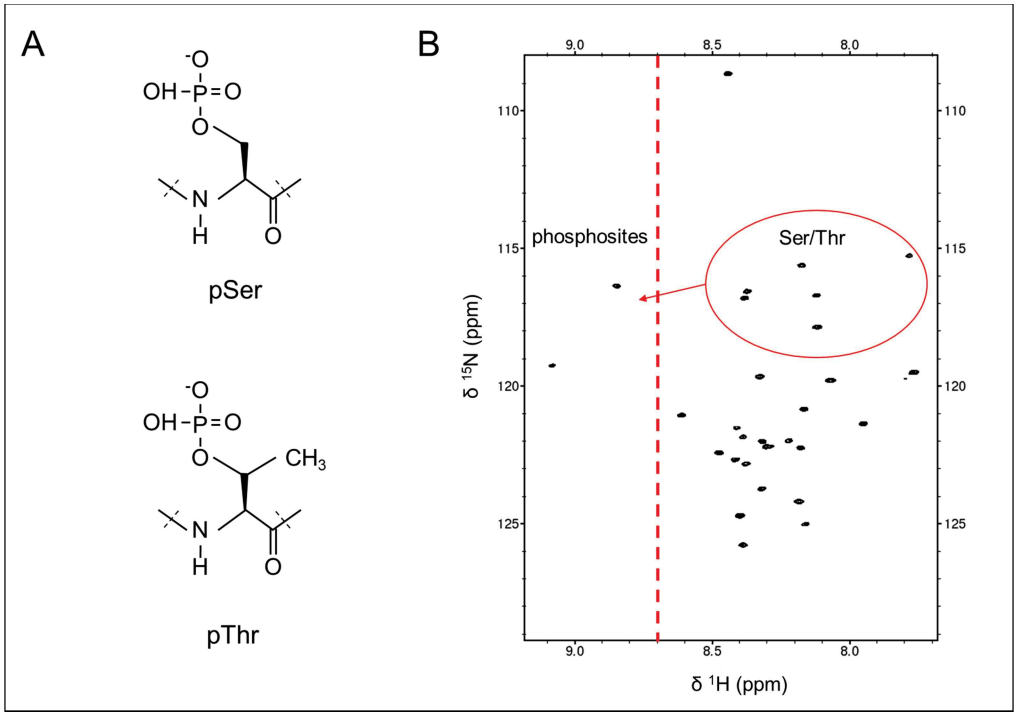


Figure 48 : Suivi de la phosphorylation des IDPs par RMN. A) La présence du groupement phosphate sur la chaîne latérale entraine une modification de l'environnement chimique local et va modifier le déplacement chimique de la fonction amide des sérines/thréonines. **B)** Cela entraine un déblindage des phospho-résidus qui se déplacent en dehors de la fenêtre spectrale classique des IDPs (limite représentée par les tirets rouges).

En plus d'être facilement détecté, le taux de phosphorylation de chaque site peut être déterminé par mesure des intensités. En effet, l'incorporation de phophate sur un résidu entraîne l'apparition d'une résonance dans sa forme phosphorylée (pX) et la disparition concomitante de sa résonance non-phosphorylée (X). La somme des intensités de ces deux résonnances (I_{pX} + I_X) étant de 100 %, une simple soustraction permet d'obtenir le taux de phosphorylation (I_{pX} = 100 % - I_X).

L'analyse concomitante des variations de déplacement chimique et d'intensité au cours du temps de réaction peut permettre d'identifier les sites phosphorylés. Pour se faire, une cinétique de phosphorylation est effectuée par acquisition d'une série de spectres ^{15}N-HSQCs correspondant à différent temps de réaction. Cette cinétique peut avoir lieu directement dans le spectromètre RMN ou à la paillasse. Cependant, les réactions de phosphorylation ont souvent lieu à ~ 25 - 30 °C et à pH ~ 7 – 8 pour

que l'activité des kinases soit optimale. Dans ces conditions, des échanges rapides des protons amides avec l'eau peuvent avoir lieu ce qui entrainerait la disparition des pics RMN (Canet *et al*, 2017). Par exemple, dans le cas du domaine C-terminal de β2AR phosphorylé par GRK2, l'acquisition de spectre à pH 7.5 entraine la disparition de nombreux résidus dont les sérines et thréonines (Figure 49). Par conséquent, toutes les expériences de phosphorylation citées par la suite dans ce manuscrit sont effectuées à la paillasse, puis une dialyse est effectuée pour passer en tampon RMN (50 mM Bis-Tris pH 6.7, 50 ou 150 mM NaCl, 0.5 mM EDTA, 1 mM TCEP).

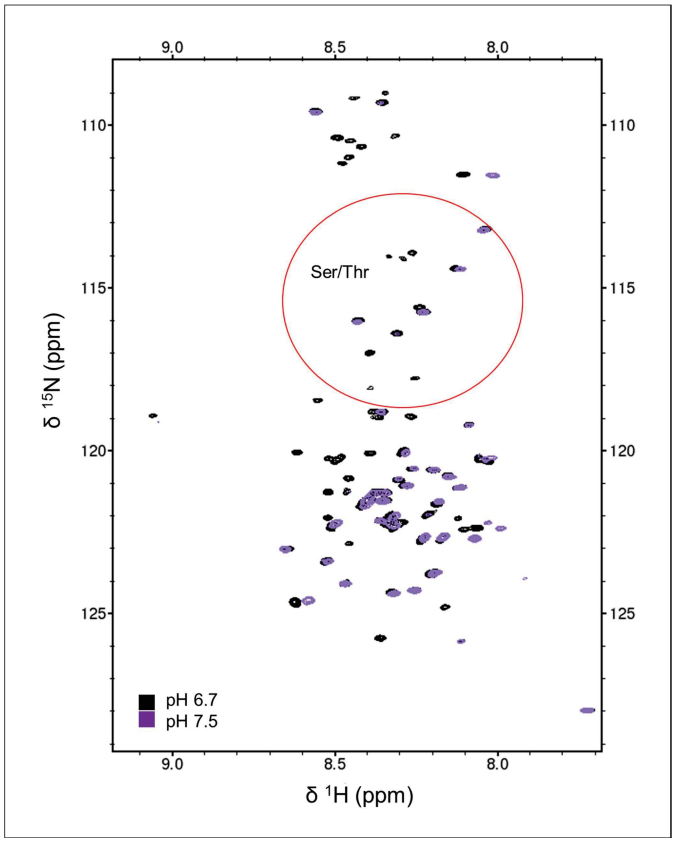


Figure 49 : Disparition de pics par échange proton amide avec l'eau à pH 7.5. Superposition des spectres ^{15}N-HSQC à pH 6.7 (noir) et 7.5 (violet) du peptide de β2AR phosphorylé par GRK2 à 30 µM. Après phosphorylation en tampon 25 mM Tris pH 7.5, un spectre RMN a directement été enregistré (violet). Après acquisition de ce spectre HSQC, l'échantillon a été dialysé en tampon RMN contenant 50 mM de Bis-Tris pH 6.7 et un nouvel HSQC a été enregistré. On peut voir sur la superposition des deux HSQCs que des pics disparaissent à pH 7.5, notamment dans la région des sérines/thréonines (encadrée en rouge). L'augmentation du pH augmente la vitesse de l'échange chimique de certains hydrogènes de la protéine avec le solvant, les rendant invisibles par RMN (Cordier *et al*, 2012). Ceci est dû au taux d'échange accru avec l'eau, ainsi la position du pic dans un régime d'échange rapide est située à la position moyenne pondérée entre la résonance de l'eau et la position amide. En raison de la grande quantité d'eau (100 M de ^{1}H) par rapport au peptide (30 µM de ^{1}H), le pic est très proche de la résonance de l'eau.

Pour en revenir à l'analyse de la cinétique de phosphorylation, si la phosphorylation est séquentielle (un site après l'autre), il est possible de suivre l'apparition d'un résidu phosphorylé et la disparition concomitante de sa résonance non-phosphorylée. De plus, tracer l'évolution de l'intensité de ces pics peut permettre d'obtenir leur constante de vitesse apparente (K_{app}) d'apparition et de disparition. Cela peut conduire à l'attribution des résidus phosphorylés. Par exemple, pour un résidu i dont l'incorporation de phosphate augmente au cours du temps, l'intensité de sa résonance initiale (non-phosphorylée) va diminuer proportionnellement à l'augmentation de l'intensité de sa résonance phosphorylée. Ainsi les vitesses d'apparition et de disparition de ces deux résonances doivent être égales.

Cette attribution peut être confirmée par l'analyse de la variation de déplacement chimique et d'intensité des résidus voisins, qui seront aussi affectés (Landrieu *et al*, 2006). Cependant, si plusieurs phosphorylations ont lieu en même temps, l'attribution par cinétique peut s'avérer fastidieuse et l'acquisition de spectre RMN 3D deviendrait nécessaire. Pour finir, la cinétique permet aussi d'obtenir l'ordre séquentiel des phosphorylations selon leur ordre d'apparition temporel sur les spectres RMN.

Concernant les RCPGs, leur domaine C-terminal désordonné (RCPG-Cter) est généralement phosphorylé sur plusieurs sites, parfois adjacents, par les GRKs. Par conséquent, identifier les résidus phosphorylés par spéctrométie de masse peut s'avérer difficile. En effet, les enzymes utilisées pour cliver les protéines en fragments, ne peuvent pas cliver entre deux sérines ou thréonines adjacentes. Ainsi, la RMN semble être la technique adéquate pour l'étude de la phosphorylation des domaines C-terminaux des RCPGs par les kinases GRK2 et GRK6.

4.3.3 Optimisation du tampon de phosphorylation

La RMN peut permettre de quantifier le taux de phosphorylation grâce à l'analyse des intensités des pics phosphorylés. Nous avons utilisé cette propriété pour mettre au point la réaction de phosphorylation. Pour cela nous avons comparé les intensités des pics phosphorylés normalisés à l'intensité d'un pic non modifié par la phosphorylation.

Pour la mise au point du tampon de phosphorylation, nous avons cherché un compromis entre le tampon de conservation de la kinase et son tampon de réaction, de manière à éviter toute étape de changement de tampon. Dans la littérature, le

tampon de phosphorylation est communément composé de 20 mM de Tris à pH 7.5, 2 mM EDTA, 8 mM $MgCl_2$ et d'ATP en excès par rapport à la concentration de la kinase (Sohlemann *et al*, 1993; Dicker *et al*, 1999; Sterne-Marr *et al*, 2013). Initialement, notre tampon de conservation était composé de 50 mM Tris pH 7.5; 150 mM NaCl, 2 mM DTT, 0.02 % Triton X100, 1mM benzamidine et 25 % de glycérol selon (Sterne-Marr *et al*, 2013). Les différences principales entre ces deux tampons sont la présence de DTT, pour prévenir la formation de ponts disulfures intra-protéiques, de Triton X100 connu pour stabiliser GRK2 (Benovic, 1991), de benzamidine pour empêcher la protéolyse et du glycérol à 25 % pour une meilleure conservation à -80 °C. En effet, après décongélation, la température de dénaturation de GRK2 est de 59.7 °C en présence de glycérol contre 55.7 °C sans glycérol (Figure S1). Cependant, de par sa propriété cryoprotectante, le glycérol pourrait réduire l'accès des peptides à la kinase et ainsi diminuer l'activité de GRK2. Pour le savoir la réaction de phosphorylation de β2AR-Cter par GRK2 a été effectuée à la paillasse avec ou sans glycérol, puis analysée en RMN. Comme attendu, le glycérol a provoqué de faibles perturbations du déplacement chimique des acides aminés (Figure 50A). Sur les deux spectres, deux sites à ~ 9 ppm en dimension ^{1}H sont observés, indiquant la présence de deux phosphorylations (Pho1 et Pho2, Figure 50A). La comparaison de leur intensité normalisée révèle une diminution de l'intensité de Pho1 en présence de glycérol (Figure 50A). Ainsi, l'activité de GRK2 est réduite en présence de glycérol, et le glycérol a donc été supprimé du tampon de conservation.

Après sélection du tampon de conservation/phosphorylation, nous nous sommes intéressés au ratio kinase/ATP adéquate pour la phosphorylation. L'ATP est couramment mis en excès, excès allant de x 200 (Sterne-Marr *et al*, 2013) à x 1 000 (Dicker *et al*, 1999) selon les protocoles. Nous avons testé de phosphoryler la région C-terminale de ^{15}N-GHSR avec un ratio 500 ou 1 000 entre la concentration de GRK2 et la concentration d'ATP. Aucune phosphorylation n'est observée à ratio 1000 en ATP, alors que deux pics (Pho1 et Pho2) sont visibles à ratio 500 (Figure 50B). Par conséquent, nous avons sélectionné le ratio 500 pour la concentration d'ATP à utiliser. Par la suite, nous avons diminué la concentration d'ATP à un excès de 250 sans que cela n'affecte l'efficacité de la phosphorylation (résultat non montré).

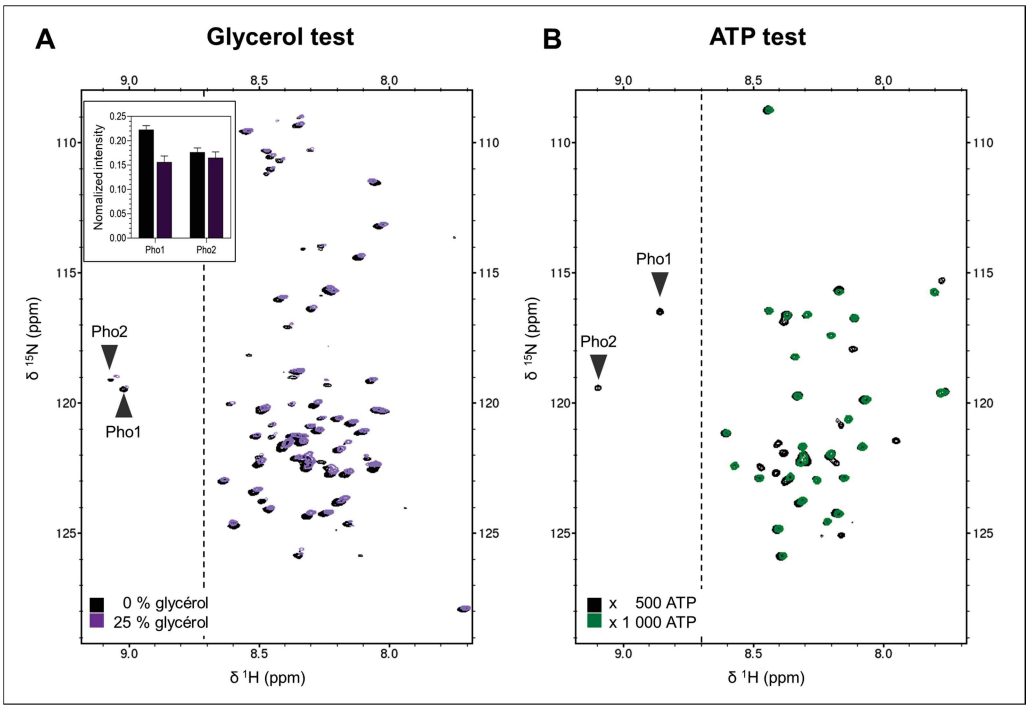


Figure 50 : Mise au point du protocole de phosphorylation des régions C-terminales des RCPGs par GRK2. Sur chaque spectre, les nouvelles résonances des résidus phosphorylés apparaissant à bas champs (après 8.7 ppm, ligne noire pointillée) sont indiquées par Pho1 et Pho2. Les spectres ont été enregistrés sur des échantillons à 10 µM préparés en Bis-Tris pH 6.7, 150 mM NaCl, à 20 °C, et sur les spectromètres 700 MHz pour le test glycérol et à 800 MHz pour le test ATP. **A)** Superposition des spectres ^{15}N-HSQC obtenus par phosphorylation de β2AR en absence (noir) ou en présence (violet) de glycérol. Les intensités normalisées des pics phosphorylés Pho1 et Pho2 dans les deux conditions sont affichées dans l'encart **B)** Superposition des spectres ^{15}N-HSQC obtenus par phosphorylation de GHSR en présence d'un ratio kinase:ATP de 500 (noir) ou 1000 (vert).

Dans l'objectif de pouvoir effectuer des cinétiques de phosphorylation, nous avons inactivé la kinase par chauffage de l'échantillon 10 min à 75 °C. Le chauffage permet de dénaturer la kinase sans affecter les peptides désordonnés des RCPGs (Annexe 2, p. 278,Figure S2). L'inactivation thermique des réactions de phosphorylation des IDPs a déjà été décrite pour la protéine Tau (Landrieu *et al*, 2006), et est classiquement proposée comme méthode d'arrêt de la réaction de phosphorylation (Julien *et al*, 2020).

Ainsi, nous avons mis en place un protocole de phosphorylation avec pour tampon de phosphorylation 50 mM Tris pH 7.5; 150 mM NaCl, 0.02 % Triton X100, 2mM DTT, 1mM benzamidine, 10 mM $MgCl_2$ et un excès de 250 en ATP. Après phosphorylation à 30 °C, la réaction est stoppée par chauffage 10 min à 75 °C. Après centrifugation, le peptide est dialysé et passé en tampon RMN pour son analyse.

4.3.4 Les sites de phosphorylation de GRK2

4.3.4.1 Identification des sites de phosphorylation de GRK2 par RMN

4.3.4.1.1 Phosphorylation de la région C-terminale de β2AR

Nous avons tout d'abord utilisé la cinétique de phosphorylation afin d'identifier les résidus phosphorylés du domaine C-terminal de β2AR par GRK2. Pour cela, la réaction de phosphorylation a été effectuée à la paillasse et elle a été stoppée au bout de 15 min, 30 min, 1h, 2h, 4h ou 16h. Après dialyse, les spectres ^{15}N-HSQCs de ces solutions ont été enregistrés sur le spectromètre 800 MHz Un premier site phosphorylé apparaît à bas champs (après 8.6 ppm en ^{1}H) au bout d'une heure de réaction (Pho1) et un deuxième site apparaît au bout de 2 h de réaction (Pho2) (Annexe 2, p. 278 Figure S3).

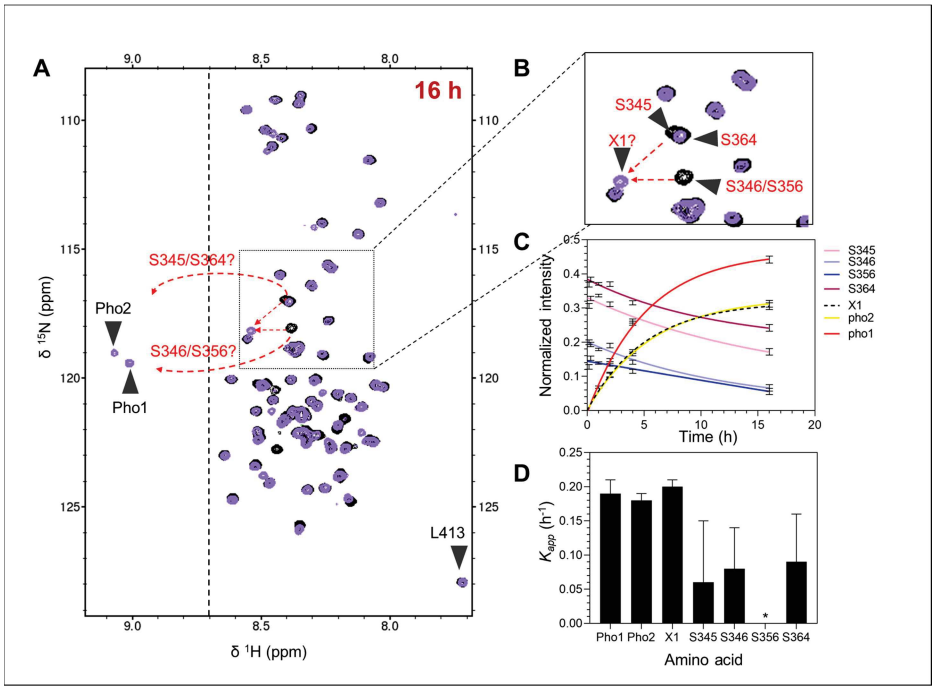


Figure 51 : Cinétique des sites phosphorylés du domaine C-terminal de β2AR (β2AR-Cter) par GRK2. A) Spectre ^{15}N-HSQC de l'échantillon phosphorylé pendant 16 heures. Les sérines potentiellement phosphorylées sont indiquées en rouge et les potentiels déplacements chimiques qu'elles ont subi sont représentés par les flèches rouges. La ligne noire pointillée représente la fin de la zone 'classique' de déplacement chimique des IDPs. Le résidu L413 qui a servi à la normalisation des intensités est affiché sur le spectre. Ce spectre a été enregistré sur un échantillon de 10 µM, dans un tampon de 50 mM Bis-Tris pH 6.7, 150 mM NaCl sur le 800 MHz, à 20 °C. **B)** Zoom de la région des sérines/thréonines issue de l'HSQC en (**A**). **C)** Suivi de l'intensité normalisée des nouveaux pics Pho1, Pho2 et X1 et des résonances non-phosphorylées des sérines S345, S346, S356 et S364. **D)** Constantes de vitesse apparentes des sites présentés en (**C**). L'étoile (*) indique que la constante n'a pas pu être déterminée car l'ajustement de la courbe d'intensité a échoué.

Après 16 h de réaction, il n'y a pas de nouveaux sites phosphorylés visibles mais l'intensité des pics précédents, Pho 1 et Pho2, a augmenté. Hélas, la phosphorylation de ces deux sites n'est pas séquentielle, et plusieurs résonances de sérines non-phosphorylées ont disparu à 16 h : celles des sérines S345 superposée à S364 et S346 superposée à S356 (Figure 51A et B). A 16 h, nous pouvons donc observer l'apparition de deux pics correspondant à des phospho-résidus, et d'un nouveau pic X1 dans la région des sérines contre la disparition des résidus S345 superposé à S364 et S346 superposé à S356. (Figure 51A et B).

Nous avons essayé d'associer les résonances des formes phosphorylées et non-phosphorylées, en étudiant l'évolution de l'intensité normalisée de ces pics au cours du temps (Figure 51C). Ces courbes ont permis l'extraction des constantes de vitesse apparentes (K_{app}) de ces sites avec : 0.19 ± 0.02, 0.18 ±0.01 et 0.20 ± 0.01 h^{-1} pour Pho1, Pho2 et X1 contre 0.06 ± 0.09, 0.08 ± 0.06 et 0.09 ± 0.07 h^{-1} pour S345, S346 et S364, respectivement. L'ajustement de la courbe d'intensité a échoué pour S356 et la constante n'a pas pu être extraite. Malheureusement, nous n'observons pas de correspondance entre les vitesses d'apparition des résidus Pho1, Pho2 et X1 et de disparition des résidus S345, S346 et S364. Cela peut être dû à la superposition de ces sérines sur les spectres HSQCs (Figure 51) qui faussent très certainement la mesure des intensités.

Pour lever l'ambiguïté, nous avons étudié les acides aminés voisins aux phospho-résidus, qui peuvent aussi subir des changements de déplacement chimique et d'intensité. Nous avons donc regardé ces deux paramètres au cours du temps. L'analyse des variations de déplacement chimique indique que la région autour des sérines S345 et S346 est affectée par la phosphorylation (Annexe 2, p. 278, Figure S4). À l'inverse aucune modification n'est visualisée autour des sérines S356 et S364. Ces résultats indiquent que la phosphorylation aurait lieu sur les sérines S345 et S346 et non sur les sérines S356 et S364. Etonnamment, la région allant de ~ Q370 à D380 est aussi affectée par la phosphorylation. Cela peut s'expliquer par la présence d'une interaction longue distance entre ces deux régions, qui a été identifiée par la mesure de PRE (Chapitre 4, Article 1 p. 129).

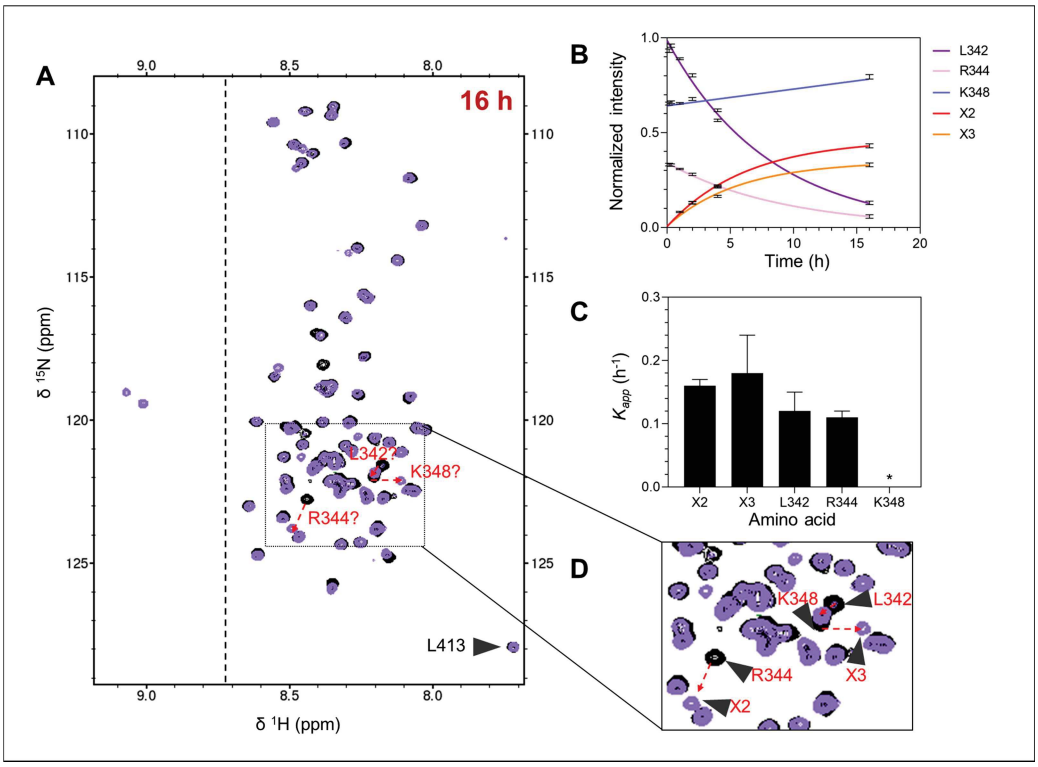


Figure 52 : Perturbation des sites voisins aux sites phosphorylés par GRK2 du domaine C-terminal de β2AR (β2AR-Cter). A) Spectre ^{15}N-HSQC de l'échantillon phosphorylé pendant 16 heures. Les résidus voisins aux sérines S345 et S346 sont indiqués en rouge et les potentiels déplacements chimiques qu'ils ont subi sont représentés par les flèches rouges. La ligne noire pointillée représente la fin de la zone 'classique' de déplacement chimique des IDPs. Le résidu L413 qui a servi à la normalisation des intensités est affiché sur le spectre. Ce spectre a été enregistré sur un échantillon de 10 µM, dans un tampon de 50 mM Bis-Tris pH 6.7, 150 mM NaCl sur le 800 MHz, à 20 °C. **B)** Suivi de l'intensité normalisée des nouveaux pics X2 et X3 et des résonances non-phosphorylées des résidus L342, R344 et K348. **C)** Constantes de vitesse apparentes des sites présentés en (**C**). L'étoile (*) indique que la constante n'a pas pu être déterminée car l'ajustement de la courbe d'intensité a échoué. **D)** Zoom de la région des résidus voisins aux sérines S345 et S346 issu de l'HSQC en (**A**).

Parmi les résidus voisins aux sérines S345 et S346, les résidus L342, R344 et K348 semblent avoir subi un changement de déplacement chimique important car leurs résonances ont disparu dans le spectre de 16 h et de nouvelles résonances, comme X2 et X3, sont apparues (Figure 52A et D). Par la suite, nous avons suivi l'intensité normalisée de ces résidus dans le temps et extrait leur constante de vitesse apparente (Figure 52B et C) : 0.16 ± 0.01 et 0.18 ± 0.06 h^{-1} pour X2 et X3 contre 0.12 ± 0.03 et 0.11 ± 0.01 h^{-1} pour L342 et R344, respectivement. Ces constantes de vitesse sont du même ordre de grandeur et pourrait donc correspondre. Ainsi, X2 correspondrait à R344 et X3 correspondrait à L342. Cependant, il est aussi possible que L342 soit déplacé vers K348 et que K348

corresponde à X3 (Figure 52D). Cela fausserait l'intensité mesurée pour K348 et expliquerait que son intensité normalisée augmente au cours du temps, au lieu de diminuer comme attendu (Figure 52B). L'intensité des résidus voisins aux sérines S356 et S364 n'est pas ou peu modifié par la phosphorylation (Annexe 2, p.278, Figure S5).

Au vu de tous ces éléments, il semblerait que les sérines S344 et S345 soient phosphorylées. Dans ce cas, X1 correspondrait à S356 qui serait affecté indirectement pas la phosphorylation. En effet, selon le modèle obtenu pour β2AR-Cter par Flexible Meccano (FM) (Chapitre 4, Article 1, p. 129), cette région forme un β-turn qui pourrait être affecté par la phosphorylation et/ou la modification des interactions longues distances discutées précédemment. Ces hypothèses devront être vérifiées par l'attribution de spectres 3D sur un échantillon doublement marqué.

4.3.4.1.2 Phosphorylation de la région C-terminale de GHSR

Tout comme pour le domaine C-terminal de β2AR, nous avons d'abord essayé d'identifier les résidus phosphorylés par GRK2 sur GHSR-Cter par cinétique. Pour cela, la réaction de phosphorylation a été effectuée à la paillasse et elle a été stoppée au bout de 0 min, 30 min, 1h, 2h, ou 4h. Après dialyse, les spectres ^{15}N-HSQC de ces solutions ont été enregistrés sur le spectromètre 800 MHz. Deux sites phosphorylés (Pho1 et Pho2) apparaissent à bas champs (après 8.7 ppm en ^{1}H) au bout de 1 h de réaction. Trois heures plus tard, aucune phosphorylation additionnelle n'est visible et les résonances des sérines S355, S356, S362 et S363 ont disparu (annexe 2, p.278, Figure S6). Deux nouvelles résonances X1 et X2 sont aussi apparues dans la région des sérines/thréonines (Figure 53C et Figure S6). La phosphorylation concomitante des sites Pho1 et Pho2 rendait leur identification par cinétique difficile, nous avons donc préparé un échantillon marqué ^{13}C/^{15}N de GHSR phosphorylé pour enregistrer des spectres RMN 3D. Ainsi, l'attribution de la forme phosphorylée de GHSR-Cter a été obtenue (Figure 53) : les sérines S356 et S362 sont les sites phosphorylés et les sites X1 et X2 correspondent respectivement aux sérines S363 et S355.

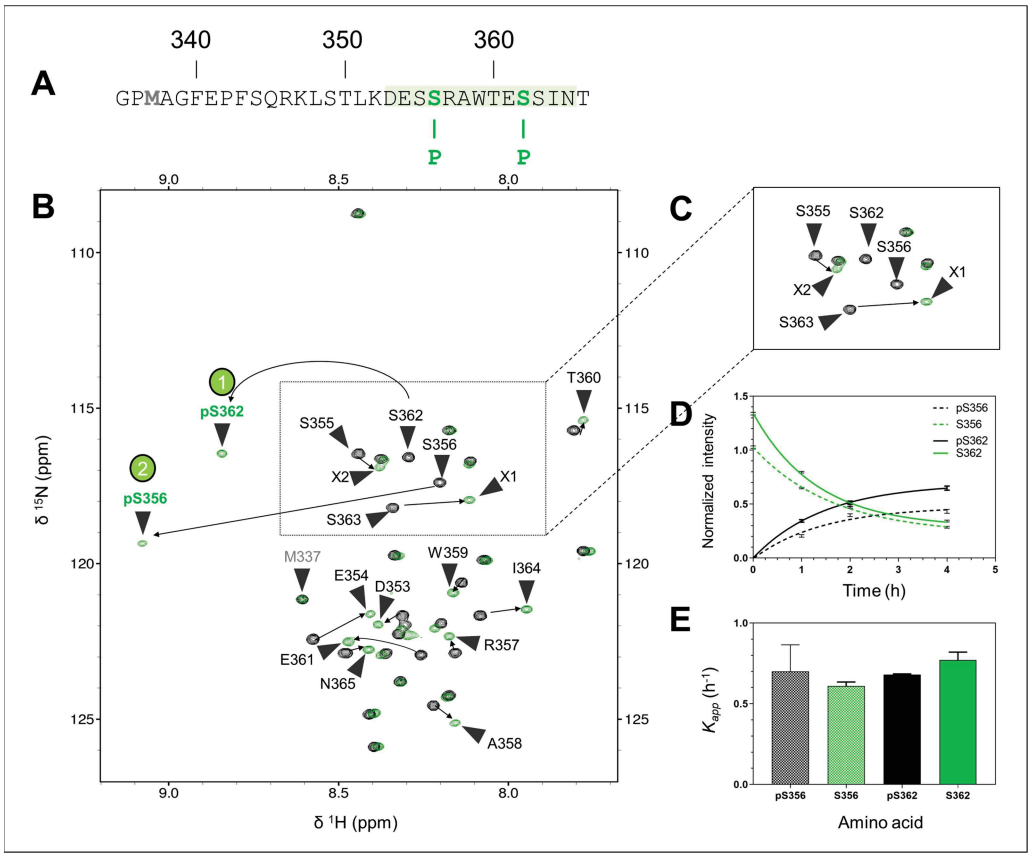


Figure 53 : Attribution du domain C-terminal de GHSR phosphorylé par GRK2. La réaction de phosphorylation a été arrêtée au bout de 9 h, de manière à avoir 100% de phosphorylation pour les résidus pS356 et pS362 (vert). **A)** Séquence de GHSR. Les sérines phosphorylées pS356 et pS362 sont indiquées en vert et les résidus très affectés par ces phosphorylations sont surlignés en vert clair. **B)** Superposition du spectre ^{15}N-HSQC de GHSR phosphorylé GRK2 (vert) au spectre ^{15}N-HSQC de GHSR non-phosphorylé (noir). Les changements de déplacements chimiques des résidus affectés par la phosphorylation (D353 à N366) sont indiqués par les flèches noires. L'ordre d'apparition des sérines phosphorylées est indiqué par les ronds verts contenant le chiffre 1 ou 2. Les résonances de la forme non-phosphorylée de GHSR des résidus S355, S356, S362, S363 ainsi que celles de la forme phosphorylée de GHSR pS356, X2, pS362 et X1 sont indiquées sur le spectre. Le spectre phosphorylé a été enregistré sur un échantillon ^{13}C/^{15}N de 100 µM, dans un tampon de 50 mM Bis-Tris pH 6.7, 50 mM NaCl sur le 800 MHz, à 20 °C. **C)** Zoom de la région des sérines/thréonines issu de l'HSQC en (**A**). **D)** Variation de l'intensité normalisée des résonances de S326 et S362 sous leur forme non-phosphorylée et phosphorylée au cours du temps. **E)** Constantes de vitesse apparentes obtenues à partir des courbes en (**D**).

