Aix Marseille Université

NNT : 2024AIXM0502

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
Le 11 Décembre 2024 par

Pierre GAUFFRE

Études structurales des enzymes du Complexe Réplication /Transcription (RTC) des Coronavirus et des virus à long génome d'ARN

Discipline
Biologie Santé

Spécialité
Biochimie structurale

École doctorale
ED62 – Science de la Vie et de la Santé

Laboratoire
Architecture et fonction des Macromolécules Biologiques
UMR 7257 CNRS
Aix-Marseille Université

Composition du jury

Dr. Sandrine BELOUZARD | Rapporteuse
CR | CIIL, Campus de l'Institut Pasteur de Lille , CNRS UMR 9017, Inserm U1019, Lille, France

Dr. Cédric LEYRAT | Rapporteur
CR | IGS, Université de Montpellier, INSERM U1191, CNRS UMR 5203, Montpellier, France

Dr. Hélène MALET | Examinatrice
MCF | IBS, Université Grenoble Alpes, CNRS UMR 5075, Grenoble, France

Dr. Bruno COUTARD | Examinateur
PR | UVE, Aix-Marseille Université, IRD 190, INSERM 1207, IHU Méditerranée Infection, Marseille, France

Dr. Patrice GOUET | Examinateur
PR | IBCP, Institut de Biologie et Chimie des Protéines, UAR 3760, UCBL, Université Claude Bernard Lyon 1, Lyon France

Dr. Jean-Claude GUILLEMOT | Président du jury
PR | Polytech Marseille, Aix-Marseille Université, Marseille, France

Dr. François FERRON | Directeur de thèse
CR | AFMB, Aix-Marseille Université, CNRS UMR7257, Marseille, France

Dr. Bruno CANARD | Co-directeur de thèse
DR | AFMB, Aix-Marseille Université, CNRS UMR7257, Marseille, France

cnrs

AMU PhD program Defence and Homeland Security

AFMB ARCHITECTURE ET FONCTION DES MACROMOLÉCULES BIOLOGIQUES

Affidavit

Je soussigné, Pierre GAUFFRE, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de François FERRON et Bruno CANARD, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille, le 15 octobre 2024

Signature : Gauffre Pierre

Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et/ou brevets et participation aux conférences

1) Liste des publications et/ou brevet [a] réalisées dans le cadre du projet de thèse :

    1. **Gauffre P**, Gaubert A, Canard B, Ferron F. “Structural studies of NSP14 SARS-CoV-2 in presence of nanobody.” article en préparation
    2. Feracci M, Eydoux C, Fattorini V, Lo Bello L, **Gauffre P**, Selisko B, Sutto-Ortiz P, Shannon A, Xia H, Shi PY, Noel M, Debart F, Vasseur JJ, Good S, Lin K, Moussa A, Sommadossi JP, Chazot A, Alvarez K, Guillemot JC, Decroly E, Ferron F, Canard B. “AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5.” Antiviral Res. 2023 Apr;212:105574. DOI : 10.1016/j.antiviral.2023.105574
    3. Shannon A*, Sama B*, **Gauffre P**,Guez T, Debart F, Vasseur JJ, Decroly E, Canard B, Ferron F. “A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome” Nucleic Acids Research, Volume 50, Issue 19, 28 October 2022, Pages 11186–11198, DOI : 10.1093/nar/gkac876
    4. Shannon, A., Fattorini, V., Sama, B., Selisko, B., Feracci, M., Falcou, C., **Gauffre**, P., El Kazzi, P., Delpal, A., Decroly, E., Alvarez, K., Eydoux, C., Guillemot, J, C., Moussa, A., S, Good, S., La Colla, P., Lin, K., Sommadossi, J, P., Zhu, Y., Yan, X., Shi, H., Ferron, F., Canrad, B. “A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase.” Nat Commun 13, 621 (2022). DOI : 10.1038/s41467-022-28113-1

[a] Cette liste comprend les articles publiés, les articles soumis à publication et les articles en préparation ainsi que les livres, chapitres de livre et/ou toutes formes de valorisation des résultats des travaux propres à la discipline du projet de thèse. La référence aux publications doit suivre les règles standards de bibliographie et doit être conforme à la charte des publications d’AMU.