L'analyse des variations d'intensités des sites phosphorylés dans le temps a permis d'obtenir l'ordre séquentiel des phosphorylations. L'intensité de pS362 est plus élevée au cours du temps que celle de pS356 (Figure 53D). Il semblerait donc que cette sérine soit phosphorylée en premier, alors que les constantes de vitesse apparente d'apparition et de disparition de ces deux sérines sont très similaires : 0.68 ± 0.33 et 0.68 ± 0.01 h^{-1} pour pS356 et pS362 et 0.61 ± 0.05 et 0.77 ± 0.10 h^{-1} pour leur résonance non-phosphorylée (Figure 53E).

4.3.4.1.3 Phosphorylation de la région C-terminale de V2R

Les premiers essais de phosphorylation du domaine C-terminal de V2R ne montraient pas de phosphorylation, soulevant l'hypothèse d'un problème d'accessibilité de la kinase GRK2 au peptide. En effet, V2R contient une cystéine qui pourrait former des ponts disufures inter-peptidiques. Nous avons donc utilisé du TCEP (tris(2-carboxyethyl)phosphine) pour prévenir la formation de ces liaisons lors de la réaction de phosphorylation. Ce test de phosphorylation a montré la présence de quatre pics à bas champs (après 8.7 ppm en ^{1}H) suggérant la présence de sites phosphorylés sur le domaine C-terminale du récepteur V2R par GRK2 (Figure 54). Visuellement, les résonances des résidus T347, T359, S362 ou encore S363 semblent être affectées par la phosphorylation (Figure 54B et C).

Pour essayer d'obtenir plus d'informations sur les sites potentiellement phosphorylés, nous avons essayé de comparer le profil des intensités normalisées des sérines et thréonines de l'échantillon phosphorylé (phospho-V2R-Cter, 5 µM) au profil du V2R-Cter non-phosphorylé (wt-V2R-Cter, 10 µM) (Figure 54D). Par exemple, comparé à la valeur moyenne, S363 a une intensité naturellement plus faible que les autres sérines chez wt-V2R-Cter. Par conséquent, sa faible intensité chez phospho-V2R-Cter n'est probablement pas due à la phosphorylation de ce site. À l'inverse, T359 a une intensité proche de la valeur moyenne chez wt-V2R-Cter alors que chez phospho-V2R-Cter son intensité est bien inférieure à la valeur moyenne. Il est donc possible que ce résidu soit affecté par la phosphorylation. Pour T347 et S362, les intensités de l'échantillon phosphorylé semblent avoir le même profil que celles de l'échantillon sauvage. Il n'est donc pas possible de savoir, de manière catégorique, si elles sont phosphorylées ou non.

Cette expérience préliminaire montre de potentiels sites phosphorylés sur le domaine C-terminal du récepteur V2R par GRK2. De nouveaux essais doivent être effectués pour valider ou non ces sites.

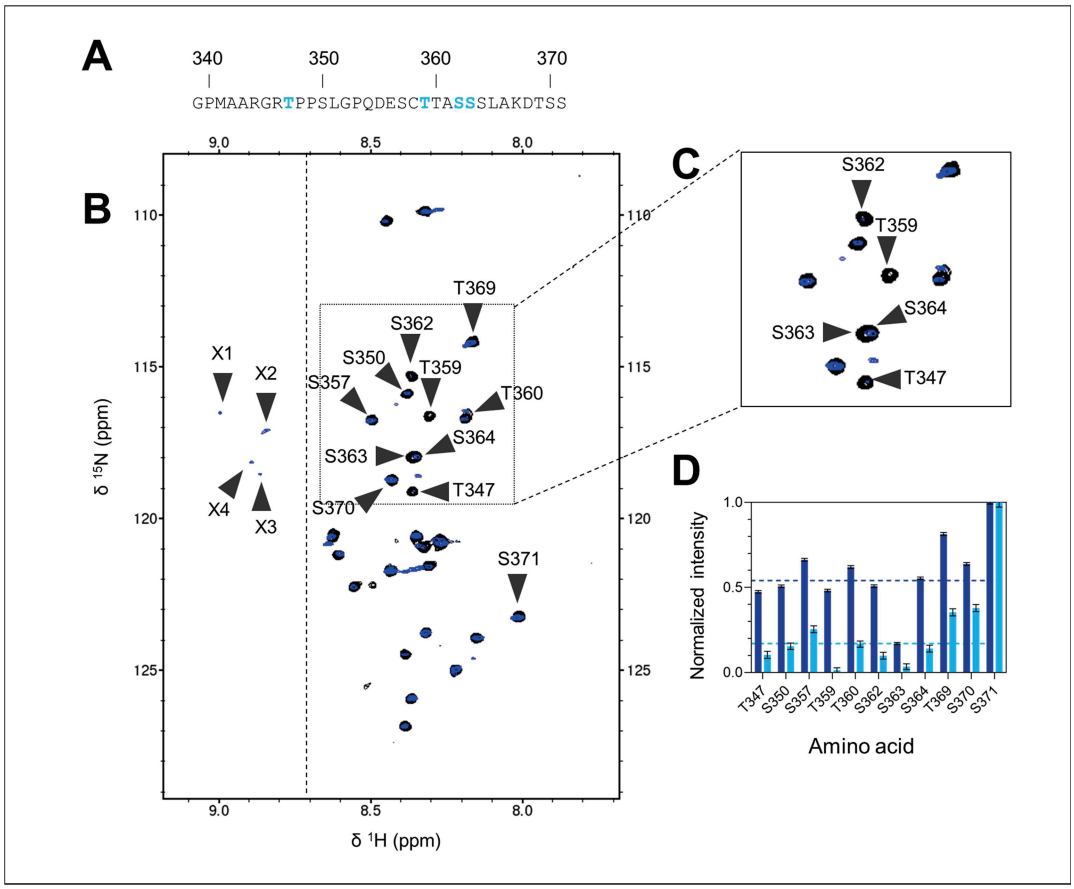


Figure 54 : Visualisation de potentiels sites phosphorylés sur le domaine C-terminal de V2R (V2R-Cter) par GRK2. A) Séquence de V2R-Cter. Les sérines et thréonines dont les résonances ne sont plus visibles dans la protéine phosphorylée sont représentées en bleu. **B)** Superposition du spectre ^{15}N-HSQC de V2R-Cter phosphorylé par GRK2 (bleu) au spectre ^{15}N-HSQC de V2R-Cter non-phosphorylé (noir). Les résonances des sérines et thréonines et des nouveaux résidus à bas champs (après 8.7 ppm, ligne noire pointillée) sont affichées sur le spectre. Le résidu S371 utilisé pour la normalisation des intensités est aussi affiché. Le spectre phosphorylé a été enregistré sur un échantillon ^{15}N de 5 µM, dans un tampon de 50 mM Bis-Tris pH 6.7, 50 mM NaCl sur le 700 MHz, à 20 °C. **C)** Zoom de la région des sérines/thréonines issu de l'HSQC en (**A**). **D)** Comparaison des intensités normalisées des sérines et thréonines chez V2R-Cter sauvage (bleu foncé) et chez V2R-Cter phosphorylé par GRK2 (bleu clair).

4.3.4.2 Comparaison aux sites de phosphorylation connus et prédits

4.3.4.2.1 Phosphorylation de la région C-terminale de β2AR

Dans la littérature, seul le phospho-barcode du récepteur β2-adrénergique (β2AR) a été totalement élucidé (Nobles *et al*, 2011). Pour se faire, Lefkowitz et son équipe ont utilisé une stratégie de 'perte de fonction' basée sur l'utilisation de petits ARNs interférents (siRNA) pour supprimer les fonctions de GRK2 et GRK6. Ainsi, dans les cellules HEK traitées à l'isoprotérénol, l'absence de GRK2 réduit la phosphorylation des sites T360, S364, S396, S401 et S411 et l'absence de GRK6 réduit la

phosphorylation des sérines S355 et S356. Ces différents motifs de phosphorylation permettent respectivement l'internalisation par GRK2, l'activation de la voie ERK par GRK6 et la désensibilisation par GRK2 et GRK6 (Chapitre 1, p. 50, Figure 9).

Dans notre essai, deux pics de phosphorylation sont observés et ils pourraient correspondre aux sérines S345, S346, S356 ou S364. Après une analyse plus approfondie des données de cinétique, il semblerait que les sérines S344 et S345 soient les sérines phosphorylées. Cependant, ces résultats sont à confirmer par acquisition de spectre 3D pour obtenir l'attribution de la protéine phosphorylée. Nous retrouvons donc le site S364 décrit par (Nobles *et al*, 2011) (Figure 55). Concernant le site S356, il est recensé dans PhosphoSiteplus comme étant phosphorylé *in vitro* et *in vivo* par GRK2, alors que ce site n'est pas décrit comme phosphorylé par (Nobles *et al*, 2011).

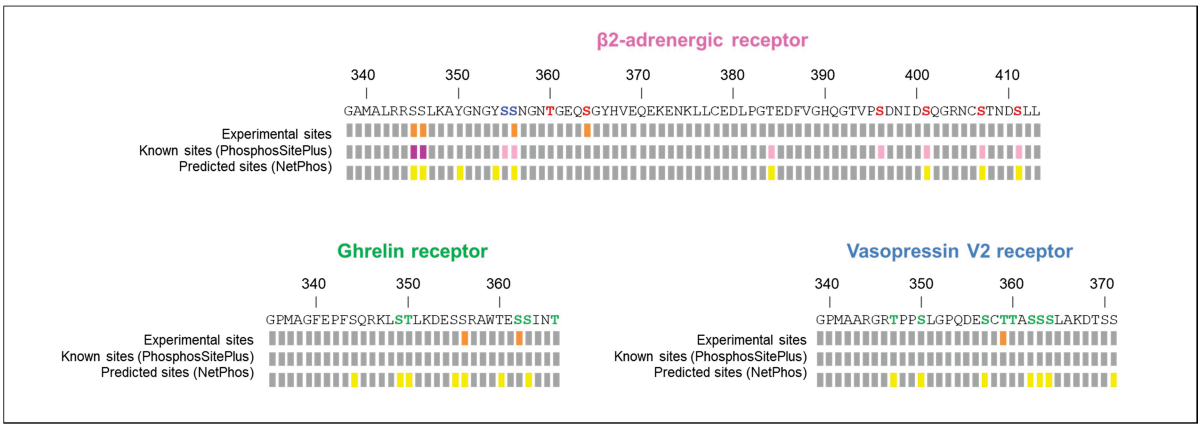


Figure 55 : Comparaison de nos sites phosphorylés déterminés par RMN (orange) aux sites connus et prédits pour être phosphorylés dans les domaines C-terminaux des récepteurs β2AR, GHSR et V2R. Les résidus identifiés comme étant phosphorylés par GRK2 (rouge), GRK6 (bleu) ou par des GRKs non-identifiées (vert) sont colorés dans la séquence selon (Innamorati *et al*, 2001; Nobles *et al*, 2011; Bouzo-Lorenzo *et al*, 2016). Le site PhosphoSitePlus **(Hornbeck *et al*, 2015)** recense les modifications post-traductionnelles décrites expérimentalement chez les protéines de mammifères. Le serveur a été utilisé pour trouver les sites décrits comme étant phosphorylés *in vitro* et *in vivo* par GRK2 (rose) et par la PKA (violet). NetPhos (Blom *et al*, 2004) a été utilisé pour prédire les sites pouvant être phosphorylés par 17 kinases différentes (jaune).

Il n'est pas surprenant de trouver différents sites phosphorylés selon les études car cela dépend de nombreux paramètres expérimentaux comme : l'agoniste utilisé, le temps de stimulation ou encore l'utilisation de molécules pouvant affecter la régulation des GRKs. Par exemple, l'étude de Lefkowitz et son équipe a été effectuée en cellules HEK, stimulées à l'isoprotérénol et en présence de siRNAs,

pour identifier le rôle des différentes kinases (Nobles *et al*, 2011). À l'inverse, Clark et son équipe (Tran *et al*, 2004) ont utilisé des anticorps spécifiques aux sites pS355 et pS356 pour suivre leur phosphorylation en cellules COS-7 (cellules provenant du tissu rénal de singe). Les auteurs ont choisi les cellules COS-7 car leur expression basale en GRKs est naturellement plus faible que pour les cellules HEK. Ils ont ainsi montré que la co-transfection des vecteurs d'expression de β2AR et de GRK2 dans ces cellules augmente par 5 la phosphorylation des sites S355 et S356, suggérant une phosphorylation dépendante de GRK2. Par ailleurs, ils ont aussi montré que le taux de phosphorylation de ces sites dépend fortement de l'agoniste utilisé pour activer le récepteur.

En ce qui concerne les sérines S345 et S346, elles sont prédites et connues pour être phosphorylées par la protéine kinase A (PKA) selon les serveurs NetPhos et PhosphoSitePlus (Figure 55). La PKA et les GRKs font partie de la même famille des kinases AGC et partagent donc des mécanismes d'activation et d'inhibition (Leroux *et al*, 2018). Comparées aux autres kinases AGC, les GRKs ont la particularité de ne phosphoryler que les récepteurs dans leur état actif et d'être impliquées dans la modulation de l'activité des RCPGs (Benovic *et al*, 1986). Curieusement, les résidus S345 et S346 ne sont phosphorylés qu'après activation et ils sont impliqués dans la désensibilisation du récepteur β2AR. D'après Bouvier et son équipe (Moffett *et al*, 2001), cela est dû à la présence d'une cystéine palmitoylée en position C341 qui réduit l'accessibilité aux sérines S345/S346 par ancrage de la région C-terminale de β2AR à la membrane. Dans notre système simplifié de peptides isolés, l'absence d'ancrage à la membrane rend les sérines S345/S346 disponibles et elles peuvent être phosphorylées par GRK2.

4.3.4.2.2 Phosphorylation de la région C-terminale de GHSR

Le code de phosphorylation du récepteur à la ghréline n'a pas été entièrement décrit à ce jour (Bouzo-Lorenzo *et al*, 2016). Pour GHSR, cinq sites phosphorylés ont été identifiés par spectrométrie de masse effectuée sur le récepteur GHSR purifié à partir de cellules HEK stimulées à la ghréline : les sérines S349, S362 et S363 et les thréonines T350 et T366. La phosphorylation de ces sites en réponse à la ghréline suggère fortement l'implication d'une GRK, mais les GRKs impliquées n'ont pas été identifiées. Pour mieux comprendre l'implication de ces différents sites dans la signalisation des RCPGs dépendante de l'arrestine, les auteurs ont créé deux

variants où les résidus S349 et T350 (variant 1) ou S362, S363 et T366 (variant 2) ont été mutés en alanine. Ils ont montré que le variant 2 interagit faiblement avec l'arrestine et que cette faible interaction peut être corrélée à l'incapacité du variant à internaliser ou à activer la voie ERK. Le variant 1 montre un couplage relativement robuste avec l'arrestine comparé au variant 2, mais cette interaction reste plus faible qu'avec le récepteur sauvage. Les auteurs ont ainsi conclu que la région composée des résidus S362, S363 et T366 est impliquée dans la liaison à l'arrestine, l'internalisation du récepteur et l'activation de la voie ERK alors que la région formée des résidus S349 et T350 joueraient un rôle dans la stabilisation du complexe RCPG:arrestine.

Lors de notre essai, la sérine S362 a été identifiée comme phosphorylée par GRK2. Par comparaison avec l'étude de Camiña et son équipe (Bouzo-Lorenzo *et al*, 2016), cela pourrait indiquer que la phosphorylation de GHSR par GRK2 est responsable de la liaison à l'arrestine. Cette hypothèse serait en accord avec la tendance générale démontrant que les GRK2 et -3 sont principalement responsables du recrutement des arrestines (Kim *et al*, 2005; Ren *et al*, 2005).

Concernant le deuxième site phosphorylé identifié par RMN, pS356, il n'est pas connu pour être phosphorylé selon le serveur PhosphoSitePlus (Figure 55). À l'inverse, le serveur NetPhos le prédit comme phosphorylé, mais la kinase à l'origine de cette potentielle phosphorylation n'est pas spécifiée. À ce jour, les sites de phosphorylation par GRK2 du domaine C-terminal de GHSR ne sont pas identifiés. Il est donc envisageable que S356 soit un nouveau site non répertorié jusqu'ici. Il serait intéressant de modifier ce résidu, en alanine par exemple, pour voir son implication dans l'interaction avec l'arrestine et ses fonctions associées.

4.3.4.2.3 Phosphorylation de la région C-terminale de V2R.

Curieusement, le phospho-barcode du récepteur V2R est le moins décrit des trois récepteurs dans la littérature, alors que ce récepteur est très souvent utilisé pour étudier la voie arrestine. En effet, Lefkowitz et son équipe ont été précurseurs du phospho-barcode modèle, en montrant l'implication différenciée des GRKs dans la voie arrestine du récepteur V2R (Ren *et al*, 2005). Pour cela, les auteurs ont utilisé des siRNAs dirigés contre les kinases GRK2, 3, 5 et 6 en cellules HEK. Ils ont ainsi montré que les GRK2 et 3 sont majoritairement responsables de la phosphorylation du récepteur V2R, de sa désensibilisation et du recrutement de l'arrestine. À

l'inverse, les kinases GRK5 et 6 sont moins impliquées dans la phosphorylation du récepteur et dans le recrutement de l'arrestine mais semble soutenir exclusivement l'activation de la voie ERK médiée par l'arrestine. Cependant, aucune information sur les sites phosphorylés et sur les kinases associées à la phosphorylation des potentiels sites n'est amenée dans cette étude.

L'équipe de Birnbaumer (Innamorati *et al*, 1998) a étudié le taux d'intégration de phosphates radioactifs dans le domaine C-terminal du récepteur V2R (V2R-Cter) après stimulation à la vasopressine (AVP) en cellules HEK. Le tronquage progressif du V2R-Cter réduisait le niveau de phosphorylation, ce qui leur a permis de mettre en évidence la présence de plusieurs sites de phosphorylation, probablement 9, dépendants des GRKs. Parmi ces résidus, le cluster de sérines S362, S363 et S364 est phosphorylé et sa troncature ou sa mutation en alanine affecte l'internalisation du récepteur.

L'essai, en présence de TCEP lors de la réaction, nous a permis de visualiser de potentiels sites phosphorylés. Cet essai était très bruité et n'a été effectué qu'une seule fois pour le moment. Lors des essais précédents aucune phosphorylation n'avait été observée et cela avait soulevé l'hypothèse d'un problème d'accessibilité de la kinase au peptide de V2R. L'ajout de TCEP pour empêcher la formation de pont disulfure intra-peptidique durant la réaction de phosphorylation semble avoir réglé le problème. La comparaison des profils d'intensité de V2R-Cter phosphorylé et non-phosphorylé a indiqué que T359 est potentiellement phosphorylé. Cependant, ce site n'est ni prédit ni connu pour être phosphorylé selon les serveurs NetPhos et PhosphoSitePlus (Figure 55).

Ces expériences préliminaires ont permis de mettre au point le protocole de phosphorylation de V2R-Cter par GRK2. De nouveaux essais peuvent maintenant être effectués pour obtenir plus d'information sur les sites phosphorylés par GRK2.

4.3.4.3 Vers un motif consensus de phosphorylation de GRK2 ?

Hormis pour V2R-Cter, nous avons obtenu des informations concernant les sites phosphorylés *in vitro* par GRK2 et nous avons comparé ces sites à ceux prédits et connus dans la littérature. Ainsi, pour β2AR-Cter nous avons identifié deux sites phosphorylés pouvant correspondre aux sérines superposées S345/S364 et/ou S346/S356, et pour GHSR-Cter les S356 et S362 sont phosphorylées par GRK2.

Certains de ces sites sont décrits dans la littérature comme étant phosphorylés et d'autres sont de nouveaux sites non répertoriés jusqu'ici. Nous espérons que ces sites phosphorylés seront confirmés, par spectrométrie de masse sur les récepteurs entiers reconstitués en nanodisque, après phosphorylation *in vitro* avec la kinase GRK2 (collaboration avec Jean-Louis Banères).

La phosphorylation étant effectuée *in vitro* et sur un modèle simplifié (peptide isolé), il est tout à fait possible que les GRKs phosphorylent leur site de prédilection, leur substrat 'préféré', en l'absence de toute forme de régulation extérieure. En effet, il est connu que les kinases ont des préférences pour la nature des résidus entourant le site de phosphorylation : résidus acides, basiques ou hydrophobes. Par exemple, les motifs consensus de phosphorylation de la protéine kinase A (PKA) sont RRX(***S***/***T***) et KRXX(***S***/***T***) où les lettres correspondent au code à une lettre des acides aminés, X correspond à n'importe quel acide aminé et où le résidu phosphorylé est indiqué en gras et en italique (Kang *et al*, 2007). Dans la suite du paragraphe, nous garderons le même code pour les motifs de phosphorylation.

Dans le cas des GRKs, les sites phosphorylés par ces kinases sont peu décrits dans la littérature. Par conséquent, l'identification de leur motif de phosphorylation ne l'est pas non plus. Cependant, des peptides synthétiques ont été utilisés comme substrat pour mesurer l'incorporation de phosphate radioactif et ainsi déterminer les préférences des GRKs (Onorato *et al*, 1991; Kunapuli *et al*, 1994; Asai *et al*, 2014). Ces études ont montré que GRK1 et GRK2 préfèrent les résidus acides autour du site phosphorylé alors que GRK5 et GRK6 préfèrent les résidus basiques (Kunapuli *et al*, 1994; Loudon & Benovic, 1994).

RCPG-Cter	Site phosphorylé	Séquence																
		← N-terminus									C-terminus →							
		-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8
β2AR	S345	/	G	A	M	A	L	R	R	S	S	L	K	A	Y	G	N	G
	S346	G	A	M	A	L	R	R	S	S	L	K	A	Y	G	N	G	Y
	S356	K	A	Y	G	N	G	Y	S	S	N	G	N	T	G	E	Q	S
	S364	S	N	G	N	T	G	E	Q	S	G	Y	H	V	E	Q	E	K
GHSR	S356	L	S	T	L	K	D	E	S	S	R	A	W	T	E	S	S	I
	S362	E	S	S	R	A	W	T	E	S	S	I	N	T	/	/	/	/

Table 13 : Recherche d'un motif consensus de phosphorylation pour GRK2. Les sérines phosphorylées ou potentiellement phosphorylées déterminées par RMN sont colorées en vert. Les acides aminés basiques (rouge) et acides (bleu) les entourant sont aussi colorés.

En utilisant cette stratégie d'étude Kang et son équipe (Asai *et al*, 2014, 2016) ont analysé l'activité de GRK2 envers plus de 17 substrats peptidiques. Il en ressort que GRK2 préfère les résidus acides positionnés à 1, 2, 3 ou 4 résidus en amont et à -1 résidu en aval du site phosphorylé. Malheureusement parmi les 17 substrats, seul deux correspondent à des peptides dérivés de domaines C-terminaux de RCPGs, l'un provient du récepteur β2AR et l'autre provient du récepteur à la rhodopsine.

Expérimentalement, nous retrouvons bien des résidus acides en amont des sites phosphorylés pS356 et pS362 de GHSR-Cter et du site potentiellement phosphorylé pS364 de β2AR-Cter (Table 13). Curieusement les sites S345, S346 et S356 de β2AR-Cter sont précédés de résidus basiques et non acides. Ainsi, aucun consensus ne ressort de ces données expérimentales. Pour pouvoir élucider les motifs consensus des différentes GRKs, il semble nécessaire d'identifier les résidus phosphorylés sur de nombreux domaines C-terminaux des RCPGs. Pour cela, la RMN peut s'avérer être un outil adapté, permettant d'identifier la phosphorylation multiple des domaines C-terminaux des RCPGs (Landrieu *et al*, 2006).

4.3.5 La phosphorylation module le repliement des RCPG-Cters et leur région d'interaction à l'arrestine-2

Comme présenté au début de cette partie, il reste à élucider si les peptides phosphorylés par GRK2 présentent 1) un phénomène de repliement similaire aux phosphomimétiques ? 2) un mode d'interaction à l'arrestine-2 similaire aux phosphomimétiques ?

1) Pour répondre à la première question, nous avons comparé les profils CSI obtenus pour la région C-terminale de GHSR dans sa forme sauvage (wt-GHSR-Cter) et dans sa forme phosphorylée (phospho-GHSR-Cter). Le profil CSI du peptide phospho-GHSR-Cter présente des valeurs négatives des résidus 352 à 360. Cela suggère une conformation étendue dans cette région. À l'inverse, cette région contient une hélice chez wt-GHSR-Cter (Figure 56). Ainsi la transition structurale d'une hélice en brin bêta opère chez les peptides phosphorylé et phosphomimétique (pm-GHSR-Cter). Le changement de structuration de pm-GHSR-Cter était localisé dans la région contenant le plus de résidus phosphomimétiques. Ici, le changement de structuration de phospho-GHSR-Cter est observé dans la région contenant les sérines phosphorylées. Ces résultats montrent l'importance de la phosphorylation

et/ou des charges dans la régulation de la conformation des domaines C-terminaux des RCPGs.

2) Un phénomène de repliement similaire à celui présent chez les phosphomimétiques a lieu lors de la phosphorylation du C-terminus de GHSR par la GRK2. Qu'en est-il de l'impact sur son interaction avec l'arrestine-2 ? Pour répondre à cette question, nous avons effectué des interactions par RMN.

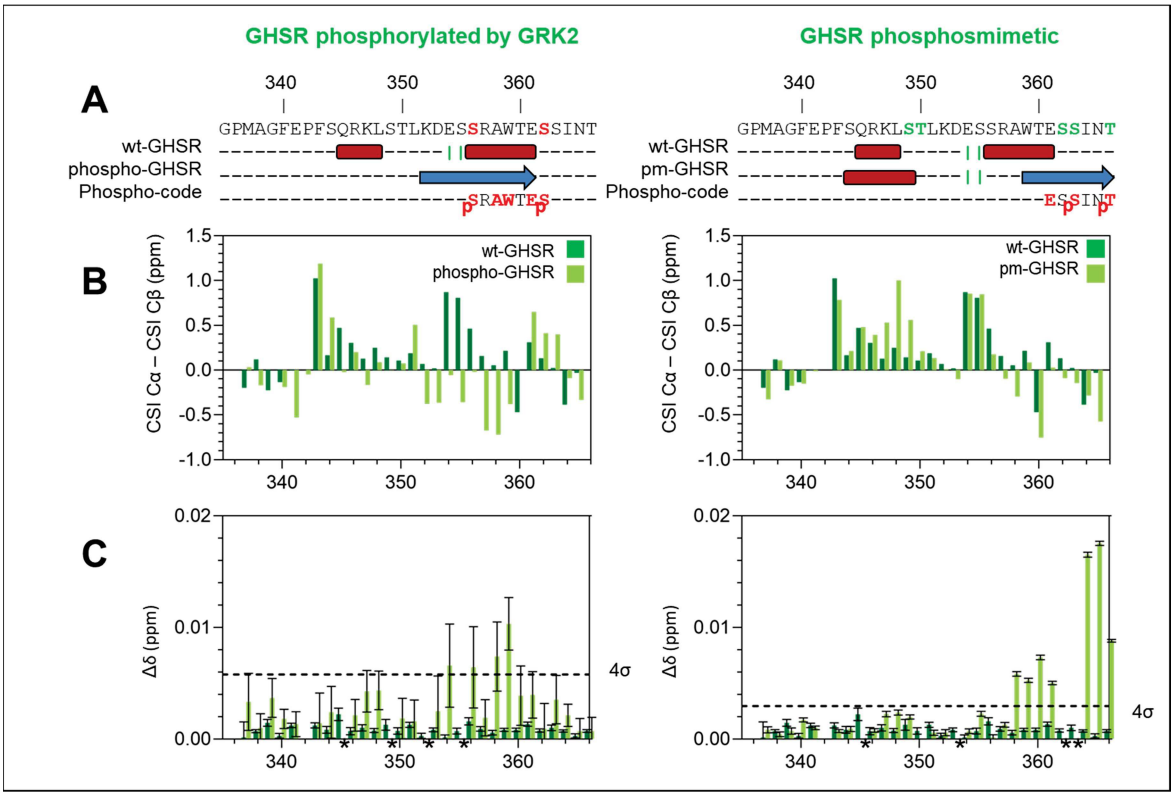


Figure 56 : Effet de la phosphorylation de GRK2 sur la conformation de la région C-terminale de GHSR (phospho-GHSR-Cter, colonne de gauche) **et sur son interaction avec l'arrestine-2, et comparaison avec les résultats du phosphomimétiques** (pm-GHSR-Cter, colonne de droite). **A)** Séquence de GHSR. Les sérines phosphorylées par GRK2 sont indiquées en rouge et les acides glutamiques remplaçant les sérines/thréonines sont colorées en vert. Sous les séquences sont représentés les profils de structuration de wt-GHSR-Cter (non-phosphorylé), de phospho-GHSR-Cter et de pm-GHSR-Cter obtenus par consensus RMN. Le profile décrit pour GHSR-Cter phosphorylé par GRK2 est basé uniquement sur le CSI. Les motifs de phosphorylation prédit par PhosCoFinder (Zhou *et al*, 2017) sont présentés en rouge : le code court pSRAWTE ou long pSRAWTEpS et le code ESpSINpT, où pS et pT sont des phospho-sérines et phospho-thréonines. **B)** Histogramme de l'index de déplacement chimique (CSI) ^{13}Cα/^{13}Cβ de wt-GHSR-Cter (vert foncé, à gauche) et phospho-GHSR-Cter (vert clair) ou pm-GHSR-Cter (vert clair, à gauche). **C)** Histogramme de la différence de déplacement chimique ^{1}H/^{15}N (ppm) observée par interaction entre l'arrestine-2 et wt-GHSR-Cter (vert foncé, à droite) ou phospho-GHSR-Cter (vert clair, à droite) ou pm-GHSR-Cter (vert clair). Les cut-off sont notés en ligne noire pointillée (4σ). Les résidus superposés dont les valeurs n'ont pas pu être mesurées sont représentés par des étoiles (*).

Tout comme pour les précédentes interactions, des spectres ^{15}N-HSQCs en absence et en présence d'arrestine (ratio peptide:arrestine = 1:10) ont été enregistrés et les différences de déplacement chimique et d'intensité ont été calculées. Des changements de déplacement chimique ainsi que des baisses d'intensités sont observées dans la région allant de S356 à E361 (Figure 56C et Figure S7). Cette région d'interaction correspond à la région autour des sérines phosphorylées, ainsi qu'à la région contenant le brin bêta identifié précédemment. Chez le phosphomimétique, la région d'interaction correspondait aussi à la région contenant les résidus phosphomimétiques structurés en brin bêta (Figure 56).

Par ailleurs, nous avons discuté dans l'article 2 que le repliement pourrait permettre l'interaction avec des poches spécifiques de l'arrestine, pour activer différentes fonctions (Chapitre 4, Article 2, p. 160). Le brin bêta de la région d'interaction de phospho-GHSR-Cter semble deux fois plus étendu (des résidus 352 à 361) que celui de la région de pm-GHSR-Cter (des résidus 361 à 366). Ainsi, ces deux régions pourraient former des interactions différentes avec les résidus de l'arrestine.

Zhou et ses collaborateurs (Zhou *et al*, 2017) ont proposé un motif de phosphorylation (ou phospho-code) qui servirait de région d'interaction de haute affinité avec l'arrestine (Chapitre 1, p. **Erreur ! Signet non défini.**). La liaison à l'arrestine se ferait grâce au motif p[S/T]Xp[S/T]XXp[S/T]/E/D (motif court) ou p[S/T]XXp[S/T]XXp[S/T]/E/D (motif long) (définition Chapitre 1, p. **Erreur ! Signet non défini.**). L'absence d'un des trois résidus chargés du motif (p[S/T] ou p[S/T]/E/D), donnerait lieu à des motifs incomplets, ou partiels, qui auraient une affinité plus faible pour l'arrestine. Tout comme pour les variants phosphomimétiques (Chapitre 4, Article 2, p. 160), la région d'interaction de GHSR-Cter phosphorylé par GRK2 avec l'arrestine, contient un motif de phosphorylation prédit par ((Zhou *et al*, 2017), PhosCoFinder) : le motif court pSRAWTE ou le motif long pSRAWTEpS (Figure 56). Ce motif est incomplet à cause de la présence de l'alaline ou du tryptophane en son centre. Par ailleurs, la région d'interaction du variant phosphomimétique de GHSR-Cter avec l'arrestine, contient le motif complet ESpSINpT (Figure 56 ; Chapitre 4, Article 2, p. 160). Cela peut expliquer que l'interaction semble plus faible pour phospho-GHSR-Cter que pour pm-GHSR-Cter. En effet, de faibles variations de déplacement chimiques sont trouvés pour

l'interaction phospho-GHSR:arrestine-2, alors que le double est mesuré pour l'interaction pm-GHSR:arrestine-2.

En résumé, notre étude valide que la phosphorylation par les GRKs module la conformation des domaines C-terminaux des RCPGs et leur région d'interaction avec l'arrestine. De plus, les variants phosphomimétiques présentent des mécanismes d'interaction et de repliement en brin bêta similaires à ceux de GHSR-Cter phosphorylé par GRK2. Par conséquent, le phosphomimétisme est une bonne stratégie pour l'étude des complexes RCPG:arrestine et de leur signalisation dépendante de l'arrestine.

Il sera intéressant d'obtenir les sites de phosphorylation par GRK6 et d'effectuer une étude similaire pour valider ses hypothèses et analyser comment les sites de phosphorylation différentiels affecteraient l'interaction.

4.4 PARTIE 4 : PROBLEMATIQUE ET RESUME

Étude du réarrangement structural global de l'arrestine

Dans cette thèse, nous avons pour objectif de déchiffrer les mécanismes moléculaires impliqués dans l'interaction des régions C-terminales des RCPGs (RCPG-Cters) avec l'arrestine-2. Dans les parties précédentes du Chapitre 3, nous avons étudié cette interaction du point de vue des RCPGs. Nous avons mis en évidence que la phosphorylation modifie la conformation des RCPG-Cters et permet leur interaction avec l'arrestine-2. Dans cette partie, nous nous intéressons au réarrangement global que subit l'arrestine lors de sa liaison aux RCPG-Cters. Pour cela, nous avons étudié la conformation de l'arrestine seule (sauvage, forme inactive), de l'arrestine phospho-indépendante R169E (forme pré-active), et de ces arrestines en complexe avec les RCPG-Cters (forme active), par SAXS. Les résultats montrent des réarangements subtiles de l'arrestine entre les formes inactive et pré-active, et suggèrent un changement de dynamique. Nous avons aussi regardé les changements de conformation de l'arrestine sauvage en complexe avec les peptides des RCPGs par FRET, grâce à notre collaboration avec l'équipe de Jean-Louis Banères.

En résumé :

Le SAXS et les expériences de digestion trypsique limitée montrent des réarangements subtiles de l'arrestine entre les formes inactive et pré-active, et suggèrent un changement de dynamique.

Le FRET permet de suivre le réarrangement global de l'arrestine lors de sa complexation avec les peptides C-terminaux des RCPGs.

Dans cette partie, j'ai contribué à la production des protéines arrestines et des RCPG-Cters. Les mesures et les analyses SAXS ont été effectuées par Assia Mouhand, post-doctorante de Nathalie Sibille (Mouhand *et al*, en préparation). Les analyses FRET ont été effectuées par Jean-Louis Banères et son équipe.

4.4.1 Analyse de la conformation de l'arrestine wt et R169E seule

Le variant R169E est un variant qui 'mime' une conformation dite 'pré-active' de l'arrestine, qui est largement utilisé dans la communauté des RCPGs (Chapitre 2, p. 71, Figure 19). Dans cette conformation pré-active, la destruction du cœur polaire de l'arrestine engendrerait un détachement de la queue C-terminale. On s'attend donc à un changement de conformation entre les arrestines wt et R169E. Pour tester cette hypothèse, nous avons effectué une expérience de digestion trypsique limitée (Figure 57A et B). La trypsine est une enzyme qui clive la liaison peptidique adjacente à une arginine ou à une lysine. Un changement de conformation d'une protéine peut masquer ou démasquer ces sites de clivages et modifier le profil de protéolyse limitée (Xiao *et al*, 2004).

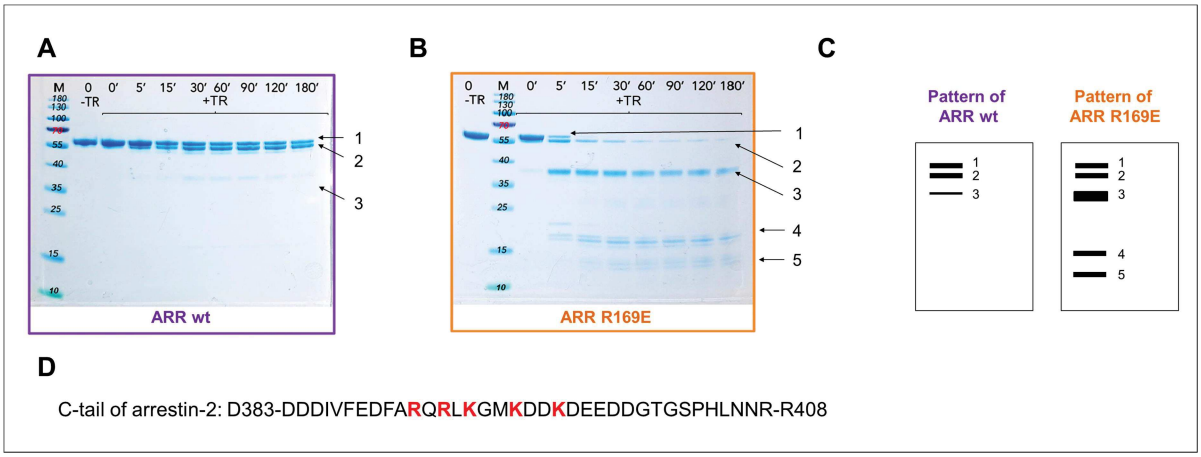
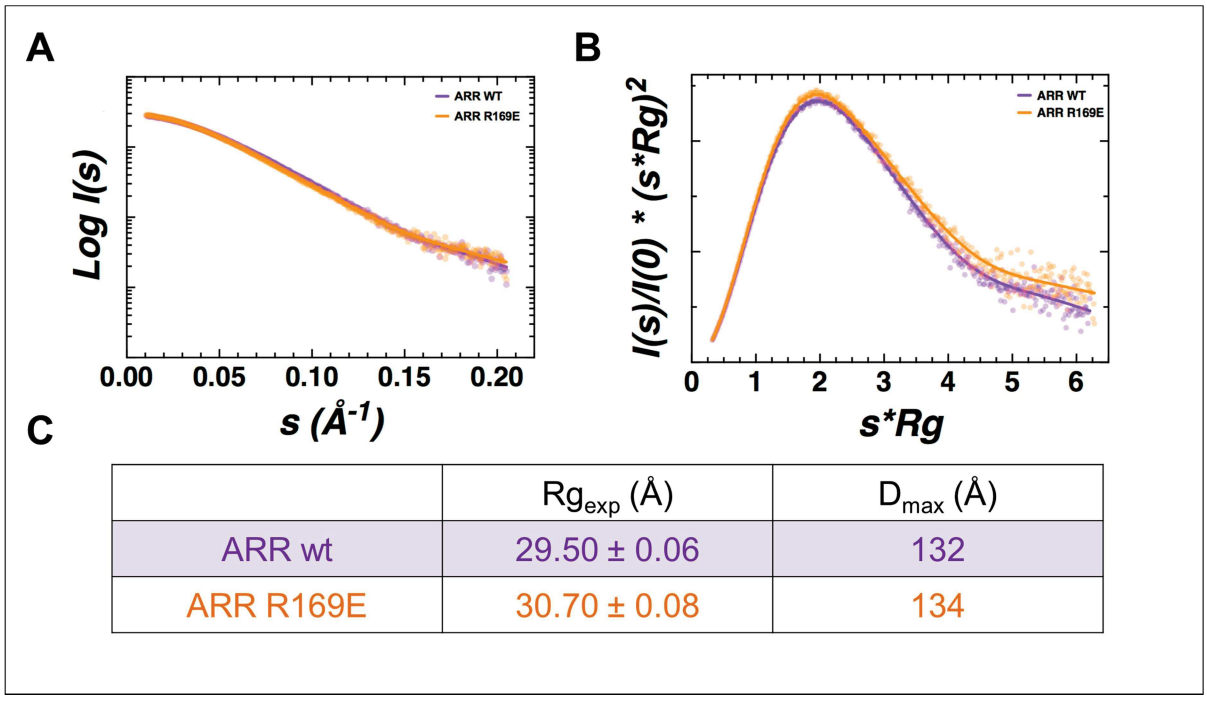


Figure 57 : Expérience de digestion trypsique limitée de l'arrestine sauvage (wt, A) et R169E (B). Les arrestines à (0.5 mg/mL) ont été digérées par la trypsine (+TR) (ratio 1:500 trypsine:arrestine) pendant 0, 5, 15, 30, 60, 90, 120 ou 180 min. Les échantillons clivés ont été déposés sur gel SDS-PAGE et les protéines ont été visualisées par coloration au bleu de Coomassie. Les différentes bandes observées sont numérotées sur la droite des gels (1 à 5). Un échantillon contrôle, sans trypsine, a aussi été déposé sur les gels (-RT) **C)** Représentation schématique du profil de digestion à la trypsine des arrestines wt et R169E. **D)** Séquence de la queue C-terminale de l'arrestine-2. Les lysines et arginines pouvant servir de site de clivage à la trypsine sont représentées en rouge.

La digestion trypsique limitée des arrestines wt et R169 montrent des profils différents (Figure 57A, B et C). La bande correspondant à l'arrestine wt non clivée (bande 1) est présente quel que soit le temps d'incubation avec la trypsine. De faibles bandes 2 et 3 commencent à apparaître au bout de 15 et 30 min. À l'inverse, la bande correspondant à l'arrestine R169E non clivée (bande 1) disparaît au bout de

15 min. Elle laisse place à de nombreuses nouvelles bandes, dont des bandes 4 et 5 qui ne sont pas présentes dans le profil de l'arrestine wt. Ces résultats montrent que l'arrestine R169E est dégradée beaucoup plus rapidement que l'arrestine wt par la trypsine. Il serait possible que la rapide dégradation de R169E soit associée à la dégradation de la queue C-terminale de l'arrestine. Cette région est 'libérée' du domaine N-terminal de l'arrestine durant son activation (Chapitre 1, p. 40, Figure 2). Elle pourrait donc être plus accessible à la trypsine dans la conformation pré-active de l'arrestine. De plus, cette région contient de nombreuses lysines et arginines (Figure 57D) qui pourraient servir de site de clivage à la trypsine.

Les expériences de digestion limitée à la trypsine confirment que les conformations des arrestines wt et R169E ne sont pas identiques et que R169E possèdent des sites de clivages plus exposés, ce qui serait en accord avec une forme plus 'active'de l'arrestine. Ensuite, nous avons analysé ces deux arrestines en SEC-SAXS (colonne d'exclusion couplée au SAXS) pour essayer de visualiser ces changements conformationnels.



	Rg_{exp} (Å)	D_{max} (Å)
ARR wt	29.50 ± 0.06	132
ARR R169E	30.70 ± 0.08	134

Figure 58 : Analyse par SEC-SAXS des arrestines sauvage (wt, violette) et R169E (orange) : A) Courbes SAXS (5 mg/mL), **B)** Kratky plots normalisés, **C)** Rayons de gyrations R_g et distances maximales (D_{max}).

Les courbes SAXS de ces deux arrestines sont très similaires (Figure 58), ce qui indique qu'il n'y a pas de réarangements structuraux importants. Par ailleurs, la différence de Rg de 1 Å est significative et indique probablement un déplacement d'équilibre de dynamique entre les formes inactive et pré-active.

Les analyses SAXS et de digestion limitée suggèrent donc que les réarrangements structuraux de l'arrestine-2 entre les formes inactives et pré-actives (R169E) sont subtils. Il est intéressant de noté que la même conclusion a été émise par l'équipe de Hargrave (Shilton *et al*, 2002) lors d'une étude par SAXS des conformations de l'arrestine-1 visuelle sauvage et R175Q (équivalent du variant R169E chez l'arrestine-2).

4.4.2 Analyse de la conformation de l'arrestine wt et R169E en complexe

4.4.2.1 Caractérisation globale des complexes par SAXS

Dans la partie précédente, nous avons montré que le SAXS met en évidence un changement de dynamique entre les formes inactives (ARR wt) et pré-active (R169E) de l'arrestine. Nous avons ensuite enregistré des données SEC-SAXS de ces arrestines en complexes avec le peptide du domaine C-terminal du récepteur à la ghréline (GHSR-Cter). Le choix du peptide a été basé sur, d'une part, l'affinité des peptides pour l'arrestine (Chapitre 4, Article 2, p. 160) et, d'autre part, sur leur rendement de production (Chapitre 2, p. 79, Table 7). Le peptide phosphomimétique de V2R est celui qui a l'affinité la plus forte pour l'arrestine. Cependant, son rendement de production est le plus faible des variants phosphomimétiques. Nous avons donc sélectionné le variant phosphomimétique de GHSR, qui a une plus faible affinité que V2R-Cter mais meilleure que β2AR-Cter, et qui présente un bon rendement de production comparé au peptide de V2R.

Dans un premier temps, nous avons enregistré les données SEC-SAXS de l'arrestine sauvage (wt) seule ou en complexe avec le variant phosphomimétique de GHSR-Cter en condition classique (le tampon de SEC ne contient pas le peptide de GHSR-Cter) (Figure 59A). En complexe, les deux protéines sont éluées séparément et présentent un volume d'élution de ~ 2 mL pour l'arrestine et de ~ 2.5 mL pour le peptide de GHSR, et les profils des Kratky plots sont similaires. Il est possible que dans ces conditions, la séparation des partenaires soit due à l'interaction transitoire

du C-ter avec l'arrestine, ainsi nous ne detectons pas d'interaction à cause de la faible affinité de nos complexes (Chapitre 4, Article 2, p. 160).

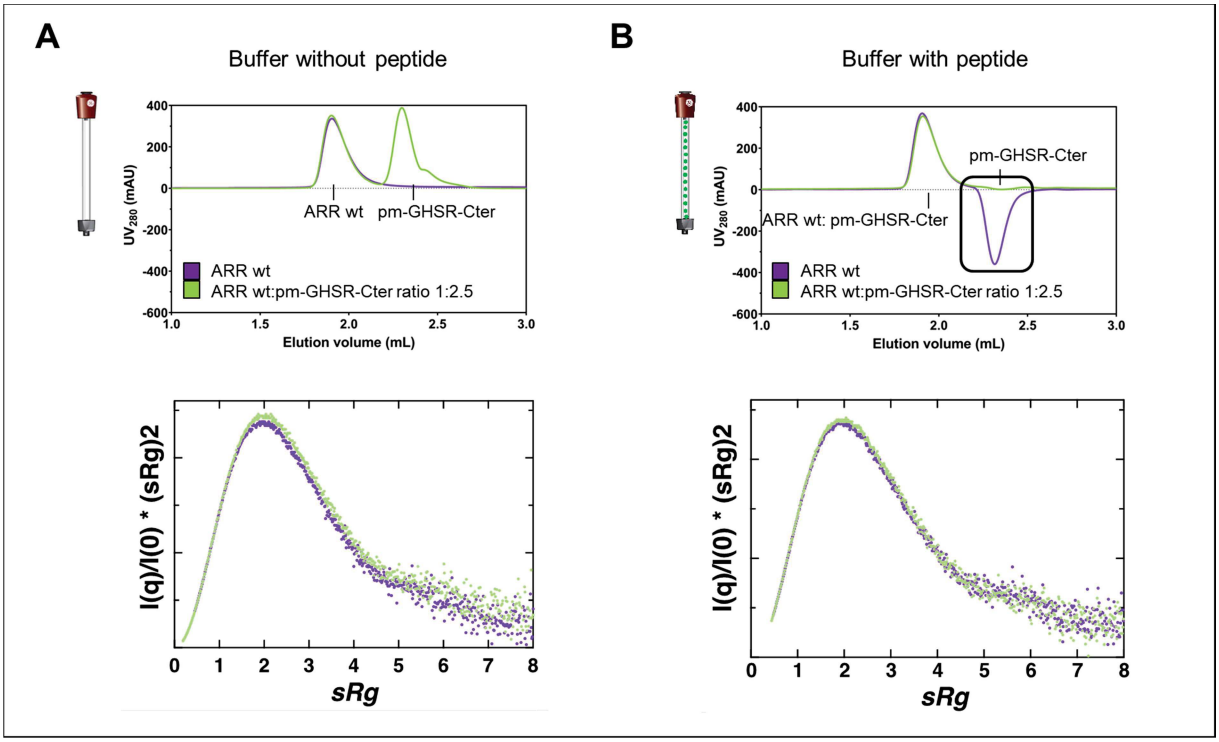


Figure 59 : Analyse par SEC-SAXS de l'arrestine-2 en complexe avec le variant phosphomimétique de GHSR-Cter (pm-GHSR-Cter), sans saturation (A) ou avec saturation (B) de la colonne en peptide. Les courbes supérieures représentent les profils d'élution à 280 nm issus de la chromatographie d'exclusion (SEC). La déplétion observée lors de l'interaction de l'arrestine avec les peptides du tampon est encadrée en noir. Les courbes inférieures correspondent aux Kratky plots normalisés de l'arrestine seule (violet) ou en complexe avec le peptide phosphomimétique de GHSR (vert).

Il existe une méthode peu commune pour étudier des complexes transitoires et/ou de faible affinité par chromatographie d'exclusion. Il s'agit de la méthode de Hummel-Dreyer (Hummel & Dreyer, 1962; Phizicky & Fields, 1995), dans laquelle le ligand est présent dans le tampon utilisé pour la chromatographie. On parle de chromatographie en condition saturante. Dans ces conditions, l'interaction entre une protéine d'intérêt et un ligand va entraîner une augmentation de la concentration de ligand dans la fraction d'élution de la protéine d'intérêt. À l'inverse, elle va engendrer une diminution de la concentration en ligand dans sa propre fraction d'élution. Ainsi, une déplétion de l'absorbance à 280 nm est observée (exemple Figure 59). Nous avons reproduit cette stratégie pour étudier l'arrestine (protéine d'intérêt) en complexe avec le peptide phosphomimétique de GHSR (ligand) par SEC-SAXS.

Pour cela, nous avons injecté l'arrestine seule ou en complexe sur une colonne saturée en peptide phosphomimétique de GHSR-Cter (Figure 59B). Le profil de l'arrestine seule (courbe violette) montre une déplétion à ~ 2.5 mL qui correspond à une diminution de la concentration en peptide de GHSR-Cter à son volume d'élution. Il y a donc une interaction entre l'arrestine et le peptide contenu dans le tampon. De plus, si l'on injecte un échantillon contenant de l'arrestine et le peptide à ratio 1:2.5, on remarque que la déplétion n'est plus présente (courbe verte) et que le signal est plat. Cela signifie que l'arrestine est déjà saturée en peptide, lors de son injection. Malgré la confirmation d'une interaction saturante entre l'arrestine wt et le peptide, les Kratky plots normalisés sont toujours similaires. Là encore, il semblerait qu'il n'y est pas de grand changement conformationnel entre les formes inactives (ARR wt) et active (en complexe) de l'arrestine.

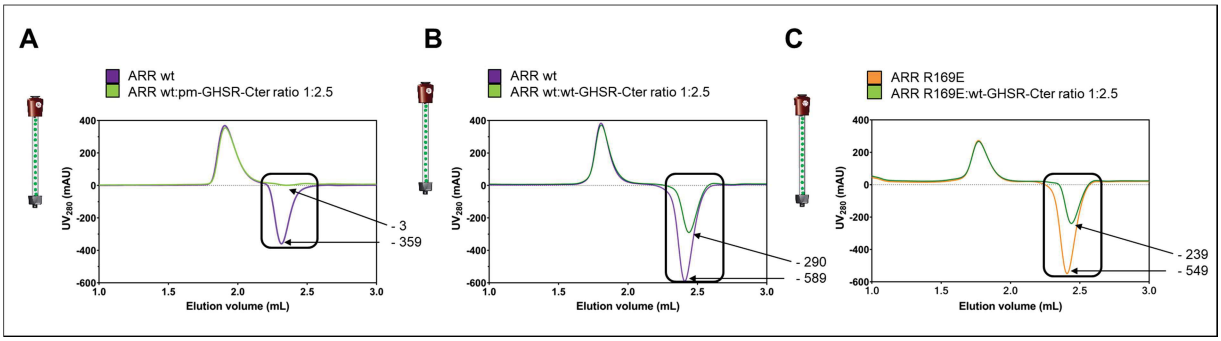


Figure 60 : Interaction entre l'arrestine sauvage (wt) ou R169E et le peptide de GHSR-Cter (sauvage, wt ; phosphomimétique, pm). Dans toutes ces expériences, la colonne a été équilibrée dans un tampon contenant 180 µM de peptide GHSR-Cter. Les déplétions observées au volume d'élution du peptide sont encadrées en noir. La figure présente les profils d'élution à 280 nm des arrestines seules wt (**A** et **B**, violet) ou R1695 (**C**, orange) et en complexe avec GHSR phosphomimétique (pm, **A**, vert) ou sauvage (wt, **B** et **C**, vert). Les flèches noires indiquent les valeurs d'absorbance maximale des dépletions en peptide de GHSR-Cter.

La même expérience de saturation sur colonne d'exclusion de taille (SEC) a été réalisée pour les arrestines wt ou R169E en complexe avec le peptide GHSR sauvage (wt-GHSR-Cter) (Figure 60). Dans cette expérience, l'injection de l'échantillon contenant l'arrestine wt ou R169E en complexe avec le peptide sauvage de GHSR (Figure 60B et C) au même ratio que le peptide phosphomimétique de GHSR (Figure 60A), induit une déplétion des peptides du tampon au lieu d'un signal plat. Ainsi, à ratio 1:2.5, les arrestines ne sont pas saturées. La comparaison des arrestines wt er R169 aurait nécessité de faire plusieurs ratios, mais de façon

relative, il semblerait que la déplétion soit moins prononcée pour l'arrestine R169E par rapport à l'arrestine wt (Figure 60B et C). Ainsi, il faudrait plus de peptides non-phosphorylés pour saturer l'arrestine wt que pour saturer le variant R169E. Cela est en accord avec la littérature étant donné que le variant R169E interagit mieux avec les RCPGs sauvages, non-phosphorylés que l'arrestine wt (Kovoor *et al*, 1999).

Ces résultats montrent qu'il existe une interaction basale avec les peptides non-phosphorylés, mais que cette interaction est plus forte lorsque le domaine C-terminal des RCPGs est chargé négativement, ce qui est en accord avec nos analyses RMN (Chapitre 4, article 2, p.).Cette interaction peut être amélioré lorsque l'arrestine est dans une conformation pré-active, comme celle de la R169E.

4.4.2.2 Caractérisation globale des complexes par FRET

Nous avons vu que les changements conformationnels impliqués dans l'activation de l'arrestine semblent être trop subtils pour être détectés par SAXS. Grâce à notre collaboration avec Jean-Louis Banères, nous avons pu étudier ces changements conformationnels par FRET (transfert d'énergie de fluorescence par résonance de type Förster). A cet effet, le double variant V167C/L191C sans aucune autre cystéine présente dans la séquence protéique de l'arrestine-2, a été produit par l'équipe de Jean-Louis Banères. Ces deux cystéines ont été marquées par un fluorophore donneur (Alexa350) et un fluorophore accepteur (Alexa488), via le groupement maléimide des Alexa. Il est connu que l'activation de l'arrestine entraîne la rotation du domaine C-terminal par rapport au domaine N-terminal (Chapitre 1, p. 40, Figure 2). C'est pourquoi la seconde sonde (C191) a été placée à l'extrémité du domaine C-terminal de manière à suivre l'activation de l'arrestine par FRET (Figure 61A). De plus, l'efficacité de FRET dépend de la distance séparant les deux sondes. La paire de fluorophores choisie présente une efficacité de 50 % lorsque la distance entre les deux sondes est de ~ 50 Å. Ces sondes sont donc bien adaptées aux caractéristiques géométriques de l'arrestine, qui présente une distance de ~ 52 Å entre les deux résidus marqués.

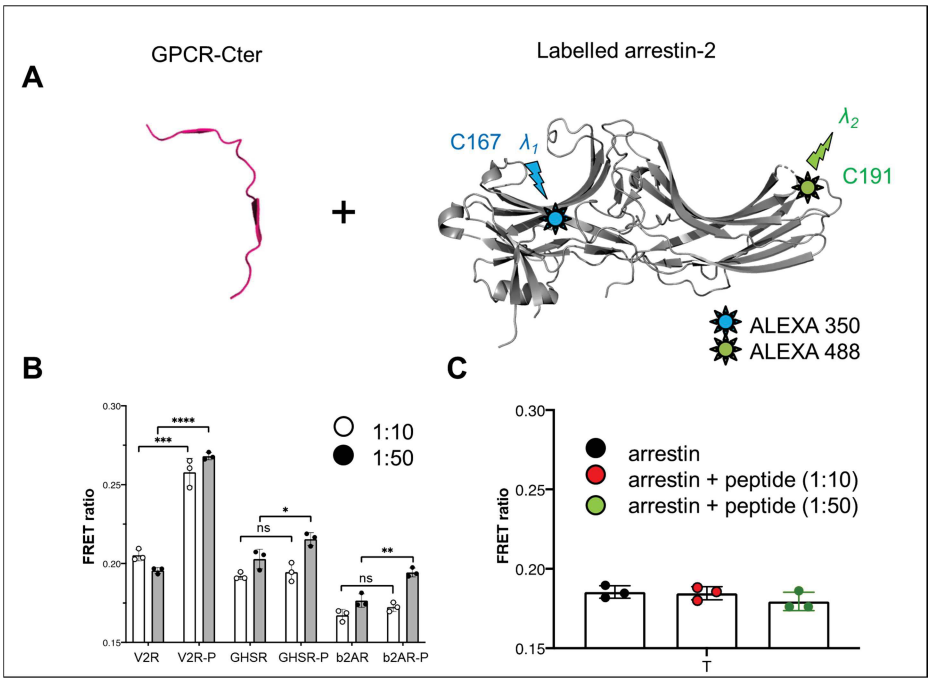


Figure 61 : Expériences de FRET avec ou sans les peptides C-terminaux des RCPGs. A) Représentation schématique de l'expérience de FRET. Le variant de l'arrestine a été marquée avec deux fluorophores différents, en position C158 et C330, pour suivre les changements conformationnels qu'elle subit lors de l'interaction. **B)** Ratio FRET mesuré lors de l'addition des peptides des RCPGs, à ratio molaire 1:10 ou 1:50 (RCPG-Cter:arrestine). Les changements significatifs sont représentés par les étoiles (*). Les domaines C-terminaux sont nommés comme suit : V2R, domaine C-terminal non-phosphorylé du récepteur V2R ; V2R-P, domaine C-terminal phosphomimétique du récepteur V2R ; GHSR, domaine C-terminal non-phosphorylé du récepteur GHSR ; GHSR-P, domaine C-terminal phosphomimétique du récepteur GHSR ; b2AR, domaine C-terminal non-phosphorylé du récepteur β2AR ; b2AR-P, domaine C-terminal phosphomimétique du récepteur β2AR. **C)** Ratios FRET mesurés lors de l'addition d'un peptide contrôle (ligand du récepteur nucléaire M3) n'interagissant pas avec l'arrestine. Dans cette expérience le taux de marquage de l'arrestine a été estimé à 20 %.

L'analyse des expériences avec ce couple de fluorophore a montré des différences de distances de FRET entre l'arrestine complexée aux domaines C-terminaux sauvage (wt) et phosphomimétique (pm) (Figure 61B), à ratio molaire 50:1 (RCPG-Cter:arrestine). À l'inverse, la même expérience avec un peptide contrôle connu pour ne pas interagir avec l'arrestine (ligand du récepteur nucléaire M3) ne montre aucune différence de FRET (Figure 61C). Ces résultats montrent que l'arrestine interagit avec les peptides phosphomimétiques, en subissant des changements conformationnels. Ainsi, il est possible de suivre le réarrangement de l'arrestine par FRET, en utilisant les peptides des RCPGs. L'utilisation de sondes à différentes positions nous permettra d'obtenir une image globale des réarrangements conformationnels de l'arrestine en fonction des différents peptides.

5 CONCLUSION

5.1 CONCLUSION GENERALE

Les Récepteurs couplés aux protéines G (RCPGs) sont des protéines transmembranaires impliquées dans la transduction du signal. Leur activation donne lieu à une grande diversité de réponses cellulaires affectant ainsi de nombreuses fonctions biologiques (régulation du rythme cardiaque, de l'homéostasie hydrique, de la faim, *etc*…). Ainsi, la famille des RCPGs est la famille la plus ciblée par les molécules pharmacologiques présentes sur le marché (Hauser *et al*, 2017). La signalisation des RCPGs est complexe car elle peut activer à la fois des voies de signalisation dépendante et indépendante des protéines G. Cette dernière résulte du couplage du récepteur à une grande famille de protéines de signalisation, les arrestines. Bien que la signalisation dépendante de l'arrestine soit une composante majeure du fonctionnement des RCPGs, nous commençons tout juste à en saisir les bases mécanistiques, même pour les RCPGs les mieux étudiés.

Le modèle actuel d'interaction RCPG:arrestine est le modèle du phospho-barcode. Dans ce modèle, l'activation du récepteur par un ligand particulier mène au recrutement de kinases spécifiques des RCPGs, les GRKs. Les GRKs phosphorylent différentes sérines et thréonines du domaine C-terminal du RCPG. Selon le motif de phosphorylation, la conformation de l'arrestine liée au domaine C-terminal du RCPG varie, ce qui module ses capacités fonctionnelles (Nobles *et al*, 2011). Cependant, les démonstrations expérimentales de ce mode de fonctionnement sont encore rares.

Fait intéressant, les domaines C-terminaux des RCPGs semblent présenter toutes les caractéristiques structurales, dynamiques et fonctionnelles des régions/protéines intrinsèquement désordonnées (IDRs/IDPs) (Jaakola *et al*, 2005). Les IDPs/IDRs font partie d'une famille de protéines dont le rôle spécifique dans les processus fondamentaux de signalisation et de régulation a commencé à être révélé il y a une vingtaine d'années (Wright & Dyson, 2015).

La majeure partie de mon travail de thèse a été d'étudier l'effet de la phosphorylation des régions C-terminales des RCPGs sur la signalisation cellulaire dépendante de l'arrestine. Pour cela, nous avons utilisé la kinase GRK2, produite au laboratoire, ainsi que des variants phosphomimétiques des domaines C-terminaux. Pour que ce travail soit représentatif, nous avons travaillé sur trois RCPGs de la

classe A : le récepteur β2-adrénergique (β2AR), le récepteur à la ghréline (GHSR) et le récepteur vasopressine V2 (V2R). Ces trois récepteurs appartiennent à deux classes différentes d'interactants vis-à-vis de l'arrestine (Oakley *et al*, 2000), et ils sont également des cibles pharmaceutiques importantes. Techniquement, la caractérisation structurale, dynamique et fonctionnelle des domaines C-terminaux des RCPGs a été faite en utilisant la résonance magnétique nucléaire en solution (RMN), couplée à la diffusion des rayons X à petits angles (SAXS) et à la spectroscopie de fluorescence (FRET).

Nous avons commencé cette étude par la caractérisation structurale par RMN des domaines C-terminaux non-phosphorylés. Nous avons montré que ces domaines sont des régions intrinsèquement désordonnées (IDRs) et qu'ils contiennent des structures secondaires transitoires.

La caractérisation structurale des variants phosphomimétiques montrent la présence d'un brin bêta qui n'est pas présent dans l'état basal des domaines C-terminaux. Ainsi, une transition structurale est induite par les mimes de la phosphorylation. Cette transition structurale a aussi été retrouvée pour le domaine C-terminal du récepteur GHSR phosphorylé par GRK2 et reste à être démontrée pour les peptides C-terminaux phosphorylés de V2R et β2AR. Cela montre que la phosphorylation régule la conformation des domaines C-terminaux des RCPGs. Ce phénomène de transition structurale dépendante de la phosphorylation est un mécanisme connu des IDRs pour réguler leur interaction avec des partenaires spécifiques (Bah *et al*, 2015).

Suite à ces analyses structurales, nous avons évalué les capacités fonctionnelles de nos peptides à interagir avec l'arrestine. Nous avons montré par RMN que les peptides phosphorylés et phosphomimétiques interagissent avec l'arrestine alors que les peptides non-phosphorylés n'interagissent pas. Cette observation était attendue car il est attesté que la phosphorylation par les GRKs est nécessaire à l'interaction (Gurevich & Gurevich, 2019a).

Dans les variants phosphomimétiques, les acides glutamiques sont en interaction avec l'arrestine. De même, les sérines phosphorylées par GRK2 du peptide de GHSR-Cter sont impliquées dans la liaison à l'arrestine. Ainsi, l'interaction a lieu là où les charges négatives sont présentes. De plus, la région d'interaction de chaque peptide phosphomimétique, et de GHSR phosphorylé par GRK2, correspond à la

région contenant le brin bêta. Cela est en accord avec les structures cystrallographiques et de cryo-microscopie éléctronique qui montrent que le domaine C-terminal de V2R forme un brin bêta lorsqu'il est lié à l'arrestine-2 (Shukla *et al*, 2013; Nguyen *et al*, 2019; Staus *et al*, 2020). Cette conformation est aussi adoptée par le domaine C-terminal du récepteur Rhodopsine (Rho) en interaction avec l'arrestine-1 (Zhou *et al*, 2017). Dans ces structures, les domaines C-terminaux de V2R et Rho interagissent avec des poches conservées de l'arrestine, nommées A, B et C (Chapitre 4, Article 2, p. 160). De plus, un phospho-peptide synthéthique du domaine C-terminal du récepteur CXCR7 (C7pp) en interaction avec l'arrestine-3, forme une boucle allongée qui interagit avec une poche en plus, P, de l'arrestine (Chapitre 4, Article 2, p. 160). Il est possible que la phosphorylation par les GRKs dicte la conformation adoptée par les domaines C-terminaux des RCPGs et leur interaction avec des poches particulières de l'arrestine. Ainsi, la conformation des RCPG-Cters pourrait être reconnue de manière spécifique par différents sous-types d'arrestines. Finalement, nous avons montré que la conformation du variant phosphomimétique de V2R ne change pas lorsqu'il est en interaction avec l'arrestine (Chapitre 4, Article 2, p. 160).

À la lumière de l'ensemble de ces données, nous proposons un mécanisme moléculaire qui complète une étape manquante du phospho-barecode modèle, en montrant que la phosphorylation par les GRKs régule la conformation des domaines C-terminaux des RCPGs et leur interaction à l'arrestine (Figure 62). Dans ce mécanisme, la phosphorylation par des GRKs spécifiques donnerait une conformation particulière au domaine C-terminal des RCPGs. Le repliement des régions C-terminales des RCPG serait reconnu par l'arrestine, et mènerait à la sélection d'une conformation spécifique de l'arrestine (sélection conformationnelle). Cette organisation structurale de l'arrestine favoriserait l'exposition d'une deuxième surface d'interaction pour un partenaire cytosolique particulier, déclanchant une voie de signalisation spécifique.

Pour finir, les variants phosphomimétiques montrent des mécanismes d'interaction et de repliement en brin bêta similaires à ceux de GHSR-Cter phosphorylé par GRK2. Nous pouvons donc conclure que les phosphomimétiques sont de très bons modèles de phosphorylation pour les régions C-terminales des RCPGs, et que cette

stratégie peut être utilisée pour étudier la signalisation des RCPGs dépendante de l'arrestine.