PDB	Titre	Auteur
9H01	nsp14 of SARS-CoV-2 in complex with a camelid nanobody	**Gauffre,P**., Canard,B, Ferron,F.
7Z2J	White Bream virus N7-Methyltransferase	Shannon,A., **Gauffre,P**., Canard,B,Ferron,F.
7Z05	White Bream virus N7-Methyltransferase	Shannon, A., **Gauffre,P**.,Canard,B,Ferron,F.
8BCR	DENV3 Methyltransferase in complexed with AT-9010 and SAH.	**Gauffre,P**.,Fattorini,V.,Canard, B.,Ferron,F.
7ED5	A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase	Shannon,A.,Fattorini,V.,Sama,B.,Selisko,B., Feracci, M., Falcou, C., **Gauffre, P**., El Kazzi,P.,Delpal,A., Decroly,E., Alvarez,K., Eydoux,C., Guillemot,JC., Moussa,A., Good,S., Colla,P.Lin,K., Sommadossi,J-P., Zhu,Y.X.,Yan, X.D., Shi,H., Ferron,F., Canard,B.

SASBD	Titre	Auteur
SASDVU8/azkeey3eio	nsp14 from SARS-CoV-2	**Gauffre,P**., Canard,B, Ferron,F.
SASDVV8/louutuw0ub	nsp14 and nsp10 from SARS-CoV-2	
SASDVW8/qdzcfurhp0	nsp14 and nsp10 from SARS-CoV-2 with vhh	
SASDVX8/oe90k4khlf	nsp14 from SARS-CoV-2 with vhh	

2) Participation aux conférences[b] et écoles d'été au cours de la période de thèse :

Présentation Oral

34th European Crystallographic Meeting (ECM) du 26 au 30 Aout 2024 à Padoue, Italie - **Gauffre P**, Gaubert A, Canard B, Ferron F "Crystallographic structure of whole nsp14 from SARS-CoV-2 stabilised by a nanobody" DOI : 10.1107/S2053273324098929

Posters

1. 34th European Crystallographic Meeting (ECM) du 26 au 30 Aout 2024 à Padoue, Italie - **Gauffre P**, Gaubert A, Canard B, Ferron F "Crystallographic structure of whole nsp14 from SARS-CoV-2 stabilised by a nanobody" HAL : hal.science/hal-04729389
2. 26e Journées Francophones de Virologie – 10 & 18 avril 2024 – Bruxelles, ULB - **Gauffre P**, Gaubert A, Canard B, Ferron F "Structure cristallographique de nsp14 entière du SARS-CoV-2 stabilisé par un nanobodie" XXVIes Journées francophones de virologie. DOI : 10.1684/vir.2024.1048
3. Biologie Structurale Intégrative (BSI) – 13 au 17 novembre 2023 – Marseille - **Gauffre P**, Gaubert A, Mate-Perez M, Canard B, Ferron F "Crystallographic structure of nsp14 full length from SARS-CoV-2, stabilised by a nanobody" HAL : hal.science/hal-04246283
4. 25e Journées Francophones de Virologie – 17 & 18 avril 2023 – Paris, Institut Pasteur - **Gauffre P**, Gaubert A, Mate-Perez M, Canard B, Ferron F "Le complexe nsp14-nsp10 du SARS-CoV-2 stabilisé par un nanobody" XXVes Journées francophones de virologie. P103 Virologie. 2023;27(2):48-9. DOI : 0.1684/vir.2023.993
5. 23e Journées Francophones de Virologie – 26 & 27 avril 2021 – Paris - **Gauffre P**, Shannon A, Sama B, Bruno Canard B, Ferron F. "Analyse structurale de la nouvelle N7-guanine-MTase du virus de la Brème bordelière" XXIIIes Journées francophones de virologie. P101 Virologie. 2021;25(1):68-74. DOI : 10.1684/vir.2021.0892

[b] Le terme « conférence » est générique. Il désigne à la fois « conférence », « congrès », « workshop », « colloques », « rencontres nationales et/ou internationales » … etc.
Indiquer si vous avez fait une présentation orale ou sous forme de poster.

Résumé

Mon travail de thèse de doctorat se concentre sur l'étude structurale des enzymes (nsp7, nsp8, et nsp12) impliquées dans le complexe de réplication et transcription (RTC) des Coronavirus, notamment SARS-CoV, SARS-CoV-2 et MERS-CoV, ainsi que sur d'autres virus à long génome d'ARN. Le RTC est essentiel pour la transcription des ARNm viraux et la réplication du génome, ce qui en fait une cible prioritaire pour le développement d'antiviraux.