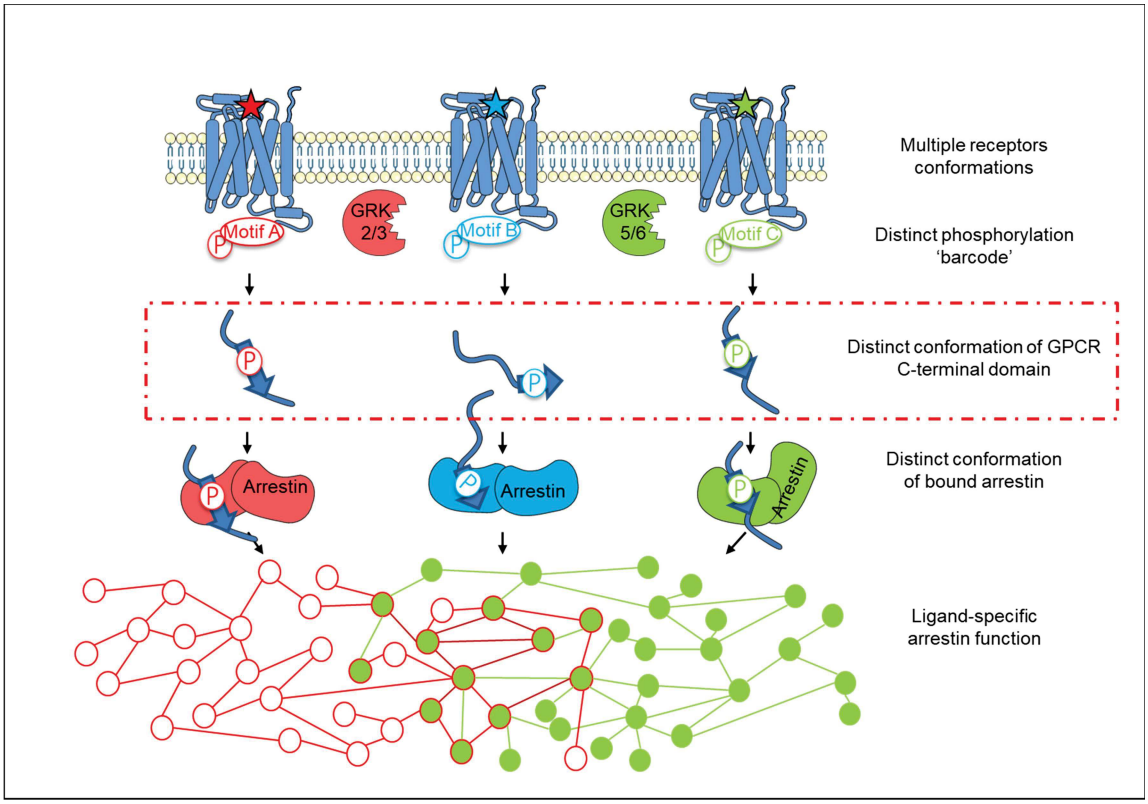


Figure 62 : La phosphorylation par les GRKs régule la conformation des domaines C-terminaux des RCPGs et leur interaction à l'arrestine. En conclusion de nos résultats nous proposons une étape supplémentaire au phospho-barecode modèle (encadrée en rouge), où la phosphorylation de différents motifs de phosphorylation par les GRKs induirait un changement de conformation des domaines C-terminaux des RCPGs. Ces changements seraient reconnus par l'arrestine qui serait alors stabilisée dans une conformation particulière (sélection conformationnelle), ce qui déclencherait différentes fonctions de l'arrestine.

5.2 PERSPECTIVES

Nous proposons un mécanisme moléculaire qui compléterait le phospho-barcode modèle (Figure 62). Cependant, plusieurs points restent à finaliser. Tout d'abord, la caractérisation des formes phosphorylées des RCPG-Cter par les kinases GRK2 et/ou GRK6 permettra de mettre en évidence les motifs de phosphorylation de chaque kinase, seule et en combinaison. Pour chacun de ces peptides phosphorylés, leur conformation pourra être étudiée par RMN et les sites d'interaction avec l'arrestine pourront être identifiés. Cela permettra de mettre en relation les motifs de phosphorylation de chaque kinase, le repliement des RCPG-Cters et leur région

d'interaction avec l'arrestine. Ces peptides pourront aussi servir à la cartographie des différents réarrangements conformationnels de l'arrestine par FRET. Ainsi, les surfaces d'interaction de l'arrestine seront mises en évidence et il sera possible de développer des inhibiteurs des surfaces d'interactions spécifiques à chaque voie de la signalisation arrestine. Le développement de ces molécules pourrait mener à une amélioration des médicaments actuels, en limitant les effets secondaires.

L'étude de la phosphorylation des récepteurs entiers reconstitués en nanodisque par les kinases GRK2 et GRK6 permettra, d'une part de vérifier les codes de phosphorylation déterminés sur les peptides, et d'autre part d'attribuer un phospho-code spécifique au récepteur entier à chaque kinase. Après vérification, il sera possible de s'affranchir de la phosphorylation *in vitro*, en utilisant des variants phosphomimétiques. En effet, nous avons montré que ces variants sont de bons modèles pour étudier la phosphorylation des RCPGs. À l'avenir, les variants pourraient permettre d'étudier plus facilement la voie arrestine et mener à l'obtention de structures de complexe RCPG:arrestine par cryo-miscorcopie électronique.

Ainsi, nous aurons décrit la voie arrestine dans un modèle simplifié composé de peptides C-terminaux isolés de RCPGs, d'arrestines, de kinases purifiées et de récepteurs entiers assemblés dans des nanodisques. Il sera ensuite nécessaire d'aller étudier cette signalisation dans un contexte cellulaire. En effet, il a été proposé que la signalisation des RCPGs soit régulée de manière différentielle selon le type de tissu et de cellule, et que le mécanisme de cette régulation soit dépendant de la phosphorylation par les GRKs (Tobin *et al*, 2008). Cela pourrait être réalisé par une expression différentielle des GRKs (Chapitre 1, p. 49, Figure 8), ou par des mécanismes de régulation de ces kinases, voire même par de l'épissage alternatif (l'épissage alternatif est un processus qui permet, à partir d'une séquence génomique unique, de produire plusieurs ARNs messagers correspondant à des protéines distinctes). En effet, très récemment, le Professeur Dunker et son équipe ont proposé que l'épissage alternatif des régions désordonnées, comme les domaines C-terminaux des RCPGs, puisse être un nouveau mécanisme de régulation de leur signalisation (Zhou *et al*, 2018). Ils ont montré que parmi les régions subissant de l'épissage alternatif, 31 % d'entre elles contiennent des sites de modification post-traductionnelles (PTM) qui pourraient être altérés. Sachant que l'épissage alternatif a souvent lieu dans des régions désordonnées, il pourrait créer

des motifs alternatifs de phosphorylation qui changeraient la signalisation cellulaire. Ainsi, la complexité de la signalisation des RCPGs augmenterait encore, avec une dimension dépendante du contexte cellulaire.

ANNEXES

Annexe 1

3. Nuclear receptors and their transcriptional coregulators

3.1 Biological context

Human nuclear receptors (NRs) constitute a family of 48 transcription factors that plays an essential role in regulating the expression of a multitude of genes that control most of fundamental biological processes such as development, organ homeostasis, metabolism, immune function, or reproduction.[151-153] This regulatory capacity of NRs occurs through their ability to recognize specific DNA sequences promoters of their target genes and their relationships with the general transcription machinery. NRs respond directly to a large variety of hormonal and metabolic substances that are hydrophobic, lipid soluble, and of small size (*e.g.* retinoic acid, estradiol, thyroid hormone, vitamin D3, etc.). They are also the target for numerous post-translational modifications including phosphorylation, SUMOylation, and acetylation.[154] They have a conserved modular structure with a variable unstructured N-terminal domain (NTD) that hosts the activation function 1 (AF-1), a well conserved zinc-finger DNA-binding domain (DBD) and a C-terminal moderately conserved ligand-binding domain (LBD) (Fig. 5). This multifunctional C-terminal domain (CTD) is responsible for ligand binding or dimerization and contains a ligand-dependent activation function (AF-2), which corresponds to transcriptional coregulator interaction surfaces that can be modulated by natural, pharmacological or environmental ligands.[155-157] The LBD also contributes to the modulation of the N-terminal AF-1 through interdomain crosstalk so that both NTD/AF-1 and CTD/AF-2 domains can recruit a range of coregulatory proteins and act individually or in a synergistic manner.[158-161] Most NRs bind to DNA as homodimers or as heterodimers with the retinoid X receptor (RXR), the obligate heterodimerization partner of more than one third of the NR family members (Fig. 5).

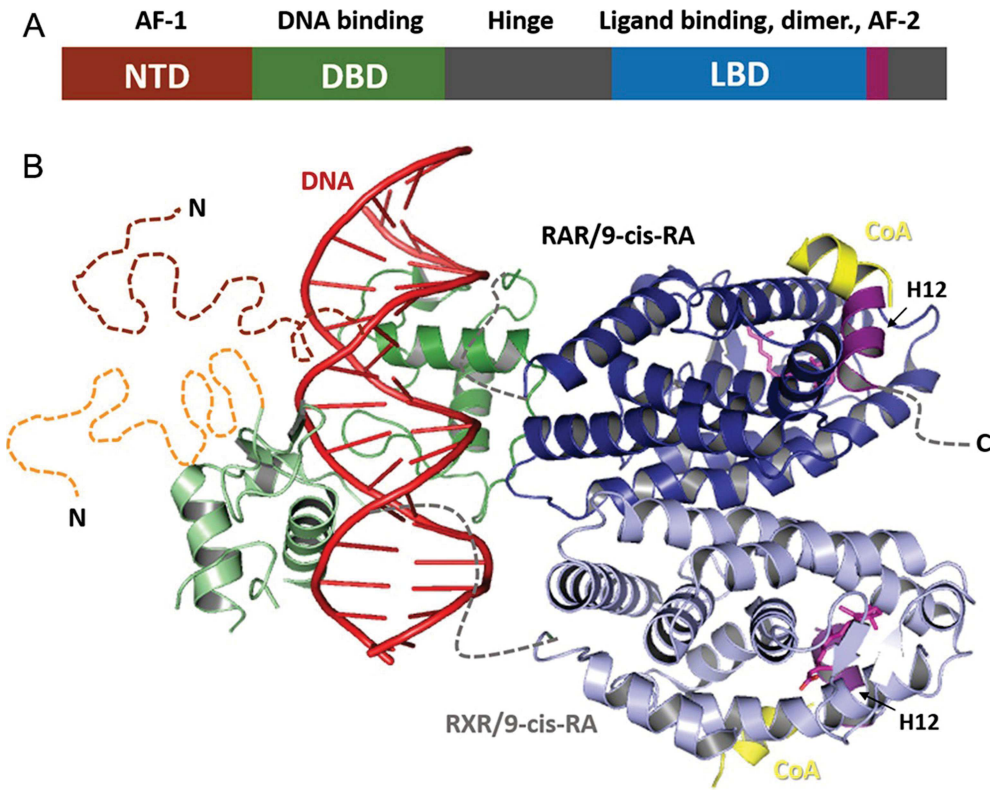


Fig. 5 Common organization of nuclear receptors. (A) Functional organization of NRs. The N-terminal domain (NTD) contains a ligand-independent activation function (AF-1). The DNA-binding domain (DBD) mediates sequence-specific DNA recognition. The ligand-binding domain (LBD) mediates ligand binding, dimerization and a ligand-dependent transactivation function (AF-2). The hinge region corresponds to a linker allowing the proper orientation of the DBD and LBD within DNA-bound dimers. (B**)** Crystal structure of RARβ/RXRα heterodimer bound to DNA.[162] LBDs (dark and light blue) and DBDs (dark and light green) are represented as ribbons. Helix H12 (AF-2) of each monomer is highlighted in purple. The coactivator peptides interacting with the heterodimer are drawn in yellow and the agonist ligands (9-cis-RA) in each subunit are drawn as pink sticks. Dotted lines denote regions with unresolved structures.

The activity of NRs depends on a diverse group of proteins termed “coregulators” that affect the transcriptional machinery in different ways and more than 350 coregulators have been reported to date. NRs may act either as repressors or activators of gene transcription depending on their ligation status that in turn determines the ability of DNA-bound receptors to recruit so-called “corepressors” or “coactivators” to target gene promoters. Coactivators and corepressors are in fact general terms defining the components of high molecular weight coregulator complexes.[163,164] They correspond to large groups of proteins with a diverse range of protein activities and enzymatic functions whose interactions with NRs and many

other transcription factors provide a complex regulatory framework for controlling the transcription of target genes. Coactivator complexes contribute to the enhancement of transcription by acetylating histones, a process that plays an important role in the opening of chromatin during transcription activation, whereas corepressors display the opposite activity by recruiting histone deacetylases. Coactivator recruitment is usually ligand dependent, whereas corepressors interact in most cases with unliganded receptors. Coactivators can be subdivided into two groups. Secondary coactivators represent a subgroup of molecules that are constituents of multisubunit coactivator complexes and that also contribute to the enhancement of NR-mediated transcription, but that do not directly contact the NRs. Primary coactivators, such as those of the SRC-1/TIF-2/RAC3 (p160) family, mediate the interaction of coactivator complexes with NRs. CBP, p300, P/CAF, and some p160 coactivators themselves are reported to act as histone acetyltransferases (HATs).[165,166] The corepressors NCoR and SMRT have been shown to reside in, or recruit, large complexes with deacetylase activity that maintain histones in a deacetylated form, a status that is associated with silent regions of the genome.[167,168] Therefore, histone acetylation and deacetylation shuffle nucleosomal targets between a condensed and relaxed chromatin configuration, the latter being requisite for transcriptional activation.

In the absence of their cognate ligand, some NRs are located in the nucleus, bind to the DNA response elements of their target genes, and recruit corepressors, whereas others are located in the cytoplasm in an inactive complex with chaperones. Ligand binding induces major structural alterations of receptor LBDs leading to (1) destabilization of corepressor or chaperone interfaces, (2) exposure of nuclear localization signals allowing nuclear translocation and DNA binding of cytoplasmic receptors, and (3) recruitment of coactivators triggering gene transcription through chromatin remodeling and activation of the general transcription machinery. Both NRs and their coregulators have been shown to frequently contain large intrinsically disordered regions that mediate their mutual interaction and play a pivotal role in transcription regulation. These unstructured proteins or domains may be functional, undergoing transitions to more ordered states or folding into stable secondary or tertiary structures upon binding to DNA response elements and protein partners. Below we review recent studies that shed light on the mechanisms governing

signaling by these flexible domains and how they can be exploited for the development of new endocrine-based therapies.

3.2 The NTD/AF-1 of NRs and its interactions with transcriptional coregulators

The NTD of NRs harbors an autonomous transcriptional activation function termed AF-1 that, when linked to a heterologous DNA-binding domain, can activate transcription in a constitutive manner, however, in the context of full-length receptors AF-1 is silent in the absence of ligand,[169-170] suggesting an allosteric communication between the N- and C-termini of NRs. When comparing NRs from different sub-families, the sequence of NTDs is highly variable, revealing a very weak evolutionary conservation. Their variable length, from a few residues like in the constitutive androstane receptor (CAR, 10 amino acids) or the vitamin D receptor (VDR, 24 amino acids) up to more than 500 residues in steroid hormone receptors such as the glucocorticoid receptor (GR, 420 amino acids), the androgen receptor (AR, 558 amino acids), the progesterone receptor (PR, 566 amino acids), and the mineralocorticoid receptor (MR, 602 amino acids), also suggests distinct transcriptional activities and biological roles.[171] Interestingly, the length of this domain has been positively correlated with the activity of the AF-1 for different members of the NR superfamily,[172] and logically, the NTD accounts for a major proportion of the total transcriptional activity of all steroid hormone receptors where AF-1 is the predominant activation domain over that of AF-2 lying in the C-terminal LBD.[161,173-176] Moreover, the NTD is subjected to alternative splicing and differential promoter usage so that the majority of NR isoforms differs in their N-terminal region. Notably, the production of translational isoforms differing only in the length of the disordered region was recently proposed as a mechanism by which NRs such as the GR can tune their transcriptional activity.[177,178] The NTD through its activation function AF-1 is involved in multiple protein-protein interactions that are critical for the modulation of the transcriptional activation of target genes in a cell-specific and promoter dependent manner.[171,179-182] As with the AF-2 domain, coactivators such as CREB-binding protein (CBP) or the p160 proteins SRC-1 and TIF-2 can also interact with and increase the transcriptional activity of the AF-1 domain.[183-185] However, unlike the LXXLL (X stands for any amino acid) binding motif for AF-2 interactions (see below for more details), no such conserved motif is known for the AF-1 binding. Note

however that the NTD of AR contains the FXXLF motif (amino acids 23–27) which was shown to mediate a physical interaction between the AF-1 and the AF-2.[172] In the majority of NRs, the NTD is subjected to post-translational modifications, such as phosphorylation and sumoylation, that play essential roles in regulating the receptor transcriptional activities.[186-188]

In contrast to the DBD and the LBD bearing the C-terminal AF-2, and due to the intrinsically disordered nature of the NTD, the structural basis for AF-1 activity are not very well defined and only now begin to be revealed. Structural disorder is a key structural feature for transactivation domains in members of the NR superfamily[50,176,178,187,189,190] as well as many other transcription factors[191] so that their interactions with transcriptional coregulators are among the best characterized examples of complexes involving IDPs.[192-194] The structural plasticity of these domains is thought to allow (1) sampling the environment until appropriate binding partners are found and (2) specific interactions through induced conformational changes upon binding to targets, to (3) result in transient/low affinity complex formation and dynamic signaling regulation and (4) facilitate post-translational modifications.[45,195] Knowledge about NR NTDs was hampered by the technical challenges of characterizing IDPs.[196,197] Only limited structural information was available from circular dichroism (CD) spectra and chemical shifts from nuclear magnetic resonance (NMR) spectroscopy showing that isolated NTDs are largely unstructured with poorly resolved chemical shifts, and with a high random coil content and minimal secondary structural elements.[182,198-200] In the last decade, complementary biophysical techniques such as small-angle X-ray scattering (SAXS), fluorescence resonance energy transfer (FRET), and hydrogen exchange mass spectrometry (HDX-MS) in combination with computational approaches such as molecular dynamics (MD) simulations have confirmed the previous observations and allowed studies in the context of full-length receptors, thereby revealing unforeseen structural features of NR NTDs.

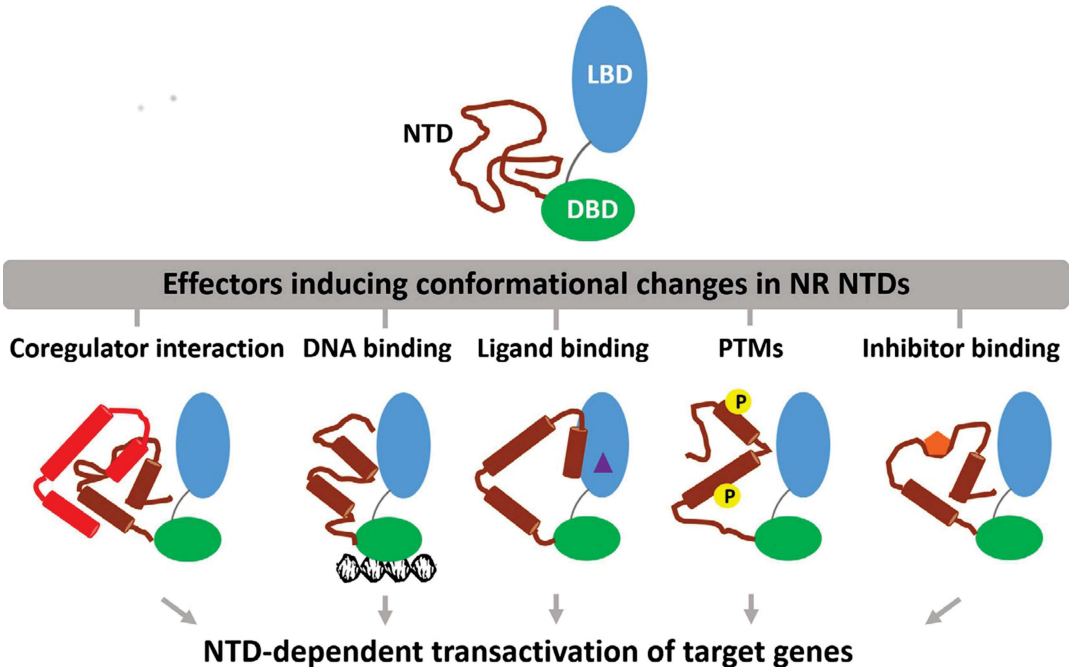


Fig. 6 NR NTDs respond to various stimuli. The non-conserved NTD of NRs is disordered but can acquire higher order structure in response to coregulator interaction, DNA binding, ligand binding to LBD, post-translational modifications (PTMs) and inhibitor binding. The NTD can adopt a multitude of conformations in response to these different signals, which ultimately translates into NTD-dependent transactivation of target genes specific of each receptor (adapted from [201]).

Such analyzes of full-length peroxisome proliferator-activated receptor gamma (PPARγ),[202] PR[189] and retinoid X receptor alpha (RXRα)[203,204] have further revealed that the disordered nature observed for the isolated NTDs is also exhibited in the context of intact NRs, indicating that this property is not abrogated via interdomain interactions. Experiments performed in the presence of secondary structure (α-helices and β-strands) stabilizing agents such as the osmolyte trimethylamine-N-oxide (TMAO) or the hydrophobic solvent trifluoroethanol (TFE),[198,200,205-208] interacting proteins,[161,182,184,189,206,208-210] or DNA50,[199,211,212] clearly demonstrated that the NTDs undergo partial folding under these circumstances and adopt some degree of α-helical conformation. Furthermore, these conformational alterations were shown to have functional consequences since when residues involved in the formation of putative α-helices were mutated, the ability of the AF-1 to form such secondary structures was remarkably reduced and the transactivation activity of this domain decreased, suggesting that the formation of α-helices is important for AF-1 to mediate gene activation[172,200,206,213,214] (Fig. 6). This was best illustrated for the case

of PR whose NTD contains a large content of random coil and is capable of adopting secondary helical structure in the presence of TMAO. HDX-MS confirmed the high disorder in the NTD within the context of full-length PR, but also showed that it can adopt a more stable tertiary folding through a direct interaction with the TATA-binding protein (TBP).[189] Mapping experiments and point mutation analysis revealed that a central region of the NTD between residues 350 and 428 appeared to be sufficient for inducing secondary structure in the presence of TBP. These data suggested that short helical element(s) embedded within the larger disordered NTD may constitute a recognition site for TBP that can induce its folding upon binding. Beyond the formation of secondary structures upon interaction, it was further shown that while mostly disordered, the NTD of NRs might be more compact than expected for an IDP and exist in a premolten globule conformation,[215,216] and that long-range metastable contacts could explain this compactness. A recent study of the estrogen receptor alpha (ERα) using SAXS, hydroxyl radical footprinting and MD simulations to generate a representative ensemble of structures identified a constitutive intra-domain interaction between two small 5-residue motifs centered around residues L33 and S118 in the NTD of this receptor.[217] In full agreement, the addition of a chemical denaturant induced a structural expansion as indicated by the significant increase of the radius of gyration of the ERα NTD monitored by SAXS.

Post-translational modifications such as phosphorylation elicit diverse effects on the biological functions of IDPs by altering the energetics of their conformational landscape and by modulating interactions with other cellular components[191,218] through stabilization and/or induction of secondary structural elements[219-221] (Fig. 6). Accordingly, most of the functionally known sites subjected to post-translational modifications in NRs are located in their NTDs, as for example in GR whose five phosphorylable serine residues are found in the NTD. In full agreement, the phosphorylation on S203, S211 and S226 of GR has been shown to induce conformational rearrangements leading to significantly enhanced AF-1-mediated transcriptional activity in a phosphorylation site-dependent manner.[209,222,223] Along this line, Peng and collaborators showed that mutation of S118, involved in a metastable contact with L33 in the NTD of ERα (see above), by the phosphomimetic aspartate residue (ERαS118D mutant) induced an expansion of the protein and increased the binding affinity to TBP, thus suggesting possible disruption of the long-

range contact and a possible link between phosphorylation, conformational change of the AF-1 and transcription.[217] Interestingly, phosphorylation of ERα at S118 was previously shown to induce chemical shifts that propagate throughout the NTD and allows recruitment of the peptidyl prolyl cis/trans isomerase Pin1. Subsequently, the *cis-trans* isomerization of P119 caused by Pin1 was found to create additional local structural changes, as revealed by NMR, leading to the enhancement of ERα transcriptional function.[180] Another recent study on the NTD of AR revealed that its interaction with the subunit 1 (RAP74) of the general transcription factor TFIIF is mediated by the sequence 423-446 within the AF-1.[224] Interestingly, further stabilization of the α-helical structure of the motif by using stapled peptides, together with phosphorylation of residue S424 dramatically increased the binding affinity to RAP74. Reminiscent of observations made on other IDP systems, phosphorylation may thus regulate NR functions by affecting the conformational dynamics of the NTD, leading to enhanced transcriptional activity.[225,226] In this respect, a mechanism for the interaction of disordered domains such as the NTD of NRs has been proposed according to which, in a first step, post-translational modifications such as phosphorylation would change their net charge thereby allowing weak polyelectrostatic interactions and, in a second step, hydrophobic interactions would direct the formation of a more stable complex where the activation domain would adopt a folded conformation for optimal interaction.[227,228]

3.3 Transcriptional coregulators and their interaction with the CTD/AF-2 of NRs

Like many proteins in signaling pathways, coregulators are large proteins belonging to the class of intrinsically disordered proteins that are functional despite a lack of stable three-dimensional structure. They act as platforms where multiple proteins attach to perform activities linked to gene transcription.[229,230] They interact with NRs through their Nuclear Receptor Interaction Domain (NRID) and recruit additional proteins to form either coactivator or corepressor complexes[231] (Fig. 7). Strikingly the intrinsic disorder of NRIDs is common to many other transcription factor interaction motifs located in these coregulator proteins.[232,233]

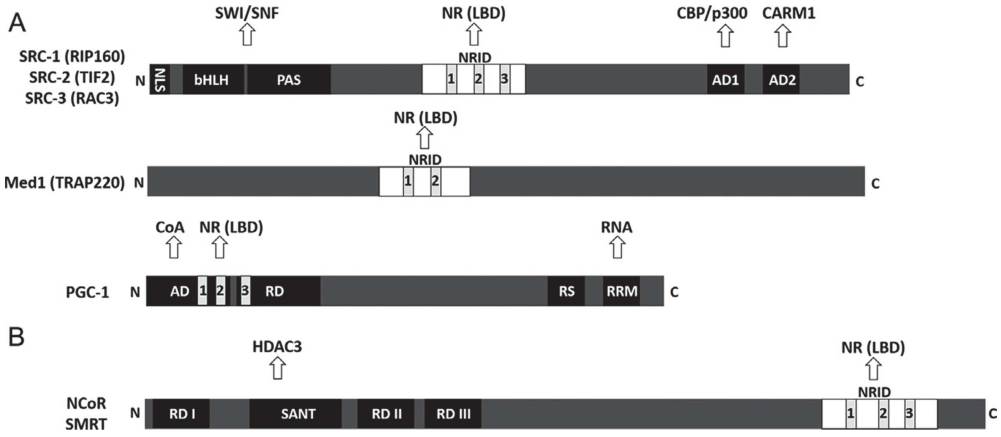


Fig. 7 Schematic representation of main transcriptional coregulators in interaction with nuclear receptors. (A**)** SRC-1, SRC-2/TIF-2 and SRC-3/RAC3 share a common organization. They have a basic helix-loop-helix/Per-ARNT-Sim homologous domain (bHLH/PAS), and a nuclear receptor (NR) interacting domain (NRID) with three LXXLL motifs critical for ligand-dependent interaction with NRs. The two activation domains AD1 and AD2 are located in the C-terminus where AD1 is responsible for interaction with the general transcriptional machinery through CBP and p300, and AD2 interacts with the histone methyltransferases coactivator-associated arginine methyltransferase 1 (CARM1). MED1/TRAP220 contains two NR boxes in its central NRID. PGC-1 contains three LXXLL motifs responsible for specific interaction with NRs. Other domains are an activation domain (AD) involved in binding to other coactivators, a repression domain (RD), an arginine-serine-rich domain (RS) and a RNA recognition motif (RRM). (B**)** The two corepressors NCoR and SMRT have common functional domains: C-terminal CoRNR boxes 1, 2 and 3 included in the NRID; the three repression domains (RDI, RDII and RDIII) that harbor intrinsic repression activity when tethered to the DNA-binding domain of GAL4 and two SANT domains.

Most of the coactivator complexes recruited by NRs do so through a conserved motif with the sequence LXXLL (the so-called NR box).[44] These LXXLL motifs form amphipathic α-helices that are responsible for direct association with NRs by recognizing a hydrophobic patch on the surface of their LBD, the structure of which is stabilized by the binding of agonist ligands. Shortly after the identification of the NR box, an analogous sequence motif (L/IXXI/VI or LXXXI/LXXXI/L) was identified in corepressor proteins.[234,235] This motif has been termed the CoRNR box and adopts a longer amphipathic helical conformation than the LXXLL motif. Crystal structures have demonstrated that peptides corresponding to NR or CoRNR boxes bind to the same hydrophobic groove on LBDs and therefore their recruitment is mutually exclusive.[236-239] The disordered NRID of coactivators and corepressors typically contains multiple NR and CoRNR boxes, respectively, connected by disordered linker regions (Fig. 7). These regions are thought to be important for NR interaction,

NR selectivity, binding, and a full transcriptional response.[240-242] Secondary structure and disorder prediction combined with computational data and in-depth biophysical and structural characterization by Size Exclusion Chromatography - Multi-Angle Light Scattering (SEC-MALS), CD, NMR and SAXS of various coactivator and corepressor NRIDs confirmed the intrinsic disorder of these regions[34,243,244] (our unpublished results). In addition, these studies showed the existence of pre-formed secondary structure elements localized in evolutionary conserved regions corresponding to NR and CoRNR boxes[34,243,244] (our unpublished results). These types of preformed secondary structures are usually important determinants for molecular recognition in IDPs.[245,246] Of note, these transient secondary structures in the unbound states resemble the secondary structures in the complexes with NRs. In addition, partial structures have been observed in flanking regions of these motifs that could play a role in NR specificity[237,247,248] and transient intramolecular contacts could exist within these NRIDs.[249]

The fact that coregulators contain multiple NR interaction motifs in a relatively close proximity, most of which appearing to be functional in terms of NR binding *in vitro*, brings up the still unsolved question of whether this multiplicity reflects redundancy or confers some specificity to the interface. Different structural studies showed that NR and CoRNR boxes are not equivalent in terms of NR affinity and specificity, which provide diversity and selectivity in the spectrum of NR/coregulator interactions for NR-specific responses upon specific ligand binding. For example, the crystal structure of PPARγ-LBD with a PGC-1 construct bearing both NR2 and NR3 boxes showed only one LXXLL motif (exclusively NR2) bound to PPARγ and the remaining portion of PGC-1 disordered.[213] This NR2 motif is also the major PGC-1 binding site to ER and RXR.[250,251] On the contrary, another study showed that the hepatocyte nuclear factor-4 alpha (HNF4α) might present a lack of strict selectivity in the interaction with the coactivator PGC-1 since the crystal structure of HNF4α-LBD in complex with the NRID of PGC-1 comprising all three LXXLL motifs revealed that only one LXXLL motif is bound to the canonical binding pocket and does not correspond to a single sequence but to an averaged structure of the different LXXLL motifs.[252] It was also discovered that two NRs, the retinoic acid receptor alpha (RARα) and Rev-erbα, that exhibit strong basal repressive activity, interact strongly with CoRNR1 but very weakly with CoRNR2, thanks to the formation of an

antiparallel β-sheet of N-terminal residues of CoRNR1 with an extended β-strand of RARα and Rev-erbα.[239,253] In addition, the differences in binding affinity of the different coregulator motifs to NRs induce directionality in the interaction of coregulators with NR heterodimers. Both receptors usually contribute to the interaction of the heterodimer with the coregulator through different thermodynamic mechanisms and to a different extent. Using a large set of biophysical and computational methods to study the interaction of RARα/RXRα heterodimer with the corepressor NCoR, we were able to show that although both NRs contribute to the interaction with the NRID in a cooperative manner, RARα plays a dominant role over RXRα.[249] The HDX-MS analysis on the complex between the heterodimer VDR/RXR and the coactivator SRC-1 supported a model of synergistic binding of one molecule of SRC-1 spanning both NRs, facilitated in the presence of agonist ligands of each subunit.[203] A combined NMR and HDX-MS analysis showed that all three LXXLL motifs within the SRC-2 NRID are involved or affected in the interaction with PPARγ/RXRα. One of the LXXLL motifs would bind to the RXRα AF-2 surface, most likely NR2 as it has the highest affinity for RXRα, while a second LXXLL motif would interact more weakly with the PPARγ AF-2 surface, and the third LXXLL motif could interact transiently with a region somewhere on the heterodimer LBD surface. Moreover, the conclusions drawn from this study suggested that ligand binding to RXRα would be the primary driver of SRC-2 NRID recruitment, even if both RXR and PPARγ agonist ligands contribute to enhance the SRC-2 NRID interaction through different thermodynamic mechanisms.[244] The observation that multiple NR interaction motifs within coregulators mediate cooperative binding to NR dimers was also made for homodimers, as exemplified by the interaction of SRC-1 with PPARγ homodimer where two motifs of SRC-1 make simultaneous contacts with both PPARγ LBDs.[236] Moreover, a SAXS analysis showed how the binding of a single PGC-1 molecule to the coactivator groove on both LBD units of estrogen-related receptor gamma (ERRγ) homodimer and the induced stabilization of the dimerization interface within the homodimer are facilitated by the intrinsic disorder in the coactivator.[243] In other cases, the binding mode observed for different coactivators to heterodimers is in favor of a fully asymmetric interaction where binding occurs to a single LBD. SAXS studies showed that RARα/RXRα and VDR/RXRα heterodimers recruit one coactivator molecule (MED1 or SRC-1 NRID) asymmetrically by a only binding the

partner of RXRα, the higher affinity of the coactivators for RARα or VDR rather than RXRα promoting this selective binding.[254] A low-resolution SAXS analysis suggested also the same asymmetric binding of the entire NRID of SRC-2 to PPARγ/RXRα where the coactivator flexibly interacts only with PPARγ using one LXXLL motif,[255] even if these data appear to be contradicted by de Vera and collaborators[244]. Several structural studies on the interaction of coactivators with homodimers revealed also an asymmetric binding, suggesting that negative cooperativity could be an essential mechanism controlling the binding of coregulators to NR homodimers. The coactivator SRC-1 seems to bind only to one monomer in RAR and ER homodimers,[255] similarly as the one-monomer interaction of PGC-1 in the complex with ERRα and ERRγ homodimers,[256] even if this last study appeared contradictory to [243]. A recent study[249] integrating information from different techniques into computational tool gave a new view on complexes between coregulator and NRs by showing that it exists an equilibrium between several conformational states of the complex formed by NCoR NRID and the heterodimer RARα/RXRα and that this equilibrium can be tuned by cognate ligands (Fig. 8). Such an approach could permit to reconcile the different views of complexes between coregulators and NR homo- and hetero-dimers and is the most promising strategy for the structural characterization of highly dynamic proteins and complexes in solution. Even if it is possible to define 3D molecular envelopes describing the low-resolution structures of flexible macromolecular assemblies involving IDPs,[243,254,255,257-259] such a structural description is probably not the most appropriate for a disordered protein or a dynamic complex with multiple conformations and orientations. Ensemble analysis of explicit or coarse-grained models provide a more accurate characterization of flexible macromolecular assemblies, particularly when combined with available structure coordinates of folded-domains determined by X-ray crystallography and data affordable by NMR or by biochemical assays.[34]

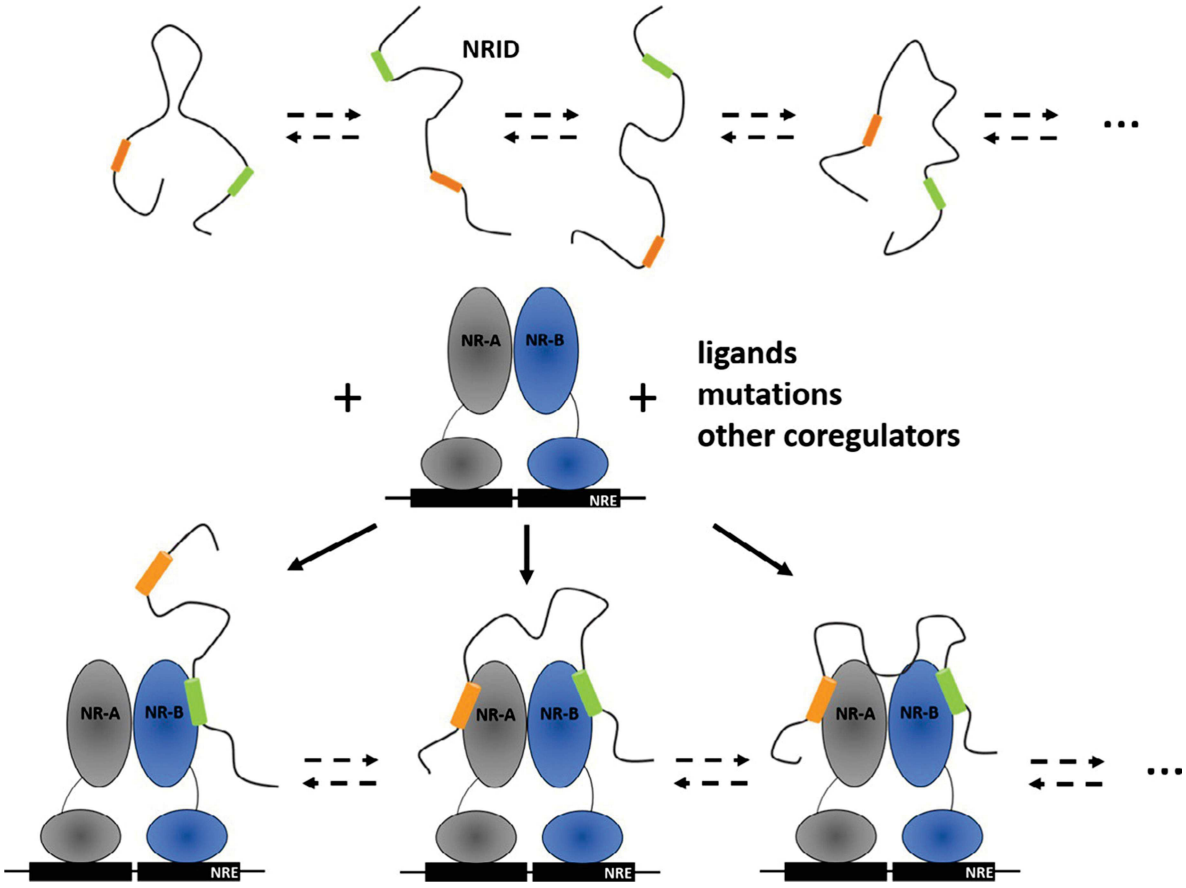


Fig. 8 Model of interaction of coregulator NRID with NR homo- and heterodimers bound to DNA. The NRID of coregulators exists as an ensemble of conformers in equilibrium with each other, which are mainly unstructured with transient secondary structures corresponding to NR interaction motifs. In interaction with NRs, the NRID undergoes disorder-to-order transition and a dynamic complex with multiple conformations and orientations is formed and equilibrium between the different states of the complex can be influenced by ligands, NR mutations and other coregulators.

Various HDX-MS analyses of complexes between coregulators and NRs showed significant changes in the conformational mobility of both partners during complexation. There is not only a stabilization of the conformational mobility of NR LBDs[243] but a disorder-to-order transition is also often observed for the coregulator.[243,249,255] Therefore, as for the AF-1, complex formation induces an increase in total secondary structure or a ternary folding of the coregulator often leading to its compaction[260] (Fig. 8). DNA may play a role in the allosteric control of coregulator binding and dimerization in NR heterodimers. Differences in NR DNA response element sequences coming from different target gene promoters are thought to propagate an allosteric conformational change, initiating in the NR DBD and functionally transmitted to other domains of full-length NRs, affecting

coregulator-NR interactions, in a ligand-specific manner.[261] For example, DNA binding to VDR/RXRα significantly impacts the conformation of the AF-2 surface in the LBDs of both components of the heterodimer,[203] modifying coregulator-NR interactions. DNA binding to PPARγ/RXRα stabilizes the conformation of the PPARγ LBD dimerization surface that interacts with the RXRα LBD, reduces the interaction conformational entropy of the complex to afford higher-affinity binding of SRC-2 NRID, and affects ligand potency for SRC-2 recruitment.[244] In addition, ligand binding to PPARγ/RXRα can allosterically affect the conformation of SRC-2 loop regions between interacting LXXLL motifs, as well as the C-terminal extended region that does not directly interact with PPARγ/RXRα. Quantitative studies on other NRs provide support for this, where the binding of the SRC-1 NRID to TR/RXR is influenced by the nature of the DNA response elements,[262-264] and where DNA and coregulator (SRC-2) cooperatively bind to the dimer of PR.[265] However, the structural mechanisms by which ligand and DNA binding to NRs may synergize to bind coregulator proteins require more studies.

3.4 Importance of NR-NTDs in diseases and as pharmaceutical targets

NRs are well-validated drug targets against cancers and metabolic diseases. The most widely used drugs target the structured pocket of the LBD to modulate coregulator interaction with the activation function AF-2. A number of studies has shown that the NTD of NRs is not only critical for several aspects of NR action but also can be exploited as drug targets, thereby opening new opportunities for endocrine-based therapies.[266-267] Lack of significant sequence homology between NRs within the NTD also suggests that targeting this domain could give selective modulators of NRs, yielding fewer side effects. Traditional views of structural drug discovery would tend to consider such intrinsically disordered regions as unattractive targets due to their lack of structure and the concomitant lack of binding pockets for small molecules to interact. However, an analysis based on the structures of intrinsically disordered proteins when folded and bound to their partners suggests they actually have a higher proportion of potential cavities where a small molecule could bind than their folded counterparts.[268] In addition, their inherent flexibility may allow them to conformationally adapt to complement a small molecule. Interestingly, the development of drugs that target disordered regions to block protein-protein

interactions is evolving and has provided potential for small molecules to have a sustained therapeutic effect.[269-270] Recent studies have shown that highly flexible and structurally dynamic NTD/AF-1 of steroid hormone receptors can be exploited as drug targets for endocrine-related cancers, thereby opening opportunities for development of novel small molecules that could block their interactions with coregulatory proteins outside the LBD/AF-2.[271]

Accordingly, few compounds that could find their application in the treatment of NTD-related diseases have been recently identified. The osmolyte trehalose has been shown to induce compaction or structure formation in the intrinsically disordered NTD of GR and to promote functionally ordered conformation of this domain, with potential clinical implications for human diseases involving disordered GR NTD.[223] A bisphenol A ester derivative (EPI-100) that specifically and covalently binds to the NTD of AR was shown to inhibit the constitutive activity of AR splice variants that lack the LBD and are suspected to be a cause of resistance in castration-recurrent prostate cancers (CRPC).[224,272,273] EPI-100 is the first small molecule inhibitor that directly binds to an intrinsically disordered region of a NR to enter clinical trials. Biophysical analysis revealed that EPI-001 alters folding of AR-NTD, thereby blocking the interaction of the AF-1 with specific coregulatory proteins. Other molecules, such as niphatenones and sintokamides, targeting AR NTD have also high potential as “selective androgen receptor modulators” to overcome drug resistance.[266] More generally, exploration of the druggability of disordered regions of NRs and coregulators, although challenging since it precluded high throughput screening of small molecules, is highly promising.[274]

The AR-NTD also plays a direct role in the pathogenesis of spinal and bulbar muscular atrophy (SBMA), a late-onset neuromuscular disorder which is caused by expansion of the polyglutamine repeat (polyQ) in this domain.[275] The mechanistic basis of this phenomenon is a matter of intense debate, but it is widely admitted that polyQ expansion cause neurotoxicity because it decreases protein solubility and leads to the formation of aggregates.[276] Later on, it has been shown that polyQ repeats undergo a length-dependent conformational change producing a highly insoluble structure when their lengths reach a threshold. In the AR, biophysical analyses have revealed that the helical propensity of polyQ tracts increases upon expansion, and above 37 residues, the AR variants form fibrillar cytotoxic

aggregates.[277,278] Using an approach combining MD simulations, NMR, CD and quantum mechanics, it has been recently shown that the helicity of the repeat correlates with its length as a result of the accumulation of unconventional interactions where glutamine side chains are involved in a hydrogen bond with the main chain carbonyl of the residue at position i-4. In addition, the study revealed that the strength of these interactions depends on the nature the acceptor residue and on its propensity to form a hydrogen bond that is strongly influenced by its side chain. Thus, the secondary structure of polyQ tracts appears to depend on their local environment.[279] These data provide a mechanistic basis of SBMA and a rationale for the development of therapeutic strategies to cure this disease.

3.5 Liquid-Liquid Phase Separation (LLPS) in NRs

Equivalently to other transcription factors, a recent report linked the ER NTD to the liquid-liquid phase separation (LLPS) phenomenon observed for this NR and provides a potential functional role for this domain.[280] LLPS has emerged in the recent years as a key cellular mechanism to regulate multiple processes, including gene transcription.[281] The formation of liquid droplets brings together many constituents through cooperative interactions of multivalent biomolecules, such as disordered proteins and nucleic acids, constituting membrane less compartments.[282] Examples of such organelles in the nucleus include nucleoli, Cajal bodies and nuclear speckles.[283] These organelles display properties of liquid droplets, such as fusion and fission, and their formation can be formally described as a LLPS normally used soft-matter studies. To our knowledge, only the NTD of the human RXRγ has been described to form liquid droplets *in vitro* in the absence of other proteins.[284] For the previously mentioned ER case, liquid droplets were observed in the presence of the coactivator MED1 NRID, containing the LXXLL motifs. Interestingly, in the presence of the agonist (estrogen) the incorporation of the ER into the MED1 droplets is enhanced. These results suggest that transcription and incorporation into liquid droplets can be two facets of the same phenomenon.[280] Taking into account that NTDs and coregulators are low-complexity disordered proteins, which can form multivalent interactions, it is tempting to hypothesize that NRs can be part of liquid droplets that compartmentalize the transcription machinery and enhance its efficiency. Notice that the RNA polymerase II can also phase separate *in vivo* through its C-terminal domain.[285] In fact, these multiprotein condensates could

explain the co-localization of multiple enhancers, enabling the simultaneous activation of multiple genes and their concerted regulation.[286]

4. Conclusions & Perspectives

The first 3D structure determination of myoglobin in 1958 by Kendrew and collaborators[287] in 1958 initiated the era of structural biology. A established paradigm defining that only well-structured proteins were functional had to be redefined when it was predicted that many proteins in eukaryotes were devoid of a globular form.[31,288] The characterization of the first partly unfolded protein under normal condition in the cell by Kriwacki and collaborators,[289] and the transcriptional coactivator, CPB and its paralog p300[290,291] were the demonstration of the relevance of IDPs.

In this chapter, the relevance of disorder is exemplified with two prototypical biological systems where the coordinated action of several players with different degrees of structuration is crucial for signaling and regulation. In these two biological machineries, GPCRs and NRs, we see the importance of disorder in orchestrating and modulating in a highly precise manner very complex biological processes. Indeed, the diversity of actions and outputs enabled by disorder is further increased thanks to PTMs and alternative splicing that widely occurs in these regions. These modifications modulate molecular recognition and specificity as well as define the cell type where their action should take place.

GPCRs and NRs are major players in cellular communication activities. As such, they represent targets of paramount interest for drug candidates, with applications in many different physiopathological processes. Mainly organized as a globular core with disordered appendices, most pharmacological and structural analyses have focused so far on the globular domains, while disordered regions are far less well characterized despite the fact that they are pivotal in coupling the signaling partners and regulating transcription. Further characterization of these domains is thus required to understand signaling and fully exploit their therapeutic potential. However, the structural characterization is still obscured by intrinsic properties of these disordered domains, including the absence of stable secondary structure elements, high conformational dynamics, the occurrence of tightly regulated PTMs and the interaction with many different proteins. In this chapter we have highlighted the importance of a better description of the disordered parts of these

complex machineries to design better modulators/inhibitors of ordered/disordered interacting surfaces to target specific pathway and decrease deleterious side effects of pharmacologic drugs.

Acknowledgments

We thank current and former members of our research groups for invaluable contributions at many stages of this work. N.S. & P.B. acknowledge support of the Labex EpiGenMed (ANR-10-LABX-12-01), N.S. & J.L.B. support of the ANR GPCteR (ANR-17-CE11-0022-01). The CBS is a member of the French Infrastructure for Integrated Structural Biology (FRISBI) supported by the French National Research Agency (ANR-10-INSB-05).

References

1. Potenza E, Di Domenico T, Walsh I, Tosatto SCE. MobiDB 2.0: An improved database of intrinsically disordered and mobile proteins. Nucleic Acids Res. 2015;43(D1):D315-D320. doi:10.1093/nar/gku982
2. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradović Z. Intrinsic disorder and protein function. Biochemistry. 2002;41(21):6573-6582. doi:10.1021/bi012159+
3. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. J Mol Biol. 2004;337(3):635-645. doi:10.1016/j.jmb.2004.02.002
4. Pejaver V, Hsu WL, Xin F, Dunker AK, Uversky VN, Radivojac P. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. Protein Sci. 2014;23(8):1077-1093. doi:10.1002/pro.2494
5. Iakoucheva LM, Radivojac P, Brown CJ, et al. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 2004;32(3):1037-1049. doi:10.1093/nar/gkh253
6. Cox CJ, Dutta K, Petri ET, et al. The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. FEBS Lett. 2002;527(1-3):303-308. doi:10.1016/S0014-5793(02)03246-5
7. Girdwood D, Bumpass D, Vaughan OA, et al. p300 transcriptional repression is mediated by SUMO modification. Mol Cell. 2003;11(4):1043-1054. doi:10.1016/S1097-2765(03)00141-2
8. Witze ES, Old WM, Resing KA, Ahn NG. Mapping protein post-translational modifications with mass spectrometry. Nat Methods. 2007;4(10). doi:10.1038/1100
9. Uversky VN. Dancing protein clouds: The strange biology and chaotic physics of intrinsically disordered proteins. J Biol Chem. 2016;291(13):6681-6688. doi:10.1074/jbc.R115.685859

10. Uversky VN, Gillespie JR, Fink AL. Why are “natively unfolded” proteins unstructured under physiologic conditions? Proteins Struct Funct Genet. 2000;41(3):415-427. doi:10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7
11. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. Proteins Struct Funct Genet. 2001;42(1):38-48. doi:10.1002/1097-0134(20010101)42:1<38::AID-PROT50>3.0.CO;2-3
12. Davey NE, Van Roey K, Weatheritt RJ, et al. Attributes of short linear motifs. Mol Biosyst. 2012;8(1):268-281. doi:10.1039/c1mb05231d
13. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015;16(1):18-29. doi:10.1038/nrm3920
14. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005;6(3):197-208. doi:10.1038/nrm1589
15. Xie H, Vucetic S, Iakoucheva LM, et al. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J Proteome Res. 2007;6(5):1882-1898. doi:10.1021/pr060392u
16. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit. 2005;18(5):343-384. doi:10.1002/jmr.747
17. Uversky VN, Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins in Human Diseases: Introducing the D 2 Concept . Annu Rev Biophys. 2008;37(1):215-246. doi:10.1146/annurev.biophys.37.032807.125924
18. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol. 2002;323(3):573-584. doi:10.1016/S0022-2836(02)00969-5
19. Landau K, Na I, Schenck R, Uversky V. Unfoldomics of prostate cancer: On the abundance and roles of intrinsically disordered proteins in prostate cancer. Asian J Androl. 2016;18(5):662-672. doi:10.4103/1008-682X.184999
20. Uversky VN. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: Another illustration of the D 2 concept. Expert Rev Proteomics. 2010;7(4):543-564. doi:10.1586/epr.10.36
21. Santamaria N, Alhothali M, Alfonso MH, Breydo L, Uversky VN. Intrinsic disorder in proteins involved in amyotrophic lateral sclerosis. Cell Mol Life Sci. 2017;74(7):1297-1318. doi:10.1007/s00018-016-2416-6
22. Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN. Abundance of intrinsic disorder in protein associated with cardiovascular disease. Biochemistry. 2006;45(35):10448-10460. doi:10.1021/bi060981d
23. Du Z, Uversky VN. A comprehensive survey of the roles of highly disordered proteins in type 2 diabetes. Int J Mol Sci. 2017;18(10). doi:10.3390/ijms18102010
24. Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. Chem Rev. 2004;104(8):3607-3622. doi:10.1021/cr030403s
25. Sibille N, Bernadó P. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. Biochem Soc Trans. 2012;40(5):956-962. doi:10.1042/BST20120149
26. Delaforge E, Cordeiro TN, Bernadó P, Sibille N. Conformational characterization of intrinsically disordered proteins and its biological

significance. Mod Magn Reson. 2018:381-399. doi:10.1007/978-3-319-28388-3_52
27. Milles S, Salvi N, Blackledge M, Jensen MR. Characterization of intrinsically disordered proteins and their dynamic complexes: From in vitro to cell-like environments. Prog Nucl Magn Reson Spectrosc. 2018;109:79-100. doi:10.1016/j.pnmrs.2018.07.001
28. Dinkel H, Van Roey K, Michael S, et al. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. 2014;42(D1). doi:10.1093/nar/gkt1047
29. Mi T, Merlin JC, Deverasetty S, et al. Minimotif Miner 3.0: Database expansion and significantly improved reduction of false-positive predictions from consensus sequences. Nucleic Acids Res. 2012;40(D1):252-260. doi:10.1093/nar/gkr1189
30. Van Der Lee R, Buljan M, Lang B, et al. Classification of intrinsically disordered regions and proteins. Chem Rev. 2014;114(13):6589-6631. doi:10.1021/cr400525m
31. Wright PE, Dyson HJ. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J Mol Biol. 1999;293(2):321-331. doi:10.1006/jmbi.1999.3110
32. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol. 2002;12(1):54-60. doi:10.1016/S0959-440X(02)00289-0
33. Spolar RS, Record MT. Coupling of local folding to site-specific binding of proteins to DNA. Science (80-). 1994;263(5148):777-784. doi:10.1126/science.8303294
34. Cordeiro TN, Herranz-Trillo F, Urbanek A, et al. Small-angle scattering studies of intrinsically disordered proteins and their complexes. Curr Opin Struct Biol. 2017;42:15-23. doi:10.1016/j.sbi.2016.10.011
35. Bernadó P, Svergun DI. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. Mol BioSyst. 2012;8(1):151-167. doi:10.1039/C1MB05275F
36. Marsh JA, Forman-Kay JD. Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. Proteins Struct Funct Bioinforma. 2012;80(2):556-572. doi:10.1002/prot.23220
37. Salmon L, Nodet G, Ozenne V, et al. NMR characterization of long-range order in intrinsically disordered proteins. J Am Chem Soc. 2010;132(24):8407-8418. doi:10.1021/ja101645g
38. Dunker AK, Brown CJ, Obradovic Z. Identification and functions of usefully disordered proteins. Adv Protein Chem. 2002;62:25-49. doi:10.1016/S0065-3233(02)62004-2
39. Robinson C V, Sali A, Baumeister W. The molecular sociology of the cell. Nature. 2007;450(7172):973-982. doi:10.1038/nature06523
40. Lagerström MC, Schiöth HB. Structural diversity of G protein-coupled receptors and significance for drug discovery. Nat Rev Drug Discov. 2008;7(4):339-357. doi:10.1038/nrd2518
41. Hauser AS, Attwood MM, Rask-Andersen M, Schioth HB, Gloriam DE. Trends in GPCR drug discovery: new agents, targets and indications. Nat Rev Drug Discov. 2017;16(12):829-842. doi:10.1038/nrd.2017.178

42. Azam S, Haque ME, Jakaria M, Jo S-H, Kim I-S, Choi D-K. G-Protein-Coupled Receptors in CNS: A Potential Therapeutic Target for Intervention in Neurodegenerative Disorders and Associated Cognitive Deficits. Cells. 2020;9(506).
43. Ottow E, Weinmann H. Nuclear Receptors as Drug Targets: A Historical Perspective of Modern Drug Discovery. In: Nuclear Receptors as Drug Targets. John Wiley & Sons, Ltd; 2008:1-23. doi:10.1002/9783527623297.ch1
44. Heery DM, Kalkhoven E, Hoare S, Parker MG. A signature motif in transcriptional co-activators mediates binding to nuclear receptors. Nature. 1997;387(6634):733-736. doi:10.1038/42750
45. Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, Asturias FJ. Malleable machines take shape in eukaryotic transcriptional regulation. Nat Chem Biol. 2008;4(12):728-737. doi:10.1038/nchembio.127
46. Rasmussen SG, Choi HJ, Rosenbaum DM, et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. Nature. 2007;450(7168):383-387. doi:10.1038/nature06325
47. Palczewski K, Kumasaka T, Hori T, et al. Crystal structure of rhodopsin: A G protein-coupled receptor. Science (80-). 2000;289(5480):739-745. doi:10.1126/science.289.5480.739
48. Granier S, Kim S, Shafer AM, et al. Structure and conformational changes in the C-terminal domain of the beta2-adrenoceptor: insights from fluorescence resonance energy transfer studies. J Biol Chem. 2007;282(18):13895-13905. doi:10.1074/jbc.M611904200
49. Fonin A V., Darling AL, Kuznetsova IM, Turoverov KK, Uversky VN. Multi-Functionality of Proteins Involved in GPCR and G Protein Signaling: Making Sense of Structure–Function Continuum with Intrinsic Disorder-Based Proteoforms. Vol 76. Springer International Publishing; 2019. doi:10.1007/s00018-019-03276-1
50. Hilser VJ, Thompson EB. Structural dynamics, intrinsic disorder, and allostery in nuclear receptors as transcription factors. J Biol Chem. 2011. doi:10.1074/jbc.R111.278929
51. Uversky VN. Functional roles of transiently and intrinsically disordered regions within proteins. FEBS J. 2015;282(7):1182-1189. doi:10.1111/febs.13202
52. Pierce KL, Premont RT, Lefkowitz RJ. Seven-transmembrane receptors. Nat Rev Mol Cell Biol. 2002;3(9):639-650. doi:10.1038/nrm908
53. Lagerstrom MC, Schioth HB. Structural diversity of G protein-coupled receptors and significance for drug discovery. Nat Rev Drug Discov. 2008;7(4):339-357. doi:10.1038/nrd2518
54. Oldham WM, Hamm HE. Heterotrimeric G protein activation by G-protein-coupled receptors. Nat Rev Mol Cell Biol. 2008;9(1):60-71. doi:10.1038/nrm2299
55. Gurevich V V, Gurevich E V. GPCR Signaling Regulation: The Role of GRKs and Arrestins. Front Pharmacol. 2019;10:125. doi:10.3389/fphar.2019.00125
56. Komolov KE, Benovic JL. G protein-coupled receptor kinases: Past, present and future. Cell Signal. 2018;41:17-24. doi:10.1016/j.cellsig.2017.07.004
57. Vilardaga JP, Jean-Alphonse FG, Gardella TJ. Endosomal generation of cAMP in GPCR signaling. Nat Chem Biol. 2014;10(9):700-706. doi:10.1038/nchembio.1611

58. Gurevich V V., Gurevich E V. Plethora of functions packed into 45 kDa arrestins: biological implications and possible therapeutic strategies. Cell Mol Life Sci. 2019;76(22):4413-4421. doi:10.1007/s00018-019-03272-5
59. Grundmann M, Merten N, Malfacini D, et al. Lack of beta-arrestin signaling in the absence of active G proteins. Nat Commun. 2018;9(1):341. doi:10.1038/s41467-017-02661-3
60. Chaumont-Dubel S, Dupuy V, Bockaert J, Becamel C, Marin P. The 5-HT6 receptor interactome: New insight in receptor signaling and its impact on brain physiology and pathologies. Neuropharmacology. 2019:107839. doi:10.1016/j.neuropharm.2019.107839
61. Schioth HB, Fredriksson R. The GRAFS classification system of G-protein coupled receptors in comparative perspective. Gen Comp Endocrinol. 2005;142(1-2):94-101. doi:10.1016/j.ygcen.2004.12.018
62. Jaakola VP, Prilusky J, Sussman JL, Goldman A. G protein-coupled receptors show unusual patterns of intrinsic unfolding. Protein Eng Des Sel. 2005;18(2):103-110. doi:10.1093/protein/gzi004
63. Cherezov V, Rosenbaum DM, Hanson MA, et al. High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science (80-). 2007;318(5854):1258-1265. doi:10.1126/science.1150577
64. Shimada I, Ueda T, Kofuku Y, Eddy MT, Wuthrich K. GPCR drug discovery: integrating solution NMR data with crystal and cryo-EM structures. Nat Rev Drug Discov. 2019;18(1):59-82. doi:10.1038/nrd.2018.180
65. Rasmussen SG, DeVree BT, Zou Y, et al. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. Nature. 2011;477(7366):549-555. doi:10.1038/nature10361
66. Garcia-Nafria J, Tate CG. Cryo-Electron Microscopy: Moving Beyond X-Ray Crystal Structures for Drug Receptors and Drug Development. Annu Rev Pharmacol Toxicol. 2020;60:51-71. doi:10.1146/annurev-pharmtox-010919-023545
67. Liang YL, Belousoff MJ, Zhao P, et al. Toward a Structural Understanding of Class B GPCR Peptide Binding and Activation. Mol Cell. 2020;77(3):656-668 e5. doi:10.1016/j.molcel.2020.01.012
68. Koehl A, Hu H, Feng D, et al. Structural insights into the activation of metabotropic glutamate receptors. Nature. 2019;566(7742):79-84. doi:10.1038/s41586-019-0881-4
69. Qi X, Liu H, Thompson B, McDonald J, Zhang C, Li X. Cryo-EM structure of oxysterol-bound human Smoothened coupled to a heterotrimeric Gi. Nature. 2019;571(7764):279-283. doi:10.1038/s41586-019-1286-0
70. Yin W, Li Z, Jin M, et al. A complex structure of arrestin-2 bound to a G protein-coupled receptor. Cell Res. 2019;29(12):971-983. doi:10.1038/s41422-019-0256-2
71. Huang W, Masureel M, Qu Q, et al. Structure of the neurotensin receptor 1 in complex with beta-arrestin 1. Nature. 2020;579(7798):303-308. doi:10.1038/s41586-020-1953-1
72. Staus DP, Hu H, Robertson MJ, et al. Structure of the M2 muscarinic receptor-beta-arrestin complex in a lipid nanodisc. Nature. 2020;579(7798):297-302. doi:10.1038/s41586-020-1954-0

73. Nguyen AH, Thomsen ARB, Cahill 3rd TJ, et al. Structure of an endosomal signaling GPCR-G protein-beta-arrestin megacomplex. Nat Struct Mol Biol. 2019;26(12):1123-1131. doi:10.1038/s41594-019-0330-y
74. Venkatakrishnan AJ, Flock T, Prado DE, Oates ME, Gough J, Madan Babu M. Structured and disordered facets of the GPCR fold. Curr Opin Struct Biol. 2014;27:129-137. doi:10.1016/j.sbi.2014.08.002
75. Hilger D, Masureel M, Kobilka BK. Structure and dynamics of GPCR signaling complexes. Nat Struct Mol Biol. 2018;25(1):4-12. doi:10.1038/s41594-017-0011-7
76. Casiraghi M, Point E, Pozza A, Moncoq K, Baneres JL, Catoire LJ. NMR analysis of GPCR conformational landscapes and dynamics. Mol Cell Endocrinol. 2019;484:69-77. doi:10.1016/j.mce.2018.12.019
77. Tastan O, Klein-Seetharaman J, Meirovitch H. The effect of loops on the structural organization of alpha-helical membrane proteins. Biophys J. 2009;96(6):2299-2312. doi:10.1016/j.bpj.2008.12.3894
78. Woods AS. The dopamine D(4) receptor, the ultimate disordered protein. J Recept Signal Transduct Res. 2010;30(5):331-336. doi:10.3109/10799893.2010.513842
79. Rahmeh R, Damian M, Cottet M, et al. Structural insights into biased G protein-coupled receptor signaling revealed by fluorescence spectroscopy. Proc Natl Acad Sci U S A. 2012;109(17):6733-6738. doi:10.1073/pnas.1201093109
80. Bichet DG. [Nephrogenic diabetes insipidus]. Nephrol Ther. 2006;2(6):387-404. doi:10.1016/j.nephro.2006.07.010
81. Rene P, Le Gouill C, Pogozheva ID, et al. Pharmacological chaperones restore function to MC4R mutants responsible for severe early-onset obesity. J Pharmacol Exp Ther. 2010;335(3):520-532. doi:10.1124/jpet.110.172098
82. Goth CK, Petäjä-Repo U, Rosenkilde MM. G Protein-Coupled Receptors in the sweet spot: glycosylation and other post-translational modifications. ACS Pharmacol Transl Sci. 2020;in press.
83. Lanctot PM, Leclerc PC, Escher E, Leduc R, Guillemette G. Role of N-glycosylation in the expression and functional properties of human AT1 receptor. Biochemistry. 1999;38(27):8621-8627. doi:10.1021/bi9830516
84. Hawtin SR, Davies AR, Matthews G, Wheatley M. Identification of the glycosylation sites utilized on the V1a vasopressin receptor and assessment of their role in receptor signalling and expression. Biochem J. 2001;357(Pt 1):73-81. doi:10.1042/0264-6021:3570073
85. Compton SJ, Sandhu S, Wijesuriya SJ, Hollenberg MD. Glycosylation of human proteinase-activated receptor-2 (hPAR2): role in cell surface expression and signalling. Biochem J. 2002;368(Pt 2):495-505. doi:10.1042/BJ20020706
86. He J, Xu J, Castleberry AM, Lau AG, Hall RA. Glycosylation of beta(1)-adrenergic receptors regulates receptor surface expression and dimerization. Biochem Biophys Res Commun. 2002;297(3):565-572. doi:10.1016/s0006-291x(02)02259-3
87. Markkanen PM, Petaja-Repo UE. N-glycan-mediated quality control in the endoplasmic reticulum is required for the expression of correctly folded delta-opioid receptors at the cell surface. J Biol Chem. 2008;283(43):29086-29098. doi:10.1074/jbc.M801880200

88. Virion Z, Doly S, Saha K, et al. Sialic acid mediated mechanical activation of beta2 adrenergic receptors by bacterial pili. Nat Commun. 2019;10(1):4752. doi:10.1038/s41467-019-12685-6
89. Mukai A, Yamamoto-Hino M, Komada M, Okano H, Goto S. Balanced ubiquitination determines cellular responsiveness to extracellular stimuli. Cell Mol Life Sci. 2012;69(23):4007-4016. doi:10.1007/s00018-012-1084-4
90. Malik R, Soh UJ, Trejo J, Marchese A. Novel roles for the E3 ubiquitin ligase atrophin-interacting protein 4 and signal transduction adaptor molecule 1 in G protein-coupled receptor signaling. J Biol Chem. 2012;287(12):9013-9027. doi:10.1074/jbc.M111.336792
91. Chini B, Parenti M. G-protein-coupled receptors, cholesterol and palmitoylation: facts about fats. J Mol Endocrinol. 2009;42(5):371-379. doi:10.1677/JME-08-0114
92. Prihandoko R, Bradley SJ, Tobin AB, Butcher AJ. Determination of GPCR Phosphorylation Status: Establishing a Phosphorylation Barcode. Curr Protoc Pharmacol. 2015;69:2 13 1-2 13 26. doi:10.1002/0471141755.ph0213s69
93. Premont RT, Macrae AD, Aparicio SAJR, Kendall HE, Welch JE, Lefkowitz RJ. The GRK4 subfamily of G protein-coupled receptor kinases. Alternative splicing, gene organization, and sequence conservation. J Biol Chem. 1999;274(41):29381-29389. doi:10.1074/jbc.274.41.29381
94. Taylor SS, Yang J, Wu J, Haste NM, Radzio-Andzelm E, Anand G. PKA: A portrait of protein kinase dynamics. Biochim Biophys Acta - Proteins Proteomics. 2004;1697(1-2):259-269. doi:10.1016/j.bbapap.2003.11.029
95. Huse M, Kuriyan J. The conformational plasticity of protein kinases. Cell. 2002;109(3):275-282. doi:10.1016/S0092-8674(02)00741-9
96. Komolov KE, Bhardwaj A, Benovic JL. Atomic structure of GRK5 reveals distinct structural features novel for G protein-coupled receptor kinases. J Biol Chem. 2015;290(34):20629-20647. doi:10.1074/jbc.M115.647297
97. Komolov KE, Du Y, Duc NM, et al. Structural and Functional Analysis of a β2-Adrenergic Receptor Complex with GRK5. Cell. 2017;169(3):407-421.e16. doi:10.1016/j.cell.2017.03.047
98. Palczewski K, Buczylko J, Kaplan MW, Polans AS, Crabb JW. Mechanism of rhodopsin kinase activation. J Biol Chem. 1991;266(20):12949-12955.
99. Chen CY, Dion SB, Kim CM, Benovic JL. β-Adrenergic receptor kinase. Agonist-dependent receptor binding promotes kinase activation. J Biol Chem. 1993;268(11):7825-7831.
100. Kannan N, Haste N, Taylor SS, Neuwald AF. The hallmark of AGC kinase functional divergence is its C-terminal tail, a cis-acting regulatory module (Proceedings of the National Academy of Sciences of the United States of America (2007) 104, (1272-1277) DOI:10.1073/pnas. 0610251104). Proc Natl Acad Sci U S A. 2008;105(26):9130. doi:10.1073/pnas.0804708105
101. Singh P, Wang B, Maeda T, Palczewski K, Tesmer JJG. Structures of rhodopsin kinase in different ligand states reveal key elements involved in G protein-coupled receptor kinase activation. J Biol Chem. 2008;283(20):14053-14062. doi:10.1074/jbc.M708974200
102. Boguth CA, Singh P, Huang CC, Tesmer JJG. Molecular basis for activation of G protein-coupled receptor kinases. EMBO J. 2010;29(19):3249-3259. doi:10.1038/emboj.2010.206

103. Yu QM, Cheng ZJ, Gan XQ, Bao G Bin, Li L, Pei G. The amino terminus with a conserved glutamic acid of G protein-coupled receptor kinases is indispensable for their ability to phosphorylate photoactivated rhodopsin. J Neurochem. 1999;73(3):1222-1227. doi:10.1046/j.1471-4159.1999.0731222.x
104. Noble B, Kallal LA, Pausch MH, Benovic JL. Development of a yeast bioassay to characterize G protein-coupled receptor kinases: Identification of an NH2-terminal region essential for receptor phosphorylation. J Biol Chem. 2003;278(48):47466-47476. doi:10.1074/jbc.M308257200
105. Pao CS, Barker BL, Benovic JL. Role of the amino terminus of G protein-coupled receptor kinase 2 in receptor phosphorylation. Biochemistry. 2009;48(30):7325-7333. doi:10.1021/bi900408g
106. Huang CC, Yoshino-Koh K, Tesmer JJG. A surface of the kinase domain critical for the allosteric activation of G protein-coupled receptor kinases. J Biol Chem. 2009;284(25):17206-17215. doi:10.1074/jbc.M809544200
107. Sterne-Marr R, Leahey PA, Bresee JE, et al. GRK2 activation by receptors: Role of the kinase large lobe and carboxyl-terminal tail. Biochemistry. 2009;48(20):4285-4293. doi:10.1021/bi900151g
108. Lodowski DT, Tesmer VM, Benovic JL, Tesmer JJG. The structure of G protein-coupled receptor kinase (GRK)-6 defines a second lineage of GRKs. J Biol Chem. 2006;281(24):16785-16793. doi:10.1074/jbc.M601327200
109. Shiraishi Y, Natsume M, Kofuku Y, et al. Phosphorylation-induced conformation of beta2-adrenoceptor related to arrestin recruitment revealed by NMR. Nat Commun. 2018;9(1):194. doi:10.1038/s41467-017-02632-8
110. Vishnivetskiy SA, Raman D, Wei J, Kennedy MJ, Hurley JB, Gurevich V V. Regulation of arrestin binding by rhodopsin phosphorylation level. J Biol Chem. 2007;282(44):32075-32083. doi:10.1074/jbc.M706057200
111. Nobles KN, Xiao K, Ahn S, et al. Distinct phosphorylation sites on the beta(2)-adrenergic receptor establish a barcode that encodes differential functions of beta-arrestin. Sci Signal. 2011;4(185):ra51. doi:10.1126/scisignal.2001707
112. Pack TF, Orlen MI, Ray C, Peterson SM, Caron MG. The dopamine D2 receptor can directly recruit and activate GRK2 without G protein activation. J Biol Chem. 2018;293(16):6161-6171. doi:10.1074/jbc.RA117.001300
113. Prihandoko R, Alvarez-Curto E, Hudson BD, et al. Distinct Phosphorylation Clusters Determine the Signaling Outcome of Free Fatty Acid Receptor 4/G Protein-Coupled Receptor 120. Mol Pharmacol. 2016;89(5):505-520. doi:10.1124/mol.115.101949
114. Bouzo-Lorenzo M, Santo-Zas I, Lodeiro M, et al. Distinct phosphorylation sites on the ghrelin receptor, GHSR1a, establish a code that determines the functions of ss-arrestins. Sci Rep. 2016;6:22495. doi:10.1038/srep22495
115. Mokrosinski J, Frimurer TM, Sivertsen B, Schwartz TW, Holst B. Modulation of constitutive activity and signaling bias of the ghrelin receptor by conformational constraint in the second extracellular loop. J Biol Chem. 2012;287(40):33488-33502. doi:10.1074/jbc.M112.383240
116. Lin X, Li M, Wang N, et al. Structural basis of ligand recognition and self-activation of orphan GPR52. Nature. 2020;579(7797):152-157. doi:10.1038/s41586-020-2019-0
117. Rajagopalan L, Rajarathnam K. Ligand selectivity and affinity of chemokine receptor CXCR1. Role of N-terminal domain. J Biol Chem. 2004;279(29):30000-30008. doi:10.1074/jbc.M313883200

118. Rajagopalan L, Rajarathnam K. Structural basis of chemokine receptor function--a model for binding affinity and ligand selectivity. Biosci Rep. 2006;26(5):325-339. doi:10.1007/s10540-006-9025-9
119. Prado GN, Suetomi K, Shumate D, et al. Chemokine signaling specificity: essential role for the N-terminal domain of chemokine receptors. Biochemistry. 2007;46(31):8961-8968. doi:10.1021/bi7004043
120. De Filippo E, Hinz S, Pellizzari V, et al. A2A and A2B adenosine receptors: The extracellular loop 2 determines high (A2A) or low affinity (A2B) for adenosine. Biochem Pharmacol. 2020;172:113718. doi:10.1016/j.bcp.2019.113718
121. Catoire LJ, Damian M, Baaden M, Guittet E, Baneres JL. Electrostatically-driven fast association and perdeuteration allow detection of transferred cross-relaxation for G protein-coupled receptor ligands with equilibrium dissociation constants in the high-to-low nanomolar range. J Biomol NMR. 2011;50(3):191-195. doi:10.1007/s10858-011-9523-3
122. Ferre G, Louet M, Saurel O, et al. Structure and dynamics of G protein-coupled receptor-bound ghrelin reveal the critical role of the octanoyl chain. Proc Natl Acad Sci U S A. 2019;116(35):17525-17530. doi:10.1073/pnas.1905105116
123. Wheatley M, Wootten D, Conner MT, et al. Lifting the lid on GPCRs: the role of extracellular loops. Br J Pharmacol. 2012;165(6):1688-1703. doi:10.1111/j.1476-5381.2011.01629.x
124. Bockaert J, Perroy J, Becamel C, Marin P, Fagni L. GPCR interacting proteins (GIPs) in the nervous system: Roles in physiology and pathologies. Annu Rev Pharmacol Toxicol. 2010;50:89-109. doi:10.1146/annurev.pharmtox.010909.105705
125. Kim JM, Hwa J, Garriga P, Reeves PJ, RajBhandary UL, Khorana HG. Light-driven activation of beta 2-adrenergic receptor signaling by a chimeric rhodopsin containing the beta 2-adrenergic receptor cytoplasmic loops. Biochemistry. 2005;44(7):2284-2292. doi:10.1021/bi048328i
126. Garcia-Nafria J, Tate CG. Cryo-EM structures of GPCRs coupled to Gs, Gi and Go. Mol Cell Endocrinol. 2019;488:1-13. doi:10.1016/j.mce.2019.02.006
127. Du Y, Duc NM, Rasmussen SGF, et al. Assembly of a GPCR-G Protein Complex. Cell. 2019;177(5):1232-1242 e11. doi:10.1016/j.cell.2019.04.022
128. Granzin J, Wilden U, Choe HW, Labahn J, Krafft B, Buldt G. X-ray crystal structure of arrestin from bovine rod outer segments. Nature. 1998;391(6670):918-921. doi:10.1038/36147
129. Hirsch JA, Schubert C, Gurevich V V, Sigler PB, Hughes H, Haven N. The 2.8˚ Crystal Structure of Visual Arrestin : The 2 . 8 A A Model for Arrestin ' s Regulation. Cell. 1999;97(2):257-269.
130. Han M, Gurevich V V., Vishnivetskiy SA, Sigler PB, Schubert C. Crystal structure of β-arrestin at 1.9 Å: Possible mechanism of receptor binding and membrane translocation. Structure. 2001;9(9):869-880. doi:10.1016/S0969-2126(01)00644-X
131. Zhan X, Gimenez LE, Gurevich V V., Spiller BW. Crystal structure of arrestin-3 reveals the basis of the difference in receptor binding between two non-visual subtypes. J Mol Biol. 2011;406(3):467-478. doi:10.1016/j.jmb.2010.12.034

132. Sutton RB, Vishnivetskiy SA, Robert J, et al. Crystal structure of cone arrestin at 2.3 Å: Evolution of receptor specificity. J Mol Biol. 2005;354(5):1069-1080. doi:10.1016/j.jmb.2005.10.023
133. Martin J. Lohse and Carsten Hoffmann. Arrestin Interactions with G Protein-Coupled Receptors. Vol 219.; 2014. doi:10.1007/978-3-642-41199-1
134. Sente A, Peer R, Srivastava A, et al. Molecular mechanism of modulating arrestin conformation by GPCR phosphorylation. Nat Struct Mol Biol. 2018;25(6):538-545. doi:10.1038/s41594-018-0071-3
135. Yang F, Yu X, Liu C, et al. Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and 19 F-NMR. Nat Commun. 2015;6(May):1-15. doi:10.1038/ncomms9202
136. Shukla AK, Westfield GH, Xiao K, et al. Visualization of arrestin recruitment by a G-protein-coupled receptor. Nature. 2014;512(7513):218-222. doi:10.1038/nature13430
137. Shukla AK, Manglik A, Kruse AC, et al. Structure of active β-arrestin-1 bound to a G-protein-coupled receptor phosphopeptide. Nature. 2013;497(7447):137-141. doi:10.1038/nature12120
138. Latorraca NR, Wang JK, Bauer B, et al. Molecular mechanism of GPCR-mediated arrestin activation. Nature. 2018;557.
139. Gurevich V V., Gurevich E V. The structural basis of the arrestin binding to GPCRs. Mol Cell Endocrinol. 2019;484(November 2018):34-41. doi:10.1016/j.mce.2019.01.019
140. Gurevich V V., Gurevich E V., Uversky VN. Arrestins: structural disorder creates rich functionality. Protein Cell. 2018;9(12):986-1003. doi:10.1007/s13238-017-0501-8
141. Kang Y, Zhou XE, Gao X, et al. Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. Nature. 2015;523(7562):561-567. doi:10.1038/nature14656
142. Kang DS, Kern RC, Puthenveedu MA, von Zastrow M, Williams JC, Benovic JL. Structure of an arrestin2-clathrin complex reveals a novel clathrin binding domain that modulates receptor trafficking. J Biol Chem. 2009;284(43):29860-29872. doi:10.1074/jbc.M109.023366
143. Gurevich V V., Gurevich E V. Analyzing the roles of multi-functional proteins in cells: The case of arrestins and GRKs. Crit Rev Biochem Mol Biol. 2015;50(5):440-452. doi:10.3109/10409238.2015.1067185
144. Lally CCM, Bauer B, Selent J, Sommer ME. C-edge loops of arrestin function as a membrane anchor. Nat Commun. 2017;8:1-12. doi:10.1038/ncomms14258
145. Kumari P, Srivastava A, Banerjee R, et al. Functional competence of a partially engaged GPCR-beta-arrestin complex. Nat Commun. 2016;7:13416. doi:10.1038/ncomms13416
146. Schoneberg T, Schulz A, Biebermann H, Hermsdorf T, Rompler H, Sangkuhl K. Mutant G-protein-coupled receptors as a cause of human diseases. Pharmacol Ther. 2004;104(3):173-206. doi:10.1016/j.pharmthera.2004.08.008
147. Zhang P, Covic L, Kuliopulos A. Pepducins and Other Lipidated Peptides as Mechanistic Probes and Therapeutics. Methods Mol Biol. 2015;1324:191-203. doi:10.1007/978-1-4939-2806-4_13

148. Zhang P, Leger AJ, Baleja JD, et al. Allosteric Activation of a G Protein-coupled Receptor with Cell-penetrating Receptor Mimetics. J Biol Chem. 2015;290(25):15785-15798. doi:10.1074/jbc.M115.636316
149. Vincenzi M, Costantini S, Scala S, et al. Conformational ensembles explored dynamically from disordered peptides targeting chemokine receptor CXCR4. Int J Mol Sci. 2015;16(6):12159-12173. doi:10.3390/ijms160612159
150. Yedvabny E, Nerenberg PS, So C, Head-Gordon T. Disordered structural ensembles of vasopressin and oxytocin and their mutants. J Phys Chem B. 2015;119(3):896-905. doi:10.1021/jp505902m
151. Germain P, Staels B, Dacquet C, Spedding M, Laudet V. Overview of nomenclature of nuclear receptors. Pharmacol Rev. 2006. doi:10.1124/pr.58.4.2
152. Gustafsson JA. Historical overview of nuclear receptors. J Steroid Biochem Mol Biol. 2016. doi:10.1016/j.jsbmb.2015.03.004
153. Weikum ER, Liu X, Ortlund EA. The nuclear receptor superfamily: A structural perspective. Protein Sci. 2018. doi:10.1002/pro.3496
154. Anbalagan M, Huderson B, Murphy L, Rowan BG. Post-translational modifications of nuclear receptors and human disease. Nucl Recept Signal. 2012. doi:10.1621/nrs.10001
155. Nagy L, Schwabe JWR. Mechanism of the nuclear receptor molecular switch. Trends Biochem Sci. 2004. doi:10.1016/j.tibs.2004.04.006
156. le Maire A, Alvarez S, Shankaranarayanan P, R de Lera A, Bourguet W, Gronemeyer H. Retinoid Receptors and Therapeutic Applications of RAR/RXR Modulators. Curr Top Med Chem. 2012;12(6):505-527. doi:10.2174/156802612799436687
157. Delfosse V, le Maire A, Balaguer P, Bourguet W. A structural perspective on nuclear receptors as targets of environmental compounds. Acta Pharmacol Sin. 2015. doi:10.1038/aps.2014.133
158. Motlagh HN, Hilser VJ. Agonism/antagonism switching in allosteric ensembles. Proc Natl Acad Sci U S A. 2012. doi:10.1073/pnas.1120519109
159. Onate SA, Boonyaratanakornkit V, Spencer TE, et al. The steroid receptor coactivator-1 contains multiple receptor interacting and activation domains that cooperatively enhance the activation function 1 (AF1) and AF2 domains of steroid receptors. J Biol Chem. 1998. doi:10.1074/jbc.273.20.12101
160. Chen W, Rogatsky I, Garabedian MJ. MED14 and MED1 differentially regulate target-specific gene activation by the glucocorticoid receptor. Mol Endocrinol. 2006. doi:10.1210/me.2005-0318
161. Choudhry MA, Ball A, McEwan IJ. The role of the general transcription factor IIF in androgen receptor-dependent transcription. Mol Endocrinol. 2006. doi:10.1210/me.2005-0486
162. Chandra V, Wu D, Li S, Potluri N, Kim Y, Rastinejad F. The quaternary architecture of RARβ-RXRα heterodimer facilitates domain-domain signal transmission. Nat Commun. 2017. doi:10.1038/s41467-017-00981-y
163. Lonard DM, O'Malley BW. The Expanding Cosmos of Nuclear Receptor Coactivators. Cell. 2006. doi:10.1016/j.cell.2006.04.021
164. Rosenfeld MG, Lunyak V V, Glass CK. Sensors and signals: A coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. Genes Dev. 2006. doi:10.1101/gad.1424806

165. Glass CK, Rosenfeld MG. The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev. 2000. doi:10.1101/gad.14.2.121
166. Lonard DM, O'Malley BW. Nuclear Receptor Coregulators: Judges, Juries, and Executioners of Cellular Regulation. Mol Cell. 2007. doi:10.1016/j.molcel.2007.08.012
167. Lazar MA. Nuclear receptor corepressors. Nucl Recept Signal. 2003. doi:10.1621/nrs.01001
168. Perissi V, Jepsen K, Glass CK, Rosenfeld MG. Deconstructing repression: Evolving models of co-repressor action. Nat Rev Genet. 2010. doi:10.1038/nrg2736
169. Tora L, Gronemeyer H, Turcotte B, Gaub MP, Chambon P. The N-terminal region of the chicken progesterone receptor specifies target gene activation. Nature. 1988. doi:10.1038/333185a0
170. Miranda TB, Morris SA, Hager GL. Complex genomic interactions in the dynamic regulation of transcription by the glucocorticoid receptor. Mol Cell Endocrinol. 2013. doi:10.1016/j.mce.2013.03.002
171. Lavery DN, McEwan IJ. Structure and function of steroid receptor AF1 transactivation domains: Induction of active conformations. Biochem J. 2005. doi:10.1042/BJ20050872
172. He B, Bai S, Hnat AT, et al. An androgen receptor NH2-terminal conserved motif interacts with the COOH terminus of the Hsp70-interacting protein (CHIP). J Biol Chem. 2004. doi:10.1074/jbc.M403117200
173. Wärnmark A, Treuter E, Wright APH, Gustafsson JÅ. Activation Functions 1 and 2 of Nuclear Receptors: Molecular Strategies for Transcriptional Activation. Mol Endocrinol. 2003. doi:10.1210/me.2002-0384
174. McEwan IJ, Lavery D, Fischer K, Watt K. Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. Nucl Recept Signal. 2007. doi:10.1621/nrs.05001
175. Huet G, Mérot Y, Percevault F, et al. Repression of the estrogen receptor-α transcriptional activity by the Rho/megakaryoblastic leukemia 1 signaling pathway. J Biol Chem. 2009. doi:10.1074/jbc.M109.045534
176. Simons SS, Edwards DP, Kumar R. Minireview: Dynamic structures of nuclear hormone receptors: New promises and challenges. Mol Endocrinol. 2014. doi:10.1210/me.2013-1334
177. Motlagh HN, Li J, Thompson EB, Hilser VJ. Interplay between allostery and intrinsic disorder in an ensemble. Biochem Soc Trans. 2012. doi:10.1042/BST20120163
178. Li J, White JT, Saavedra H, et al. Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor. Elife. 2017. doi:10.7554/eLife.30688
179. Metivier R, Petit FG, Valotaire Y, Pakdel F. Function of N-terminal transactivation domain of the estrogen receptor requires a potential α-helical structure and is negatively regulated by the A domain. Mol Endocrinol. 2000. doi:10.1210/mend.14.11.0546
180. Rajbhandari P, Finn G, Solodin NM, et al. Regulation of Estrogen Receptor N-Terminus Conformation and Function by Peptidyl Prolyl Isomerase Pin1. Mol Cell Biol. 2012. doi:10.1128/mcb.06073-11

181. Sadovsky Y, Webb P, Lopez G, et al. Transcriptional activators differ in their responses to overexpression of TATA-box-binding protein. Mol Cell Biol. 1995. doi:10.1128/mcb.15.3.1554
182. Wärnmark A, Wikström A, Wright APH, Gustafsson JÅ, Härd T. The N-terminal Regions of Estrogen Receptor α and β Are Unstructured in Vitro and Show Different TBP Binding Properties. J Biol Chem. 2001. doi:10.1074/jbc.M107875200
183. Kitagawa H, Yanagisawa J, Fuse H, et al. Ligand-Selective Potentiation of Rat Mineralocorticoid Receptor Activation Function 1 by a CBP-Containing Histone Acetyltransferase Complex. Mol Cell Biol. 2002. doi:10.1128/mcb.22.11.3698-3706.2002
184. Khan SH, Awasthi S, Guo C, et al. Binding of the N-terminal region of coactivator TIF2 to the intrinsically disordered AF1 domain of the glucocorticoid receptor is accompanied by conformational reorganizations. J Biol Chem. 2012. doi:10.1074/jbc.M112.411330
185. Simons SS, Kumar R. Variable steroid receptor responses: Intrinsically disordered AF1 is the key. Mol Cell Endocrinol. 2013. doi:10.1016/j.mce.2013.06.007
186. Gioeli D, Paschal BM. Post-translational modification of the androgen receptor. Mol Cell Endocrinol. 2012. doi:10.1016/j.mce.2011.07.004
187. Kumar R, McEwan IJ. Allosteric modulators of steroid hormone receptors: Structural dynamics and gene regulation. Endocr Rev. 2012. doi:10.1210/er.2011-1033
188. Kumar R, Thompson EB. Role of phosphorylation in the modulation of the glucocorticoid receptor's intrinsically disordered domain. Biomolecules. 2019. doi:10.3390/biom9030095
189. Kumar R, Moure CM, Khan SH, et al. Regulation of the structurally dynamic N-terminal domain of progesterone receptor by protein-induced folding. J Biol Chem. 2013. doi:10.1074/jbc.M113.491787
190. Kumar R, Thompson EB. Folding of the glucocorticoid receptor N-terminal transactivation function: Dynamics and regulation. Mol Cell Endocrinol. 2012. doi:10.1016/j.mce.2011.03.024
191. Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015. doi:10.1038/nrm3920
192. Brzovic PS, Heikaus CC, Kisselev L, et al. The Acidic Transcription Activator Gcn4 Binds the Mediator Subunit Gal11/Med15 Using a Simple Protein Interface Forming a Fuzzy Complex. Mol Cell. 2011. doi:10.1016/j.molcel.2011.11.008
193. Di Lello P, Jenkins LMM, Jones TN, et al. Structure of the Tfb1/p53 Complex: Insights into the Interaction between the p62/Tfb1 Subunit of TFIIH and the Activation Domain of p53. Mol Cell. 2006. doi:10.1016/j.molcel.2006.05.007
194. Feng H, Jenkins LMM, Durell SR, et al. Structural Basis for p300 Taz2-p53 TAD1 Binding and Modulation by Phosphorylation. Structure. 2009. doi:10.1016/j.str.2008.12.009
195. Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol. 2009. doi:10.1016/j.sbi.2008.12.003
196. Oldfield CJ, Dunker AK. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. Annu Rev Biochem. 2014. doi:10.1146/annurev-biochem-072711-164947

197. Van Der Lee R, Buljan M, Lang B, et al. Classification of intrinsically disordered regions and proteins. Chem Rev. 2014. doi:10.1021/cr400525m
198. Kumar R, Lee JC, Bolen DW, Thompson EB. The Conformation of the Glucocorticoid Receptor AF1/taul Domain Induced by Osmolyte Binds Co-regulatory Proteins. J Biol Chem. 2001. doi:10.1074/jbc.M100825200
199. Bain DL, Franden MA, McManaman JL, Takimoto GS, Horwitz KB. The N-terminal region of human progesterone B-receptors: Biophysical and biochemical comparison to A-receptors. J Biol Chem. 2001. doi:10.1074/jbc.M102611200
200. Reid J, Kelly SM, Watt K, Price NC, McEwan IJ. Conformational analysis of the androgen receptor amino-terminal domain involved in transactivation. Influence of structure-stabilizing solutes and protein-protein interactions. J Biol Chem. 2002. doi:10.1074/jbc.M201003200
201. Aagaard MM, Siersbæk R, Mandrup S. Molecular basis for gene-specific transactivation by nuclear receptors. Biochim Biophys Acta - Mol Basis Dis. 2011. doi:10.1016/j.bbadis.2010.12.018
202. Chandra V, Huang P, Hamuro Y, et al. Structure of the intact PPAR-γ-RXR-α nuclear receptor complex on DNA. Nature. 2008. doi:10.1038/nature07413
203. Zhang J, Chalmers MJ, Stayrook KR, et al. DNA binding alters coactivator interaction surfaces of the intact VDR-RXR complex. Nat Struct Mol Biol. 2011. doi:10.1038/nsmb.2046
204. Belorusova AY, Osz J, Petoukhov M V, et al. Solution Behavior of the Intrinsically Disordered N-Terminal Domain of Retinoid X Receptor α in the Context of the Full-Length Protein. Biochemistry. 2016. doi:10.1021/acs.biochem.5b01122
205. Baskakov I V, Kumar R, Srinivasan G, Ji YS, Wayne Bolen D, Thompson EB. Trimethylamine N-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor. J Biol Chem. 1999. doi:10.1074/jbc.274.16.10693
206. Fischer K, Kelly SM, Watt K, Price NC, McEwan IJ. Conformation of the mineralocorticoid receptor N-terminal domain: Evidence for induced and stable structure. Mol Endocrinol. 2010. doi:10.1210/me.2010-0005
207. Kumar R, Serrette JM, Khan SH, Miller AL, Thompson EB. Effects of different osmolytes on the induced folding of the N-terminal activation domain (AF1) of the glucocorticoid receptor. Arch Biochem Biophys. 2007. doi:10.1016/j.abb.2007.06.019
208. Kumar R, Betney R, Li J, Thompson EB, McEwan IJ. Induced alpha-Helix Structure in AF1 of the Androgen Receptor upon Binding Transcription Factor TFIIF. Biochemistry. 2004. doi:10.1021/bi035934p
209. Kumar R, Volk DE, Li J, Lee JC, Gorenstein DG, Thompson EB. TATA box binding protein induces structure in the recombinant glucocorticoid receptor AF1 domain. Proc Natl Acad Sci U S A. 2004. doi:10.1073/pnas.0407160101
210. Wardell SE, Kwok SC, Sherman L, Hodges RS, Edwards DP. Regulation of the Amino-Terminal Transcription Activation Domain of Progesterone Receptor by a Cofactor-Induced Protein Folding Mechanism. Mol Cell Biol. 2005. doi:10.1128/mcb.25.20.8792-8808.2005
211. Wood JR. Allosteric Modulation of Estrogen Receptor Conformation by Different Estrogen Response Elements. Mol Endocrinol. 2001. doi:10.1210/me.15.7.1114

212. Brodie J, McEwan IJ. Intra-domain communication between the N-terminal and DNA-binding domains of the androgen receptor: Modulation of androgen response element DNA binding. J Mol Endocrinol. 2005. doi:10.1677/jme.1.01723
213. Li W, Cavasotto CN, Cardozo T, et al. Androgen receptor mutations identified in prostate cancer and androgen insensitivity syndrome display aberrant ARTt-27 coactivator function. Mol Endocrinol. 2005. doi:10.1210/me.2005-0134
214. Betney R, McEwan IJ. Role of conserved hydrophobic amino acids in androgen receptor AF-1 function. J Mol Endocrinol. 2003. doi:10.1677/jme.0.0310427
215. Lavery DN, McEwan IJ. The human androgen receptor AF1 transactivation domain: Interactions with transcription factor IIF and molten-globule-like structural characteristics. Biochem Soc Trans. 2006. doi:10.1042/BST0341054
216. Lavery DN, McEwan IJ. Structural characterization of the native NH2-terminal transactivation domain of the human androgen receptor: A collapsed disordered conformation underlies structural plasticity and protein-induced folding. Biochemistry. 2008. doi:10.1021/bi702221e
217. Peng Y, Cao S, Kiselar J, et al. A Metastable Contact and Structural Disorder in the Estrogen Receptor Transactivation Domain. Structure. 2019. doi:10.1016/j.str.2018.10.026
218. Bah A, Forman-Kay JD. Modulation of intrinsically disordered protein function by post-translational modifications. J Biol Chem. 2016. doi:10.1074/jbc.R115.695056
219. Baker JMR, Hudson RP, Kanelis V, et al. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. Nat Struct Mol Biol. 2007. doi:10.1038/nsmb1278
220. He Y, Chen Y, Mooney SM, et al. Phosphorylation-induced conformational ensemble switching in an intrinsically disordered cancer/testis antigen. J Biol Chem. 2015. doi:10.1074/jbc.M115.658583
221. Bah A, Vernon RM, Siddiqui Z, et al. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature. 2015. doi:10.1038/nature13999
222. Garza AMS, Khan SH, Kumar R. Site-Specific Phosphorylation Induces Functionally Active Conformation in the Intrinsically Disordered N-Terminal Activation Function (AF1) Domain of the Glucocorticoid Receptor. Mol Cell Biol. 2010. doi:10.1128/mcb.00552-09
223. Khan SH, McLaughlin WA, Kumar R. Site-specific phosphorylation regulates the structure and function of an intrinsically disordered domain of the glucocorticoid receptor. Sci Rep. 2017. doi:10.1038/s41598-017-15549-5
224. De Mol E, Szulc E, Di Sanza C, et al. Regulation of Androgen Receptor Activity by Transient Interactions of Its Transactivation Domain with General Transcription Regulators. Structure. 2018. doi:10.1016/j.str.2017.11.007
225. Kulkarni P, Jolly MK, Jia D, et al. Phosphorylation-induced conformational dynamics in an intrinsically disordered protein and potential role in phenotypic heterogeneity. Proc Natl Acad Sci U S A. 2017. doi:10.1073/pnas.1700082114
226. Dunker AK, Uversky VN. Signal transduction via unstructured protein conduits. Nat Chem Biol. 2008. doi:10.1038/nchembio0408-229
227. Borg M, Mittag T, Pawson T, Tyers M, Forman-Kay JD, Hue SC. Polyelectrostatic interactions of disordered ligands suggest a physical basis

for ultrasensitive. Proc Natl Acad Sci U S A. 2007. doi:10.1073/pnas.0702580104
228. Mittag T, Kay LE, Forman-Kaya JD. Protein dynamics and conformational disorder in molecular recognition. J Mol Recognit. 2010. doi:10.1002/jmr.961
229. Csizmok V, Follis AV, Kriwacki RW, Forman-Kay JD. Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling. Chem Rev. 2016. doi:10.1021/acs.chemrev.5b00548
230. Hegyi H, Schad E, Tompa P. Structural disorder promotes assembly of protein complexes. BMC Struct Biol. 2007. doi:10.1186/1472-6807-7-65
231. Millard CJ, Watson PJ, Fairall L, Schwabe JWR. An evolving understanding of nuclear receptor coregulator proteins. J Mol Endocrinol. 2013. doi:10.1530/JME-13-0227
232. Ahmad KF, Melnick A, Lax S, et al. Mechanism of SMRT Corepressor Recruitment by the BCL6 BTB Domain. Mol Cell. 2003. doi:10.1016/S1097-2765(03)00454-4
233. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. Intrinsic disorder in transcription factors. Biochemistry. 2006. doi:10.1021/bi0602718
234. Hu X, Lazar MA. The CoRNR motif controls the recruitment of compressors by nuclear hormone receptors. Nature. 1999. doi:10.1038/47069
235. Nagy L, Kao HY, Love JD, et al. Mechanism of corepressor binding and release from nuclear hormone receptors. Genes Dev. 1999;13(24):3209-3216. doi:10.1101/gad.13.24.3209
236. Nolte RT, Wisely GB, Westin S, et al. Ligand binding and co-activator assembly of the peroxisome proliferator- activated receptor-γ. Nature. 1998. doi:10.1038/25931
237. Xu HE, Stanley TB, Montana VG, et al. Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARα. Nature. 2002. doi:10.1038/415813a
238. Wang CY, Li CW, Chen JD, Welsh WJ. Structural model reveals key interactions in the assembly of the pregnane X receptor/corepressor complex. Mol Pharmacol. 2006. doi:10.1124/mol.106.022368
239. Le Maire A, Teyssier C, Erb C, et al. A unique secondary-structure switch controls constitutive gene repression by retinoic acid receptor. Nat Struct Mol Biol. 2010. doi:10.1038/nsmb.1855
240. Chang C, Norris JD, Grøn H, et al. Dissection of the LXXLL Nuclear Receptor-Coactivator Interaction Motif Using Combinatorial Peptide Libraries: Discovery of Peptide Antagonists of Estrogen Receptors α and β. Mol Cell Biol. 1999. doi:10.1128/mcb.19.12.8226
241. Darimont BD, Wagner RL, Apriletti JW, et al. Structure and specificity of nuclear receptor-coactivator interactions. Genes Dev. 1998. doi:10.1101/gad.12.21.3343
242. McInerney EM, Rose DW, Flynn SE, et al. Determinants of coactivator LXXLL motif specificity in nuclear receptor transcriptional activation. Genes Dev. 1998. doi:10.1101/gad.12.21.3357
243. Devarakonda S, Gupta K, Chalmers MJ, et al. Disorder-to-order transition underlies the structural basis for the assembly of a transcriptionally active PGC-1α/ERRγ complex. Proc Natl Acad Sci U S A. 2011. doi:10.1073/pnas.1113813108

244. de Vera IMS, Zheng J, Novick S, et al. Synergistic Regulation of Coregulator/Nuclear Receptor Interaction by Ligand and DNA. Structure. 2017. doi:10.1016/j.str.2017.07.019
245. Mohan A, Oldfield CJ, Radivojac P, et al. Analysis of Molecular Recognition Features (MoRFs). J Mol Biol. 2006. doi:10.1016/j.jmb.2006.07.087
246. Iešmantavičius V, Dogan J, Jemth P, Teilum K, Kjaergaard M. Helical propensity in an intrinsically disordered protein accelerates ligand binding. Angew Chemie - Int Ed. 2014. doi:10.1002/anie.201307712
247. Hu X, Li Y, Lazar MA. Determinants of CoRNR-Dependent Repression Complex Assembly on Nuclear Hormone Receptors. Mol Cell Biol. 2001. doi:10.1128/mcb.21.5.1747-1758.2001
248. Perissi V, Staszewski LM, McInerney EM, et al. Molecular determinants of nuclear receptor-corepressor interaction. Genes Dev. 1999. doi:10.1101/gad.13.24.3198
249. Cordeiro TN, Sibille N, Germain P, et al. Interplay of Protein Disorder in Retinoic Acid Receptor Heterodimer and Its Corepressor Regulates Gene Expression. Structure. 2019. doi:10.1016/j.str.2019.05.001
250. Delerive P, Wu Y, Burris TP, Chin WW, Suen CS. PGC-1 functions as a transcriptional coactivator for the retinoid X receptors. J Biol Chem. 2002. doi:10.1074/jbc.M109409200
251. Bourdoncle A, Labesse G, Margueron R, Castet A, Cavaillès V, Royer CA. The nuclear receptor coactivator PGC-1α exhibits modes of interaction with the estrogen receptor distinct from those of SRC-1. J Mol Biol. 2005. doi:10.1016/j.jmb.2005.01.048
252. Rha GB, Wu G, Shoelson SE, Chi YC. Multiple binding modes between HNF4α and the LXXLL motifs of PGC-1α lead to full activation. J Biol Chem. 2009. doi:10.1074/jbc.M109.052506
253. Phelan CA, Gampe RT, Lambert MH, et al. Structure of Rev-erbα bound to N-CoR reveals a unique mechanism of nuclear receptor-co-repressor interaction. Nat Struct Mol Biol. 2010. doi:10.1038/nsmb.1860
254. Rochel N, Ciesielski F, Godet J, et al. Common architecture of nuclear receptor heterodimers on DNA direct repeat elements with different spacings. Nat Struct Mol Biol. 2011. doi:10.1038/nsmb.2054
255. Osz J, Pethoukhov M V., Sirigu S, Svergun DI, Moras D, Rochel N. Solution structures of PPAR γ 2/RXR α complexes. PPAR Res. 2012. doi:10.1155/2012/701412
256. Takacs M, Petoukhov M V., Atkinson RA, et al. The Asymmetric Binding of PGC-1α to the ERRα and ERRγ Nuclear Receptor Homodimers Involves a Similar Recognition Mechanism. PLoS One. 2013. doi:10.1371/journal.pone.0067810
257. Longhi S, Receveur-Bréchot V, Karlin D, et al. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J Biol Chem. 2003. doi:10.1074/jbc.M300518200
258. Marsh JA, Dancheck B, Ragusa MJ, Allaire M, Forman-Kay JD, Peti W. Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. Structure. 2010. doi:10.1016/j.str.2010.05.015

259. Gosselin P, Oulhen N, Jam M, et al. The translational repressor 4E-BP called to order by eIF4E: New structural insights by SAXS. Nucleic Acids Res. 2011. doi:10.1093/nar/gkq1306
260. Pavlin MR, Brunzelle JS, Fernandez EJ. Agonist ligands mediate the transcriptional response of nuclear receptor heterodimers through distinct stoichiometric assemblies with coactivators. J Biol Chem. 2014. doi:10.1074/jbc.M114.575423
261. Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, Yamamoto KR. DNA binding site sequence directs glucocorticoid receptor structure and activity. Science (80-). 2009. doi:10.1126/science.1164265
262. Takeshita A, Yen PM, Ikeda M, et al. Thyroid hormone response elements differentially modulate the interactions of thyroid hormone receptors with two receptor binding domains in the steroid receptor coactivator-1. J Biol Chem. 1998. doi:10.1074/jbc.273.34.21554
263. Putcha BDK, Fernandez EJ. Direct interdomain interactions can mediate allosterism in the thyroid receptor. J Biol Chem. 2009. doi:10.1074/jbc.M109.026682
264. Velasco LFR, Togashi M, Walfish PG, et al. Thyroid hormone response element organization dictates the composition of active receptor. J Biol Chem. 2007. doi:10.1074/jbc.M610700200
265. Heneghan AF, Connaghan-Jones KD, Miura MT, Bain DL. Coactivator assembly at the promoter: Efficient recruitment of SRC2 is coupled to cooperative DNA binding by the progesterone receptor. Biochemistry. 2007. doi:10.1021/bi700850v
266. Monaghan AE, McEwan IJ. A sting in the tail: The N-terminal domain of the androgen receptor as a drug target. Asian J Androl. 2016. doi:10.4103/1008-682X.181081
267. Shamilov R, Aneskievich BJ. Intrinsic Disorder in Nuclear Receptor Amino Termini: From Investigational Challenge to Therapeutic Opportunity. Nucl Recept Res. 2019. doi:10.32527/2019/101417
268. Zhang Z, Boskovic Z, Hussain MM, et al. Chemical perturbation of an intrinsically disordered region of TFIID distinguishes two modes of transcription initiation. Elife. 2015. doi:10.7554/eLife.07777
269. Dunker AK, Uversky VN. Drugs for “protein clouds”: Targeting intrinsically disordered transcription factors. Curr Opin Pharmacol. 2010. doi:10.1016/j.coph.2010.09.005
270. Santofimia-Castaño P, Rizzuti B, Xia Y, et al. Targeting intrinsically disordered proteins involved in cancer. Cell Mol Life Sci. 2019. doi:10.1007/s00018-019-03347-3
271. Kumar R. Steroid hormone receptors and prostate cancer: Role of structural dynamics in therapeutic targeting. Asian J Androl. 2016. doi:10.4103/1008-682X.183380
272. Myung JK, Banuelos CA, Fernandez JG, et al. An androgen receptor N-terminal domain antagonist for treating prostate cancer. J Clin Invest. 2013. doi:10.1172/JCI66398
273. Andersen RJ, Mawji NR, Wang J, et al. Regression of Castrate-Recurrent Prostate Cancer by a Small-Molecule Inhibitor of the Amino-Terminus Domain of the Androgen Receptor. Cancer Cell. 2010. doi:10.1016/j.ccr.2010.04.027

274. Bushweller JH. Targeting transcription factors in cancer — from undruggable to reality. Nat Rev Cancer. 2019. doi:10.1038/s41568-019-0196-7
275. Spada ARL, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 1991. doi:10.1038/352077a0
276. Chen S, Ferrone FA, Wetzel R. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. Proc Natl Acad Sci U S A. 2002. doi:10.1073/pnas.182276099
277. Eftekharzadeh B, Piai A, Chiesa G, et al. Sequence Context Influences the Structure and Aggregation Behavior of a PolyQ Tract. Biophys J. 2016. doi:10.1016/j.bpj.2016.04.022
278. Davies P, Watt K, Kelly SM, Clark C, Price NC, McEwan IJ. Consequences of poly-glutamine repeat length for the conformation and folding of the androgen receptor amino-terminal domain. J Mol Endocrinol. 2008. doi:10.1677/JME-08-0042
279. Escobedo A, Topal B, Kunze MBA, et al. Side chain to main chain hydrogen bonds stabilize a polyglutamine helix in a transcription factor. Nat Commun. 2019. doi:10.1038/s41467-019-09923-2
280. Boija A, Klein IA, Sabari BR, et al. Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. Cell. 2018. doi:10.1016/j.cell.2018.10.042
281. Brangwynne CP, Eckmann CR, Courson DS, et al. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science (80-). 2009. doi:10.1126/science.1172046
282. Banjade S, Wu Q, Mittal A, Peeples WB, Pappu R V, Rosen MK. Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. Proc Natl Acad Sci U S A. 2015. doi:10.1073/pnas.1508778112
283. Zhu L, Brangwynne CP. Nuclear bodies: The emerging biophysics of nucleoplasmic phases. Curr Opin Cell Biol. 2015. doi:10.1016/j.ceb.2015.04.003
284. Sołtys K, Ożyhar A. Ordered structure-forming properties of the intrinsically disordered AB region of hRXRγ and its ability to promote liquid-liquid phase separation. J Steroid Biochem Mol Biol. 2020. doi:10.1016/j.jsbmb.2019.105571
285. Boehning M, Dugast-Darzacq C, Rankovic M, et al. RNA polymerase II clustering through carboxy-terminal domain phase separation. Nat Struct Mol Biol. 2018. doi:10.1038/s41594-018-0112-y
286. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. A Phase Separation Model for Transcriptional Control. Cell. 2017. doi:10.1016/j.cell.2017.02.007
287. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 1958;181(4610):662-666. doi:10.1038/181662a0
288. Dunker AK, Garner E, Guilliot S, et al. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pac Symp Biocomput. 1998:473-484.
289. Kriwacki RW, Hengst L, Tennant L, Reed SI, Wright PE. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: Conformational disorder

mediates binding diversity. Proc Natl Acad Sci U S A. 1996;93(21):11504-11509. doi:10.1073/pnas.93.21.11504
290. Shikama N, Ivon J, La Thangue N. The p300/CBP family: Integrating signals with transcription factors and chromatin. Trends Cell Biol. 1997;7(6):230-236. doi:10.1016/S0962-8924(97)01048-9
291. Janknecht R, Hunter T. Transcriptional control: Versatile molecular glue. Curr Biol. 1996;6(8):951-954. doi:10.1016/S0960-9822(02)00636-X

Annexe 2

Partie 3 : La phosphorylation par les GRKs module la région d'interaction RCPG:arrestine

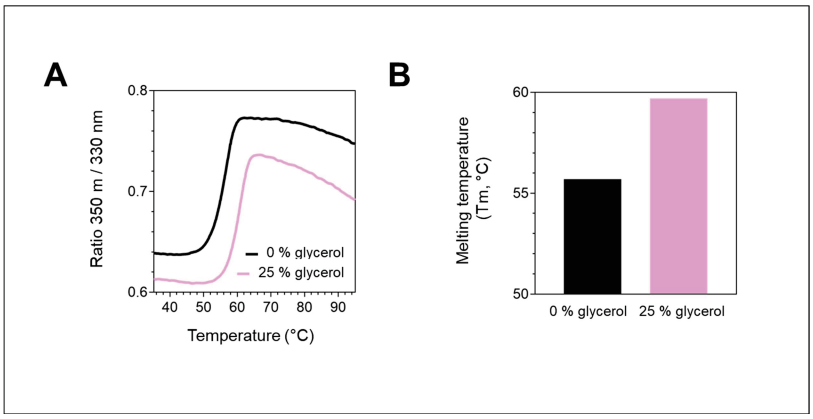


Figure S1 : Reliée à la Figure 50. Effet du glycérol sur la stabilité de GRK2. A) Ratio 350/330 nm issu de la fluorescence intrinsèque des résidus tryptophanes et tyrosines en fonction de la température. **B)** Histogrammes représentant la température de dénaturation de GRK2 sans et avec 25 % de glycérol après conservation. La température de dénaturation est plus élevée en présence de glycérol (Tm = 55.7 °C sans glycérol ; Tm = 59.7 °C avec glycérol). Le glycérol semble permettre une meilleure stabilité de GRK2 après décongélation.

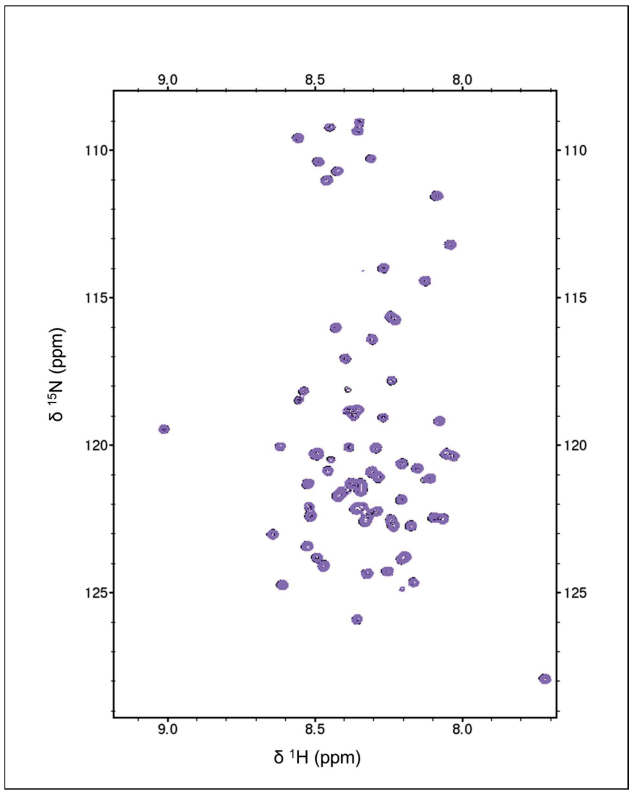


Figure S2 : Reliée à la Figure 50. Le chauffage dont on se sert pour stopper la réaction de phosphorylation n'affecte pas la protéine. Après phosphorylation par GRK2, l'échantillon de 10 µM de β2AR-Cter a été divisé en deux. Une aliquote a été chauffée 10 min à 75 °C puis dialysée en tampon RMN, alors que l'autre a directement été mise en dialyse sans chauffage. Après dialyse, les deux échantillons ont été analysés sur le 800 MHz, à 20 °C, en tampon 50 mM Bis-Tris pH 6.7, 150 mM NaCl, 2 mM DTT. Le spectre correspondant à l'échantillon non-chauffé est en noir et le spectre correspondant à l'échantillon chauffé est en violet. Ces deux spectres sont parfaitement superposés, ce qui indique que le chauffage n'affecte pas la protéine.

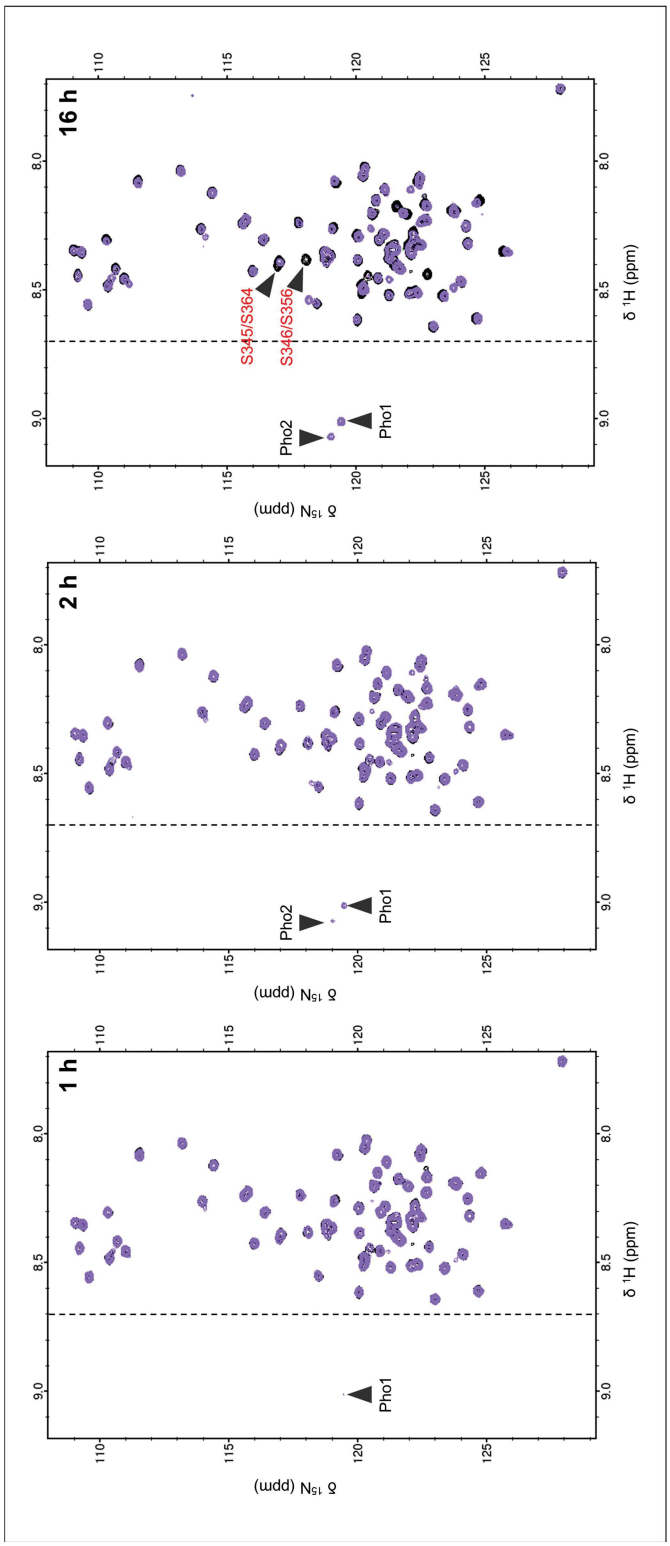


Figure S3 : Reliée à la Figure 51. Cinétique de phosphorylation du peptide β2AR par GRK2. Les temps de réaction de phosphorylation sont indiqués en haut à gauche des spectres. Les spectres phosphorylés (violet) sont superposés à un spectre de référence non-phosphorylé (noir). Sur chaque spectre, les nouvelles résonances apparaissant à bas champs (après 8.7 ppm, ligne noire pointillée) sont indiquées Pho1 et Pho2 en fonction de leur ordre d'apparition. Sur le spectre correspondant à 16h de réaction, les sérines potentiellement phosphorylées sont indiquées en rouge. Les spectres ont été enregistrés sur des échantillons à 10 µM, sur le 800 MHz, à 20°C, dans 50 mM Bis-Tris pH 6.7, 150 mM NaCl.

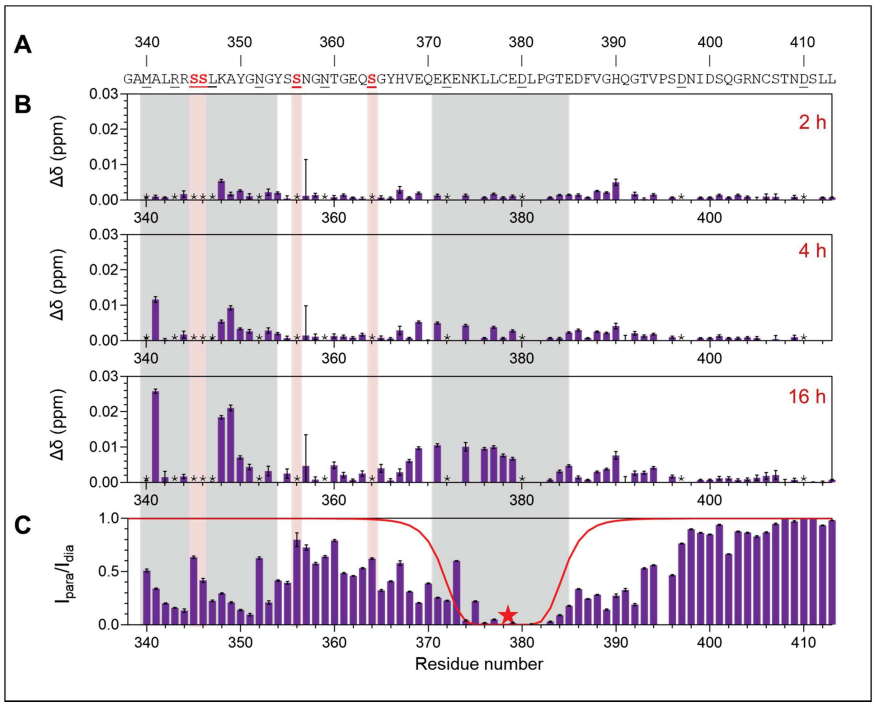


Figure S4 : Reliée à la Figure 52. La phosphorylation entraine des changements de déplacement chimique dans les régions en interaction longue distance. A) Séquence de β2AR-Cter. Les sérines potentiellement phosphorylées sont indiquées en rouge et les résidus superposés dans les spectres ^{15}N-HSQC sous soulignés et indiqués par une étoile noire dans les histogrammes. **B)** La déviation de déplacement chimique au cours du temps (2h, 4h et 16h) montrent deux régions affectées par la phosphorylation : la région autour des sérines 345-346, et la région allant de ~ 370 à 380. Cette observation inattendue peut s'expliquer par la présence d'interaction longue distance entre ces régions identifiées précédemment par les PREs (**C**), chapitre 4, Article 1, p. 129)

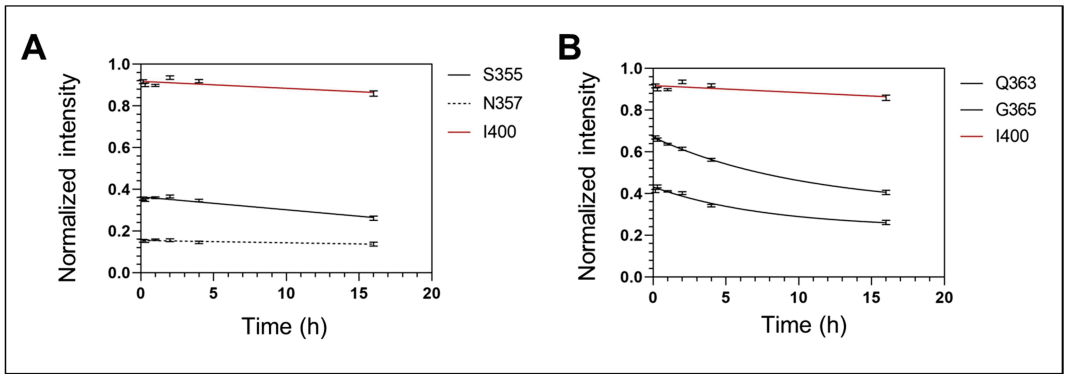


Figure S5 : Reliée à la Figure 52. Suivi de l'intensité normalisée des résidus voisins aux sérines S356 (A) et S364 (B) dans le temps. Comparé à l'intensité de I400 (résidu contrôle non affecté par la phosphorylation), les voisins de la sérine S356 ne sont pas affectés par la phosphorylation. Pour ceux de S364, une légère diminution de l'intensité de Q363 et G365 peut être observée.

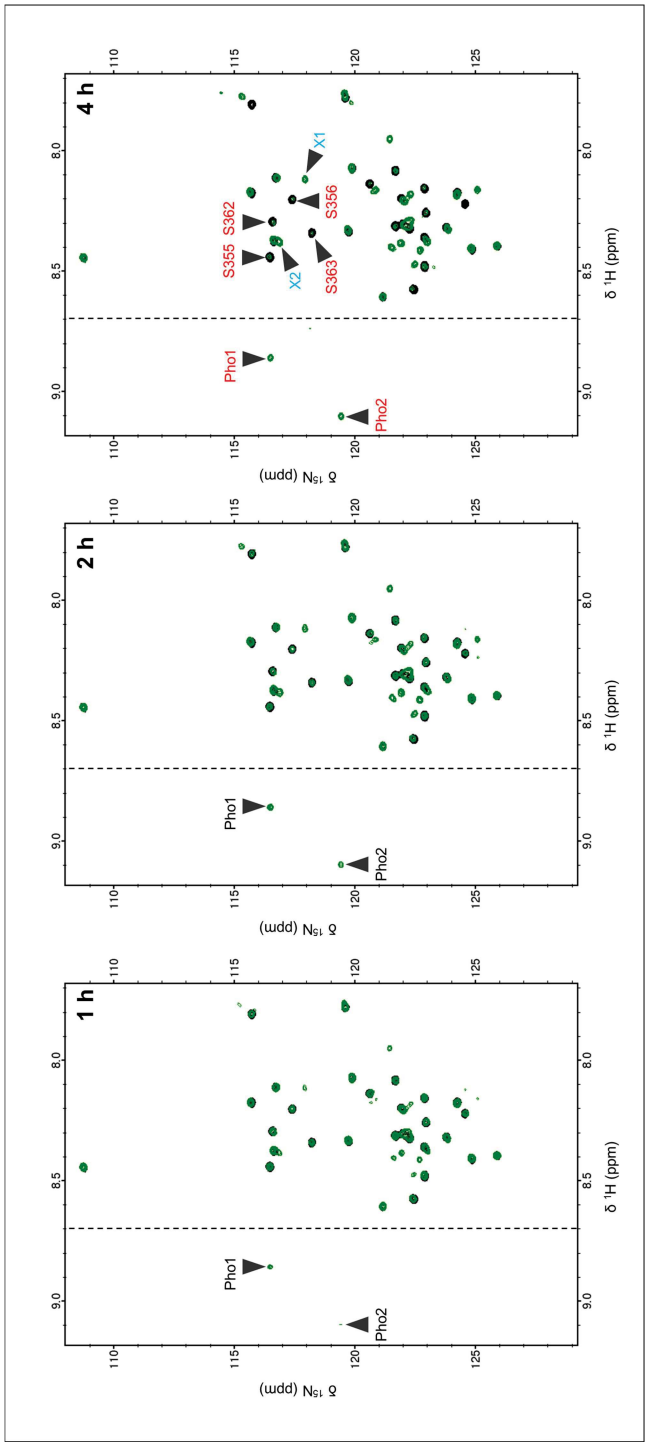


Figure S6 : Reliée à la Figure 53. Cinétique de phosphorylation du peptide GHSR par GRK2. Les temps de réaction de phosphorylation sont indiqués en haut à gauche des spectres. Les spectres phosphorylés (vert) sont superposés à un spectre de référence non-phosphorylé (noir). Sur chaque spectre, les nouvelles résonances apparaissant à bas champs (après 8.7 ppm, ligne noire pointillée) sont indiquées Pho1 et Pho2 en fonction de leur ordre d'apparition. Sur le spectre correspondant à 4 h de réaction, les sérines pouvant correspondre à Pho1 et Pho2 sont indiquées en rouge. Deux nouvelles résonances apparaissent dans la région des sérines, X1 et X2, et sont représentées en bleu. Les spectres ont été enregistrés sur des échantillons à 10 µM, sur le 800 MHz, à 20°C, dans 50 mM Bis-Tris pH 6.7, 150 mM NaCl.

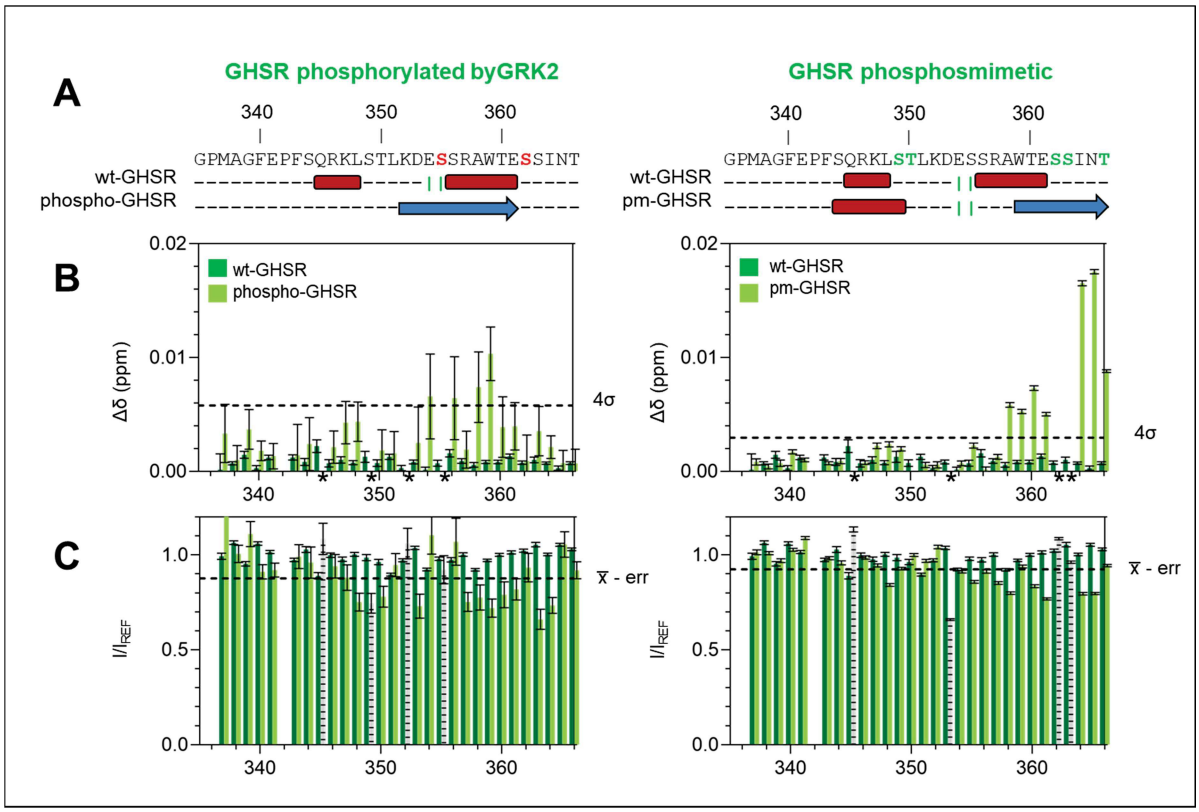


Figure S7 : Reliée à la Figure 56. Interaction entre l'arrestine et GHSR-Cter phosphorylé par GRK2 (pjospho-GHSR, à gauche) ou GHSR phosphomimétique (pm-GHSR, à droite). A) Séquence de GHSR-Cter. Les sérines phosphorylées par GRK2 sont indiquées en rouge et les résidus phosphomimétiques sont indiqués en vert. Les profils de structuration de GHSR non-phosphorylé (wt-GHSR), phospho-GHSR et pm-GHSR, obtenus par RMN, sont représentés sous la séquence. **B)** Histogramme de la différence de déplacement chimique $^{1}H/^{15}N$ (ppm) observée par interaction entre l'arrestine 2 et wt-GHSR (vert foncé) ou phospho-GHSR (vert clair) ou pm-GHSR (vert clair). **C)** Histogramme de la différence d'intensité observée par interaction entre l'arrestine 2 et wt-GHSR (vert foncé) ou phospho-GHSR (vert clair) ou pm-GHSR (vert clair). Les interactions ont été enregistrées à 20 °C, pH 6.7, 50 mM NaCl. Les cut-offs sont notés en ligne noire pointillée. Les résidus superposés dans les ^{15}N-HSQCs sont représentés par des étoiles rouges dans les histogrammes (**B**) et sont hachés en gris sur les histogrammes (**C**).

BIBLIOGRAPHIE

Ahmed MR, Bychkov E, Gurevich VV, Benovic JL & Gurevich EV (2008) Altered expression and subcellular distribution of GRK subtypes in the dopamine-depleted rat basal ganglia is not normalized by l-DOPA treatment: GRKs expression and dopamine depletion. *J Neurochem* 104: 1622–1636

Ahn S, Shenoy SK, Luttrell LM & Lefkowitz RJ (2020) SnapShot: β-Arrestin Functions. *Cell* 182: 1362-1362.e1

Altosaar K, Balaji P, Bond RA, Bylund DB, Cotecchia S, Devost D, Doze VA, Eikenburg DC, Gora S, Goupil E, *et al* (2019) Adrenoceptors (version 2019.4) in the IUPHAR/BPS Guide to Pharmacology Database. *IUPHARBPS Guide Pharmacol CITE* 2019

Asai D, Murata M, Toita R, Kawano T, Nakashima H & Kang JH (2016) Role of amino acid residues surrounding the phosphorylation site in peptide substrates of G protein-coupled receptor kinase 2 (GRK2). *Amino Acids* 48: 2875–2880

Asai D, Toita R, Murata M, Katayama Y, Nakashima H & Kang J-H (2014) Peptide substrates for G protein-coupled receptor kinase 2. *FEBS Lett* 588: 2129–2132

Attramadal H, Arriza JL, Aoki C, Dawson TM, Codina J, Kwatra MM, Snyder SH, Caron MG & Lefkowitz RJ (1992) Beta-arrestin2, a novel member of the arrestin/beta-arrestin gene family. *J Biol Chem* 267: 17882–17890

Bah A, Vernon RM, Siddiqui Z, Krzeminski M, Muhandiram R, Zhao C, Sonenberg N, Kay LE & Forman-Kay JD (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* 519: 106–109

Bax A, Kontaxis G & Tjandra N (2001) Dipolar Couplings in Macromolecular Structure Determination. In *Methods in Enzymology* pp 127–174. Elsevier

Benovic JL (1991) [28] Purification and characterization of β-adrenergic receptor kinase. In *Methods in Enzymology* pp 351–362. Academic Press

Benovic JL & Gomez J (1993) Molecular cloning and expression of GRK6. A new member of the G protein-coupled receptor kinase family. *J Biol Chem* 268: 19521–19527

Benovic JL, Kühn H, Weyand I, Codina J, Caron MG & Lefkowitz RJ (1987) Functional desensitization of the isolated beta-adrenergic receptor by the beta-adrenergic receptor kinase: potential role of an analog of the retinal protein arrestin (48-kDa protein). *Proc Natl Acad Sci U S A* 84: 8879–8882

Benovic JL, Onorato J, Lohse MJ, Dohlman HG, Staniszewski C, Caron MG & Lefkowitz RJ (1990) Synthetic peptides of the hamster beta 2-adrenoceptor as substrates and inhibitors of the beta-adrenoceptor kinase. *Br J Clin Pharmacol* 30 Suppl 1: 3S-12S

Benovic JL, Onorato JJ, Arriza JL, Stone WC, Lohse M, Jenkins NA, Gilbert DJ, Copeland NG, Caron MG & Lefkowitz RJ (1991) Cloning, expression, and chromosomal localization of beta-adrenergic receptor kinase 2. A new member of the receptor kinase family. *J Biol Chem* 266: 14939–14946

Benovic JL, Strasser RH, Caron MG & Lefkowitz RJ (1986) Beta-adrenergic receptor kinase: identification of a novel protein kinase that phosphorylates the agonist-occupied form of the receptor. *Proc Natl Acad Sci U S A* 83: 2797–2801

Bernadó P, Bertoncini CW, Griesinger C, Zweckstetter M & Blackledge M (2005a) Defining Long-Range Order and Local Disorder in Native α-Synuclein Using Residual Dipolar Couplings. *J Am Chem Soc* 127: 17968–17969

Bernadó P, Blanchard L, Timmins P, Marion D, Ruigrok RWH & Blackledge M (2005b) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci* 102: 17002–17007

Bizien T, Durand D, Roblina P, Thureau A, Vachette P & Pérez J (2016) A Brief Survey of State-of-the-Art BioSAXS. *Protein Pept Lett* 23: 217–231

Blanchet CE & Svergun DI (2013) Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu Rev Phys Chem* 64: 37–54

Blom N, Sicheritz-Pontén T, Gupta R, Gammeltoft S & Brunak S (2004) Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *PROTEOMICS* 4: 1633–1649

Bockaert J & Philippe Pin J (1999) Molecular tinkering of G protein-coupled receptors: an evolutionary success. *EMBO J* 18: 1723–1729

Boguth CA, Singh P, Huang C & Tesmer JJG (2010) Molecular basis for activation of G protein-coupled receptor kinases. *EMBO J* 29: 3249–3259

Bouzo-Lorenzo M, Santo-Zas I, Lodeiro M, Nogueiras R, Casanueva FF, Castro M, Pazos Y, Tobin AB, Butcher AJ & Camiña JP (2016) Distinct phosphorylation sites on the ghrelin receptor, GHSR1a, establish a code that determines the functions of ß-arrestins. *Sci Rep* 6: 22495

Brinkley M (1992) A brief survey of methods for preparing protein conjugates with dyes, haptens and crosslinking reagents. *Bioconjug Chem* 3: 2–13

Buchan DWA & Jones DT (2019) The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res* 47: W402–W407

Butcher AJ, Prihandoko R, Kong KC, McWilliams P, Edwards JM, Bottrill A, Mistry S & Tobin AB (2011) Differential G-protein-coupled receptor phosphorylation provides evidence for a signaling bar code. *J Biol Chem* 286: 11506–11518

Bychkov ER, Ahmed MR, Gurevich VV, Benovic JL & Gurevich EV (2011) Reduced expression of G protein-coupled receptor kinases in schizophrenia but not in schizoaffective disorder. *Neurobiol Dis* 44: 248–258

Cahill TJ, Thomsen ARB, Tarrasch JT, Plouffe B, Nguyen AH, Yang F, Huang L-Y, Kahsai AW, Bassoni DL, Gavino BJ, *et al* (2017) Distinct conformations of GPCR–β-arrestin complexes mediate desensitization, signaling, and endocytosis. *Proc Natl Acad Sci* 114: 2562–2567

Canet E, Kateb F, Segawa TF, Sehgal AA, Bodenhausen G & Pelupessy P (2017) Mesurer l'échange rapide de protons par RMN. *Actual Chim* 416: 7

Carman CV & Benovic JL (1998) G-protein-coupled receptors: turn-ons and turn-offs. *Curr Opin Neurobiol* 8: 335–344

Caron MG & Barak LS (2019) A Brief History of the β-Arrestins. In *Beta-Arrestins: Methods and Protocols*, Scott MGH & Laporte SA (eds) pp 3–8. New York, NY: Springer

Chaturvedi M, Maharana J & Shukla AK (2020) Terminating G-Protein Coupling: Structural Snapshots of GPCR-β-Arrestin Complexes. *Cell* 0

Chemes LB, Alonso LG, Noval MG & de Prat-Gay G (2012) Circular Dichroism Techniques for the Analysis of Intrinsically Disordered Proteins and Domains. In *Intrinsically Disordered Protein Analysis: Volume 1, Methods and Experimental Tools*, Uversky VN & Dunker AK (eds) pp 387–404. Totowa, NJ: Humana Press

Chen CY, Dion SB, Kim CM & Benovic JL (1993) Beta-adrenergic receptor kinase. Agonist-dependent receptor binding promotes kinase activation. *J Biol Chem* 268: 7825–7831

Chen K & Tjandra N (2011) The Use of Residual Dipolar Coupling in Studying Proteins by NMR. In *NMR of Proteins and Small Biomolecules*, Zhu G (ed) pp 47–67. Berlin, Heidelberg: Springer Berlin Heidelberg

Chen Q, Plasencia M, Li Z, Mukherjee S, Patra D, Chen C-L, Klose T, Yao X-Q, Kossiakoff AA, Chang L, *et al* (2021) Structures of rhodopsin in complex with G-protein-coupled receptor kinase 1. *Nature* 595: 600–605

Cheng J, Sweredoski MJ & Baldi P (2005) Accurate Prediction of Protein Disordered Regions by Mining Protein Structure Data. *Data Min Knowl Discov* 11: 213–222

Ciccarone VC, Polayes DA & Luckow VA (1998) Generation of Recombinant Baculovirus DNA in E.coli Using a Baculovirus Shuttle Vector. *Methods Mol Med* 13: 213–235

Clore GM (2015) Practical Aspects of Paramagnetic Relaxation Enhancement in Biological Macromolecules. *Methods Enzymol* 564: 485–497

Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017a) Structural Characterization of Highly Flexible Proteins by Small-Angle Scattering. *Adv Exp Med Biol* 1009: 107–129

Cordeiro TN, Herranz-Trillo F, Urbanek A, Estaña A, Cortés J, Sibille N & Bernadó P (2017b) Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr Opin Struct Biol* 42: 15–23

Cordier F, Chaffotte A, Terrien E, Préhaud C, Theillet F-X, Delepierre M, Lafon M, Buc H & Wolff N (2012) Ordered Phosphorylation Events in Two Independent Cascades of the PTEN C-tail Revealed by NMR. *J Am Chem Soc* 134: 20533–20543

Craft CM, Whitmore DH & Wiechmann AF (1994) Cone arrestin identified by targeting expression of a functional family. *J Biol Chem* 269: 4613–4619

Crépieux P, Poupon A, Langonné-Gallay N, Reiter E, Delgado J, Schaefer MH, Bourquard T, Serrano L & Kiel C (2017) A Comprehensive View of the β-Arrestinome. *Front Endocrinol* 8: 32

Davenport AP, Holst B, Kleinz M, Maguire JJ & Sivertsen BB (2019) Ghrelin receptor (version 2019.4) in the IUPHAR/BPS Guide to Pharmacology Database. *IUPHARBPS Guide Pharmacol CITE* 2019

Delaforge E, Cordeiro TN, Bernadó P & Sibille N (2018) Conformational Characterization of Intrinsically Disordered Proteins and Its Biological Significance. In *Modern Magnetic Resonance*, Webb GA (ed) pp 381–399. Cham: Springer International Publishing

Di Cera E (2020) Mechanisms of ligand binding. *Biophys Rev* 1: 011303

Dicker F, Quitterer U, Winstel R, Honold K & Lohse MJ (1999) Phosphorylation-independent inhibition of parathyroid hormone receptor signaling by G protein-coupled receptor kinases. *Proc Natl Acad Sci* 96: 5476–5481

Drozdetskiy A, Cole C, Procter J & Barton GJ (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* 43: W389–W394

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, *et al* (2001) Intrinsically disordered protein. *J Mol Graph Model* 19: 26–59

Edeling MA, Mishra SK, Keyel PA, Steinhauser AL, Collins BM, Roth R, Heuser JE, Owen DJ & Traub LM (2006) Molecular switches involving the AP-2 beta2 appendage regulate endocytic cargo selection and clathrin coat assembly. *Dev Cell* 10: 329–342

Fredriksson R (2003) The G-Protein-Coupled Receptors in the Human Genome Form Five Main Families. Phylogenetic Analysis, Paralogon Groups, and Fingerprints. *Mol Pharmacol* 63: 1256–1272

Fuxreiter M, Simon I, Friedrich P & Tompa P (2004) Preformed Structural Elements Feature in Partner Recognition by Intrinsically Unstructured Proteins. *J Mol Biol* 338: 1015–1026

Gaemers S & Bax A (2001) Morphology of Three Lyotropic Liquid Crystalline Biological NMR Media Studied by Translational Diffusion Anisotropy. *J Am Chem Soc* 123: 12343–12352

Gaidarov I, Krupnick JG, Falck JR, Benovic JL & Keen JH (1999) Arrestin function in G protein-coupled receptor endocytosis requires phosphoinositide binding. *EMBO J* 18: 871–881

Geourjon C & Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Bioinformatics* 11: 681–684

Gimenez LE, Kook S, Vishnivetskiy SA, Ahmed MR, Gurevich EV & Gurevich VV (2012) Role of Receptor-attached Phosphates in Binding of Visual and Non-visual Arrestins to G Protein-coupled Receptors. *J Biol Chem* 287: 9028–9040

Granzin J, Stadler A, Cousin A, Schlesinger R & Batra-Safferling R (2015) Structural evidence for the role of polar core residue Arg175 in arrestin activation. *Sci Rep* 5: 15808

Grzesiek S, Bax A, Clore GM, Gronenborn AM, Hu J-S, Kaufman J, Palmer I, Stahl SJ & Wingfield PT (1996) The solution structure of HIV-1 Nef reveals an unexpected fold and permits delineation of the binding surface for the SH3 domain of Hck tyrosine protein kinase. *Nat Struct Mol Biol* 3: 340–345

Guillien M, le Maire A, Mouhand A, Bernadó P, Bourguet W, Banères J-L & Sibille N (2020) IDPs and their complexes in GPCR and nuclear receptor signaling. In *Progress in Molecular Biology and Translational Science* pp 105–155. Elsevier

Gurevich VV, Dion SB, Onorato JJ, Ptasienski J, Kim CM, Sterne-Marr R, Hosey MM & Benovic JL (1995) Arrestin interactions with G protein-coupled receptors. Direct binding studies of wild type and mutant arrestins with rhodopsin, beta 2-adrenergic, and m2 muscarinic cholinergic receptors. *J Biol Chem* 270: 720–731

Gurevich VV & Gurevich EV (2019a) GPCR Signaling Regulation: The Role of GRKs and Arrestins. *Front Pharmacol* 10

Gurevich VV & Gurevich EV (2019b) The structural basis of the arrestin binding to GPCRs. *Mol Cell Endocrinol* 484: 34–41

Han M, Gurevich VV, Vishnivetskiy SA, Sigler PB & Schubert C Crystal Structure of beta-Arrestin at 1.9 A˚: Possible Mechanism of Receptor Binding and Membrane Translocation. 12

Hansen MR, Mueller L & Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* 5: 1065–1074

Hanson J, Paliwal K, Litfin T, Yang Y & Zhou Y (2019a) Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks. *Bioinformatics* 35: 2403–2410

Hanson J, Paliwal K & Zhou Y (2018) Accurate Single-Sequence Prediction of Protein Intrinsic Disorder by an Ensemble of Deep Recurrent and Convolutional Architectures. *J Chem Inf Model* 58: 2369–2376

Hanson K, Paliwal K, Litfin T & Zhou Y (2019b) SPOT-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning - ScienceDirect.

Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB & Gloriam DE (2017) Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* 16: 829–842

He Q-T, Xiao P, Huang S-M, Jia Y-L, Zhu Z-L, Lin J-Y, Yang F, Tao X-N, Zhao R-J, Gao F-Y, *et al* (2021) Structural studies of phosphorylation-dependent interactions between the V2R receptor and arrestin-2. *Nat Commun* 12: 2396

Heffernan R, Yang Y, Paliwal K & Zhou Y (2017) Capturing non-local interactions by long short-term memory bidirectional recurrent neural networks for improving prediction of protein secondary structure, backbone angles, contact numbers and solvent accessibility. *Bioinforma Oxf Engl* 33: 2842–2849

Hendus-Altenburger R, Fernandes CB, Bugge K, Kunze MBA, Boomsma W & Kragelund BB (2019) Random coil chemical shifts for serine, threonine and tyrosine phosphorylation over a broad pH range. *J Biomol NMR* 73: 713–725

Hendus-Altenburger R, Pedraz-Cuesta E, Olesen CW, Papaleo E, Schnell JA, Hopper JTS, Robinson CV, Pedersen SF & Kragelund BB (2016) The human Na+/H+ exchanger 1 is a membrane scaffold protein for extracellular signal-regulated kinase 2. *BMC Biol* 14: 31

Hirsch JA, Schubert C, Gurevich VV & Sigler PB (1999) A Model for Arrestin's Regulation: The 2.8 Å Crystal Structure of Visual Arrestin. *Cell* 97: 257–269

Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V & Skrzypek E (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* 43: D512-520

Huang W, Masureel M, Qu Q, Janetzko J, Inoue A, Kato HE, Robertson MJ, Nguyen KC, Glenn JS, Skiniotis G, *et al* (2020) Structure of the neurotensin receptor 1 in complex with β-arrestin 1. *Nature* 579: 303–308

Hummel JP & Dreyer WJ (1962) Measurement of protein-binding phenomena by gel filtration. *Biochim Biophys Acta* 63: 530–532

Inglese J, Freedman NJ, Koch WJ & Lefkowitz RJ (1993) Structure and mechanism of the G protein-coupled receptor kinases. *J Biol Chem* 268: 23735–23738

Innamorati G, Le Gouill C, Balamotis M & Birnbaumer M (2001) The long and the short cycle. Alternative intracellular routes for trafficking of G-protein-coupled receptors. *J Biol Chem* 276: 13096–13103

Innamorati G, Sadeghi H, Eberle AN & Birnbaumer M (1997) Phosphorylation of the V2 Vasopressin Receptor *. *J Biol Chem* 272: 2486–2492

Innamorati G, Sadeghi HM, Tran NT & Birnbaumer M (1998) A serine cluster prevents recycling of the V2 vasopressin receptor. *Proc Natl Acad Sci U S A* 95: 2222–2226

Ishida T & Kinoshita K (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* 35: W460–W464

Jaakola V-P, Prilusky J, Sussman JL & Goldman A (2005) G protein-coupled receptors show unusual patterns of intrinsic unfolding. *Protein Eng Des Sel* 18: 103–110

Jensen MR, Markwick PRL, Meier S, Griesinger C, Zweckstetter M, Grzesiek S, Bernadó P & Blackledge M (2009) Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure* 17: 1169–1185

Jerabek-Willemsen M, André T, Wanner R, Roth HM, Duhr S, Baaske P & Breitsprecher D (2014) MicroScale Thermophoresis: Interaction analysis and beyond. *J Mol Struct* 1077: 101–113

Jones DT & Cozzetto D (2015) DISOPRED3: precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* 31: 857–863

Julien M, Bouguechtouli C, Alik A, Ghouil R, Zinn-Justin S & Theillet F-X (2020) Multiple Site-Specific Phosphorylation of IDPs Monitored by NMR. In *Intrinsically Disordered Proteins: Methods and Protocols*, Kragelund BB & Skriver K (eds) pp 793–817. New York, NY: Springer US

Kang DS, Kern RC, Puthenveedu MA, von Zastrow M, Williams JC & Benovic JL (2009) Structure of an Arrestin2-Clathrin Complex Reveals a Novel Clathrin Binding Domain That Modulates Receptor Trafficking. *J Biol Chem* 284: 29860–29872

Kang DS, Tian X & Benovic JL (2014) Role of β-arrestins and arrestin domain-containing proteins in G protein-coupled receptor trafficking. *Curr Opin Cell Biol* 0: 63–71

Kang J-H, Jiang Y, Toita R, Oishi J, Kawamura K, Han A, Mori T, Niidome T, Ishida M, Tatematsu K, *et al* (2007) Phosphorylation of Rho-associated kinase (Rho-kinase/ROCK/ROK) substrates by protein kinases A and C. *Biochimie* 89: 39–47

Kang Y, Zhou XE, Gao X, He Y, Liu W, Ishchenko A, Barty A, White TA, Yefanov O, Han GW, *et al* (2015) Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser. *Nature* 523: 561–567

Karplus M & Anderson DH (1959) Valence-Bond Interpretation of Electron-Coupled Nuclear Spin Interactions; Application to Methane. *J Chem Phys* 30: 6–10

Kim J, Ahn S, Ren X-R, Whalen EJ, Reiter E, Wei H & Lefkowitz RJ (2005) Functional antagonism of different G protein-coupled receptor kinases for beta-arrestin-mediated angiotensin II receptor signaling. *Proc Natl Acad Sci U S A* 102: 1442–1447

Kjaergaard M, Brander S & Poulsen FM (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J Biomol NMR* 49: 139–149

Kjaergaard M & Poulsen FM (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *J Biomol NMR* 50: 157–165

Kleckner IR & Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* 1814: 942–968

Kohout TA & Lefkowitz RJ (2003) Regulation of G Protein-Coupled Receptor Kinases and Arrestins During Receptor Desensitization. *Mol Pharmacol* 63: 9–18

Komolov KE, Bhardwaj A & Benovic JL (2015) Atomic Structure of GRK5 Reveals Distinct Structural Features Novel for G Protein-coupled Receptor Kinases. *J Biol Chem* 290: 20629–20647

Komolov KE, Du Y, Duc NM, Betz RM, Rodrigues JPGLM, Leib RD, Patra D, Skiniotis G, Adams CM, Dror RO, *et al* (2017) Structural and Functional Analysis of a β2-Adrenergic Receptor Complex with GRK5. *Cell* 169: 407-421.e16

Kovoor A, Celver J, Abdryashitov RI, Chavkin C & Gurevich VV (1999) Targeted Construction of Phosphorylation-independent β-Arrestin Mutants with Constitutive Activity in Cells. *J Biol Chem* 274: 6831–6834

Krueger KM, Daaka Y, Pitcher JA & Lefkowitz RJ (1997) The Role of Sequestration in G Protein-coupled Receptor Resensitization. *J Biol Chem* 272: 5–8

Krupnick JG, Goodman OB, Keen JH & Benovic JL (1997a) Arrestin/Clathrin Interaction. *J Biol Chem* 272: 15011–15016

Krupnick JG, Gurevich VV & Benovic JL (1997b) Mechanism of quenching of phototransduction. Binding competition between arrestin and transducin for phosphorhodopsin. *J Biol Chem* 272: 18125–18131

Ku H (1966) Notes on the use of propagation of error formulas. *J Res Natl Bur Stand-C Eng Instrum* 70C: 263–273

Kühn H & Dreyer WJ (1972) Light dependent phosphorylation of rhodopsin by ATP. *FEBS Lett* 20: 1–6

Kunapuli P, Onorato JJ, Hosey MM & Benovic JL (1994) Expression, purification, and characterization of the G protein-coupled receptor kinase GRK5. *J Biol Chem* 269: 1099–1105

Lagerström MC & Schiöth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat Rev Drug Discov* 7: 339–357

Landrieu I, Lacosse L, Leroy A, Wieruszeski J-M, Trivelli X, Sillen A, Sibille N, Schwalbe H, Saxena K, Langer T, *et al* (2006) NMR Analysis of a Tau Phosphorylation Pattern. *J Am Chem Soc* 128: 3575–3583

Lee CW, Ferreon JC, Ferreon ACM, Arai M & Wright PE (2010) Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. *Proc Natl Acad Sci U S A* 107: 19290–19295

Lee W, Tonelli M & Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31: 1325–1327

Lee Y, Warne T, Nehmé R, Pandey S, Dwivedi-Agnihotri H, Chaturvedi M, Edwards PC, García-Nafría J, Leslie AGW, Shukla AK, *et al* (2020) Molecular basis of β-arrestin coupling to formoterol-bound β 1 -adrenoceptor. *Nature* 583: 862–866

Leroux AE, Schulze JO & Biondi RM (2018) AGC kinases, mechanisms of regulation and innovative drug development. *Semin Cancer Biol* 48: 1–17

Liggett SB (2011) Phosphorylation barcoding as a mechanism of directing GPCR signaling. *Sci Signal* 4: pe36

Lin FT, Miller WE, Luttrell LM & Lefkowitz RJ (1999) Feedback regulation of beta-arrestin1 function by extracellular signal-regulated kinases. *J Biol Chem* 274: 15971–15974

Lodowski DT, Tesmer VM, Benovic JL & Tesmer JJG (2006) The Structure of G Protein-coupled Receptor Kinase (GRK)-6 Defines a Second Lineage of GRKs. *J Biol Chem* 281: 16785–16793

Lohse MJ, Andexinger S, Pitcher J, Trukawinski S, Codina J, Faure JP, Caron MG & Lefkowitz RJ (1992) Receptor-specific desensitization with purified proteins. Kinase dependence and receptor specificity of beta-arrestin and arrestin in the beta 2-adrenergic receptor and rhodopsin systems. *J Biol Chem* 267: 8558–8564

Loudon RP & Benovic JL (1994) Expression, purification, and characterization of the G protein-coupled receptor kinase GRK6. *J Biol Chem* 269: 22691–22697

Loudon RP & Benovic JL (1997) Altered Activity of Palmitoylation-deficient and Isoprenylated Forms of the G Protein-coupled Receptor Kinase GRK6. *J Biol Chem* 272: 27422–27427

Luckow VA, Lee SC, Barry GF & Olins PO (1993) Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *J Virol* 67: 4566–4579

Luttrell LM, Ferguson SSG, Daaka Y, Miller WE, Maudsley S, Rocca GJD, Lin F-T, Kawakatsu H, Owada K, Luttrell DK, *et al* (1999) β-Arrestin-Dependent Formation of β2 Adrenergic Receptor-Src Protein Kinase Complexes. *Science* 283: 655–661

Luttrell LM, Roudabush FL, Choy EW, Miller WE, Field ME, Pierce KL & Lefkowitz RJ (2001) Activation and targeting of extracellular signal-regulated kinases by β-arrestin scaffolds. *Proc Natl Acad Sci* 98: 2449–2454

Manning M, Stoev S, Chini B, Durroux T, Mouillac B & Guillon G (2008) Peptide and non-peptide agonists and antagonists for the vasopressin and oxytocin V1a, V1b, V2 and OT receptors: research tools and potential therapeutic agents. *Prog Brain Res* 170

Min K, Yoon H-J, Park JY, Baidya M, Dwivedi-Agnihotri H, Maharana J, Chaturvedi M, Chung KY, Shukla AK & Lee HH (2020) Crystal Structure of β-Arrestin 2 in Complex with CXCR7 Phosphopeptide. *Structure*

Minton AP (2016) Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. *Anal Biochem* 501: 4–22

Moerke NJ (2009) Fluorescence Polarization (FP) Assays for Monitoring Peptide-Protein or Nucleic Acid-Protein Binding. *Curr Protoc Chem Biol* 1: 1–15

Moffett S, Rousseau G, Lagacé M & Bouvier M (2001) The palmitoylation state of the beta(2)-adrenergic receptor regulates the synergistic action of cyclic AMP-dependent protein kinase and beta-adrenergic receptor kinase involved in its phosphorylation and desensitization. *J Neurochem* 76: 269–279

Moore CAC, Milano SK & Benovic JL (2007) Regulation of Receptor Trafficking by GRKs and Arrestins. *Annu Rev Physiol* 69: 451–482

Murakami A, Yajima T, Sakuma H, McLaren MJ & Inana G (1993) X-arrestin: a new retinal arrestin mapping to the X chromosome. *FEBS Lett* 334: 203–209

Nguyen AH, Thomsen ARB, Cahill TJ, Huang R, Huang L-Y, Marcink T, Clarke OB, Heissel S, Masoudi A, Ben-Hail D, *et al* (2019) Structure of an endosomal signaling GPCR–G protein–β-arrestin megacomplex. *Nat Struct Mol Biol*: 1–9

Nielsen JT & Mulder FAA (2018) POTENCI: prediction of temperature, neighbor and pH-corrected chemical shifts for intrinsically disordered proteins. *J Biomol NMR* 70: 141–165

Nobles KN, Guan Z, Xiao K, Oas TG & Lefkowitz RJ (2007) The active conformation of beta-arrestin1: direct evidence for the phosphate sensor in the N-domain and conformational differences in the active states of beta-arrestins1 and -2. *J Biol Chem* 282: 21370–21381

Nobles KN, Xiao K, Ahn S, Shukla AK, Lam CM, Rajagopal S, Strachan RT, Huang T-Y, Bressler EA, Hara MR, *et al* (2011) Distinct Phosphorylation Sites on the 2-Adrenergic Receptor Establish a Barcode That Encodes Differential Functions of -Arrestin. *Sci Signal* 4: ra51–ra51

Oakley RH, Laporte SA, Holt JA, Caron MG & Barak LS (2000) Differential Affinities of Visual Arrestin, βArrestin1, and βArrestin2 for G Protein-coupled Receptors Delineate Two Major Classes of Receptors. *J Biol Chem* 275: 17201–17210

O'Dowd BF, Hnatowich M, Caron MG, Lefkowitz RJ & Bouvier M (1989) Palmitoylation of the human β2-adrenergic receptor. *J Biol Chem* 264: 7564–7569

Oh K-I, Lee K-K, Park E-K, Jung Y, Hwang G-S & Cho M (2012) A comprehensive library of blocked dipeptides reveals intrinsic backbone conformational propensities of unfolded proteins. *Proteins* 80: 977–990

Onorato JJ, Palczewski K, Regan JW, Caron MG, Lefkowitz RJ & Benovic JL (1991) Role of acidic amino acids in peptide substrates of the .beta.-adrenergic receptor kinase and rhodopsin kinase. *Biochemistry* 30: 5118–5125

Ottiger M, Delaglio F & Bax A (1998) Measurement ofJand Dipolar Couplings from Simplified Two-Dimensional NMR Spectra. *J Magn Reson* 131: 373–378

Ozenne V, Bauer F, Salmon L, Huang J -r., Jensen MR, Segard S, Bernado P, Charavay C & Blackledge M (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics* 28: 1463–1470

Palczewski K, Buczyłko J, Kaplan MW, Polans AS & Crabb JW (1991) Mechanism of rhodopsin kinase activation. *J Biol Chem* 266: 12949–12955

Palczewski K, McDowell JH & Hargrave PA (1988) Rhodopsin kinase: substrate specificity and factors that influence activity. *Biochemistry* 27: 2306–2313

Phizicky EM & Fields S (1995) Protein-protein interactions: methods for detection and analysis. *Microbiol Rev* 59: 94–123

Pitcher JA, Payne ES, Csortos C, DePaoli-Roach AA & Lefkowitz RJ (1995) The G-protein-coupled receptor phosphatase: a protein phosphatase type 2A with a distinct subcellular distribution and substrate specificity. *Proc Natl Acad Sci U S A* 92: 8343–8347

Pitcher JA, Tesmer JJ, Freeman JL, Capel WD, Stone WC & Lefkowitz RJ (1999) Feedback inhibition of G protein-coupled receptor kinase 2 (GRK2) activity by extracellular signal-regulated kinases. *J Biol Chem* 274: 34531–34534

Premont RT, Macrae AD, Aparicio SAJR, Kendall HE, Welch JE & Lefkowitz RJ (1999) The GRK4 Subfamily of G Protein-coupled Receptor Kinases ALTERNATIVE SPLICING, GENE ORGANIZATION, AND SEQUENCE CONSERVATION. *J Biol Chem* 274: 29381–29389

Premont RT, Macrae AD, Stoffel RH, Chung N, Pitcher JA, Ambrose C, Inglese J, MacDonald ME & Lefkowitz RJ (1996) Characterization of the G Protein-coupled Receptor Kinase GRK4 IDENTIFICATION OF FOUR SPLICE VARIANTS. *J Biol Chem* 271: 6403–6410

Qanbar R & Bouvier M (2003) Role of palmitoylation/depalmitoylation reactions in G-protein-coupled receptor function. *Pharmacol Ther* 97: 1–33

Ramakrishnan C & Ramachandran GN (1965) Stereochemical Criteria for Polypeptide and Protein Chain Conformations. *Biophys J* 5: 909–933

Ren X-R, Reiter E, Ahn S, Kim J, Chen W & Lefkowitz RJ (2005) Different G protein-coupled receptor kinases govern G protein and beta-arrestin-mediated signaling of V2 vasopressin receptor. *Proc Natl Acad Sci U S A* 102: 1448–1453

Robinson MJ & Cobb MH (1997) Mitogen-activated protein kinase pathways. *Curr Opin Cell Biol* 9: 180–186

Romero P, Obradovic Z, Li X, Garner EC, Brown CJ & Dunker AK (2001) Sequence complexity of disordered protein. *Proteins* 42: 38–48

Rückert M & Otting G (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J Am Chem Soc* 122: 7793–7797

Sadeghi HM, Innamorati G, Dagarag M & Birnbaumer M (1997) Palmitoylation of the V2 Vasopressin Receptor. *Mol Pharmacol* 52: 21–29

Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M & Blackledge M (2010) NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *J Am Chem Soc* 132: 8407–8418

Schöneberg T, Schulz A, Biebermann H, Hermsdorf T, Römpler H & Sangkuhl K (2004) Mutant G-protein-coupled receptors as a cause of human diseases. *Pharmacol Ther* 104: 173–206

Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig K-P & Kalbitzer HR (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J Biomol NMR* 39: 275–289

Schwarzinger S, Kroon GJA, Foss TR, Chung J, Wright PE & Dyson HJ (2001) Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *J Am Chem Soc* 123: 2970–2978

Senicourt L, Maire A le, Allemand F, Carvalho JE, Guee L, Germain P, Schubert M, Bernadó P, Bourguet W & Sibille N (2020) Structural insights into the cooperative interaction of the intrinsically disordered co-activator TIF2 with retinoic acid receptor heterodimer (RXR/RAR). *bioRxiv*: 2020.10.28.359091

Seyedabadi M, Gharghabi M, Gurevich EV & Gurevich VV (2021) Receptor-Arrestin Interactions: The GPCR Perspective. *Biomolecules* 11: 218

Shen Y, Roche J, Grishaev A & Bax A (2018) Prediction of nearest neighbor effects on backbone torsion angles and NMR scalar coupling constants in disordered proteins. *Protein Sci Publ Protein Soc* 27: 146–158

Shilton BH, McDowell JH, Smith WC & Hargrave PA (2002) The solution structure and activation of visual arrestin studied by small-angle X-ray scattering. *Eur J Biochem* 269: 3801–3809

Shukla AK, Manglik A, Kruse AC, Xiao K, Reis RI, Tseng W-C, Staus DP, Hilger D, Uysal S, Huang L-Y, *et al* (2013) Structure of active β-arrestin1 bound to a G protein-coupled receptor phosphopeptide. *Nature* 497: 137–141

Shukla AK, Westfield GH, Xiao K, Reis RI, Huang L-Y, Tripathi-Shukla P, Qian J, Li S, Blanc A, Oleskie AN, *et al* (2014) Visualization of arrestin recruitment by a G Protein-Coupled Receptor. *Nature* 512: 218–222

Sibille N & Bernadó P (2012) Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem Soc Trans* 40: 955–962

Singh P, Wang B, Maeda T, Palczewski K & Tesmer JJG (2008) Structures of Rhodopsin Kinase in Different Ligand States Reveal Key Elements Involved in G Protein-coupled Receptor Kinase Activation. *J Biol Chem* 283: 14053–14062

Smith GE, Fraser MJ & Summers MD (1983) Molecular Engineering of the Autographa californica Nuclear Polyhedrosis Virus Genome: Deletion Mutations Within the Polyhedrin Gene. *J Virol* 46: 584–593

Smith JS & Rajagopal S (2016) The β-Arrestins: Multifunctional Regulators of G Protein-coupled Receptors. *J Biol Chem* 291: 8969–8977

Sohlemann P, Hekman M, Buchen C, Elce JS & Lohse MJ (1993) Purification and functional characterization of β-adrenergic receptor kinase expressed in insect cells. *FEBS Lett* 324: 59–62

Staus DP, Hu H, Robertson MJ, Kleinhenz ALW, Wingler LM, Capel WD, Latorraca NR, Lefkowitz RJ & Skiniotis G (2020) Structure of the M2 muscarinic receptor–β-arrestin complex in a lipid nanodisc. *Nature* 579: 297–302

Sterne-Marr R, Baillargeon AI, Michalski KR & Tesmer JJG (2013) Expression, Purification, and Analysis of G-Protein-Coupled Receptor Kinases. *Methods Enzymol* 521: 347–366

Szczepek M, Beyrière F, Hofmann KP, Elgeti M, Kazmin R, Rose A, Bartl FJ, Stetten D von, Heck M, Sommer ME, *et al* (2014) Crystal structure of a common GPCR-binding interface for G protein and arrestin. *Nat Commun* 5: 1–8

Tamiola K, Acar B & Mulder FAA (2010) Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *J Am Chem Soc* 132: 18000–18003

Teilum K, Kragelund BB & Poulsen FM (2002) Transient structure formation in unfolded acyl-coenzyme A-binding protein observed by site-directed spin labelling. *J Mol Biol* 324: 349–357

Thomsen ARB, Plouffe B, Cahill TJ, Shukla AK, Tarrasch JT, Dosey AM, Kahsai AW, Strachan RT, Pani B, Mahoney JP, *et al* (2016) GPCR-G Protein-β-Arrestin Super-Complex Mediates Sustained G Protein Signaling. *Cell* 166: 907–919

Tobin AB, Butcher AJ & Kong KC (2008) Location, location, location...site-specific GPCR phosphorylation offers a mechanism for cell-type-specific signalling. *Trends Pharmacol Sci* 29: 413–420

Tompa P, Schad E, Tantos A & Kalmar L (2015) Intrinsically disordered proteins: emerging interaction specialists. *Curr Opin Struct Biol* 35: 49–59

Tovo-Rodrigues L, Roux A, Hutz MH, Rohde LA & Woods AS (2014) Functional characterization of G-protein-coupled receptors: A bioinformatics approach. *Neuroscience* 277: 764–779

Tran TM, Friedman J, Qunaibi E, Baameur F, Moore RH & Clark RB (2004) Characterization of agonist stimulation of cAMP-dependent protein kinase and G protein-coupled receptor kinase phosphorylation of the beta2-adrenergic receptor using phosphoserine-specific antibodies. *Mol Pharmacol* 65: 196–206

Uversky VN (2002) What does it mean to be natively unfolded? *Eur J Biochem* 269: 2–12

Uversky VN (2012) Size-Exclusion Chromatography in Structural Analysis of Intrinsically Disordered Proteins. In *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools*, Uversky VN & Dunker AK (eds) pp 179–194. New York, NY: Springer

Uversky VN (2015) Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J* 282: 1182–1189

Uversky VN, Gillespie JR & Fink AL (2000) Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* 41: 415–427

Van Roey K, Gibson TJ & Davey NE (2012) Motif switches: decision-making in cell regulation. *Curr Opin Struct Biol* 22: 378–385

Venkatakrishnan A, Flock T, Prado DE, Oates ME, Gough J & Madan Babu M (2014) Structured and disordered facets of the GPCR fold. *Curr Opin Struct Biol* 27: 129–137

Violin JD, Ren X-R & Lefkowitz RJ (2006) G-protein-coupled receptor kinase specificity for beta-arrestin recruitment to the beta2-adrenergic receptor revealed by fluorescence resonance energy transfer. *J Biol Chem* 281: 20577–20588

Vuister GW & Bax A (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins. *J Am Chem Soc* 115: 7772–7777

Walsh I, Martin AJM, Di Domenico T & Tosatto SCE (2012) ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics* 28: 503–509

Weiss ER, Raman D, Shirakawa S, Ducceschi MH, Bertram PT, Kraft TW & Osawa S The Cloning of GRK7, a Candidate Cone Opsin Kinase, from Cone- and Rod-Dominant Mammalian Retinas. *Mol Vis*: 8

Wen J, Arakawa T & Philo JS (1996) Size-Exclusion Chromatography with On-Line Light-Scattering, Absorbance, and Refractive Index Detectors for Studying Proteins and Their Interactions. *Anal Biochem* 240: 155–166

Wilden U, Hall SW & Kühn H (1986) Phosphodiesterase activation by photoexcited rhodopsin is quenched when rhodopsin is phosphorylated and binds the intrinsic 48-kDa protein of rod outer segments. *Proc Natl Acad Sci U S A* 83: 1174–1178

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc* 73: 1–16

Wishart DavidS & Sykes BrianD (1994a) The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR* 4

Wishart DS, Bigam CG, Holm A, Hodges RS & Sykes BD (1995) 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* 5: 67–81

Wishart DS & Sykes BD (1994b) [12] Chemical shifts as a tool for structure determination. In *Methods in Enzymology* pp 363–392. Elsevier

Wisler JW, Xiao K, Thomsen ARB & Lefkowitz RJ (2014) Recent developments in biased agonism. *Curr Opin Cell Biol* 0: 18–24

Woody RW (1996) Theory of Circular Dichroism of Proteins. In *Circular Dichroism and the Conformational Analysis of Biomolecules*, Fasman GD (ed) pp 25–67. Boston, MA: Springer US

Woody RW (2010) Circular Dichroism of Intrinsically Disordered Proteins. In *Instrumental Analysis of Intrinsically Disordered Proteins* pp 303–321. John Wiley & Sons, Ltd

Wright PE & Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 16: 18–29

Xiao K, Shenoy SK, Nobles K & Lefkowitz RJ (2004) Activation-dependent Conformational Changes in β-Arrestin 2. *J Biol Chem* 279: 55744–55753

Xue B, Dunbrack RL, Williams RW, Dunker AK & Uversky VN (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta BBA - Proteins Proteomics* 1804: 996–1010

Yan R, Xu D, Yang J, Walker S & Zhang Y (2013) A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. *Sci Rep* 3: 2619

Yang F, Yu X, Liu C, Qu C-X, Gong Z, Liu H-D, Li F-H, Wang H-M, He D-F, Yi F, *et al* (2015) Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and 19 F-NMR. *Nat Commun* 6: 1–15

von Zastrow M & Kobilka BK (1992) Ligand-regulated internalization and recycling of human beta 2-adrenergic receptors between the plasma membrane and endosomes containing transferrin receptors. *J Biol Chem* 267: 3530–3538

Zhang J, Barak LS, Anborgh PH, Laporte SA, Caron MG & Ferguson SSG (1999) Cellular Trafficking of G Protein-coupled Receptor/β-Arrestin Endocytic Complexes. *J Biol Chem* 274: 10999–11006

Zhou J, Zhao S & Dunker AK (2018) Intrinsically Disordered Proteins Link Alternative Splicing and Post-translational Modifications to Complex Cell Signaling and Regulation. *J Mol Biol* 430: 2342–2359

Zhou XE, He Y, de Waal PW, Gao X, Kang Y, Van Eps N, Yin Y, Pal K, Goswami D, White TA, *et al* (2017) Identification of Phosphorylation Codes for Arrestin Recruitment by G Protein-Coupled Receptors. *Cell* 170: 457-469.e13