Le premier chapitre détaille les méthodes et stratégies d'optimisation de la production et de la purification des enzymes du RTC pour chacun des trois virus, en mettant un accent particulier sur le MERS-CoV. Ces travaux ont permis de reconstituer *in vitro* ce complexe et de développer une caractérisation précise des mécanismes structuraux et fonctionnels des RTC de ces virus. Ils ont notamment contribué à la caractérisation d'un analogue nucléotidique, l'AT-9010 qui cible la nsp12 du SARS-CoV-2.

Le deuxième chapitre aborde la caractérisation structurale par cristallographie aux rayons X de trois méthyltransférases issues de différents virus. La première étude porte sur la résolution d'un complexe entre la méthyltransférase NS5 de la dengue et l'AT-9010, afin de caractériser ce potentiel inhibiteur. La seconde concerne l'étude structurale de la première N7-méthyltransférase virale d’un virus à ARN+ identifiée dans un locus inhabituel du génome d'un Nidovirus . La troisième caractérisation examine la nsp14 de SARS-CoV-2 en complexe avec un anticorps à domaine unique, visant à stabiliser cette enzyme flexible. NSP14 possède une activité N7-guanine méthyltransferase non canonique ainsi qu'un domaine exoribonucléase. Cette enzyme joue un rôle clé dans la correction des erreurs de réplication et la protection des ARN viraux, représentant ainsi une cible stratégique pour le développement d'inhibiteurs. Les résultats cristallographiques sont complétés par diverses analyses biophysiques, enzymatiques et structurales, notamment par des techniques de SEC-SAXS.

En conclusion, mon travail de thèse vise à approfondir notre compréhension des mécanismes d'action des enzymes du RTC des Coronavirus, et a permis de mettre en évidence de nouveaux modes d'action d'un inhibiteur potentiel contre ces différents virus.

Mes résultats ouvrent des perspectives prometteuses pour le développement de thérapies ciblées contre ces virus, contribuant potentiellement à la lutte contre de futures pandémies.

Mots clés : Réplication virale, nsp12, RTC, exonucléase, méthyltransférase, nsp14, Coronavirus, Nidovirus , SARS-CoV, MERS, SARS-CoV-2.

Abstract

My doctoral research focuses on the structural study of enzymes (nsp7, nsp8, and nsp12) involved in the replication and transcription complex (RTC) of Coronaviruses, including SARS-CoV, SARS-CoV-2, and MERS-CoV, as well as other long RNA genome viruses. The RTC is essential for the transcription of viral mRNAs and genome replication, making it a priority target for antiviral development.

The first chapter details the methods and strategies for optimizing the production and purification of RTC enzymes for each of the three viruses, with particular emphasis on MERS-CoV. This work enabled the *in vitro* reconstitution of the complex and the precise characterization of the structural and functional mechanisms of the RTCs of these viruses. Notably, it contributed to the characterization of a nucleotide analog, AT-9010, which targets nsp12 of SARS-CoV-2.

The second chapter addresses the structural characterization by X-ray crystallography of three methyltransferases from different viruses. The first study involves resolving the complex between the dengue virus NS5 methyltransferase and AT-9010 to characterize this potential inhibitor. The second study pertains to the structural analysis of the first viral N7-methyltransferase from a positive-sense RNA virus identified in an unusual locus within a Nidovirus genome. The third characterization examines SARS-CoV-2 nsp14 in complex with a single-domain antibody aimed at stabilizing this flexible enzyme. NSP14 possesses a non-canonical N7-guanine methyltransferase domain as well as an exoribonuclease domain. This enzyme plays a key role in replication error correction and protection of viral RNA, thus representing a strategic target for inhibitor development. The crystallographic results are complemented by various biophysical, enzymatic, and structural analyses, including SEC-SAXS techniques.

In conclusion, my doctoral research aims to deepen our understanding of the mechanisms of action of RTC enzymes in Coronaviruses and has highlighted new modes of action of a potential inhibitor against these various viruses.

My results open promising prospects for the development of targeted therapies against these viruses, potentially contributing to the fight against future pandemics.

Keywords : Viral replication, nsp12, RTC, exonuclease, methyltransferase, nsp14, Coronavirus, Nidovirus , SARS-CoV, MERS, SARS-CoV-2.

Remerciements

Je tiens à remercier tout d'abord les deux directeurs successifs de notre laboratoire, **Yves Bourne** et **Juan Reguera**, pour m'avoir accueilli dans ce magnifique laboratoire qu'est le laboratoire d'Architecture et Fonction des Macromolécules Biologiques (**AFMB**), où j'ai réellement pu m'épanouir au cours de ces quatre années de thèse.

Je tiens sincèrement à remercier mes encadrants de thèse, **François Ferron** et **Bruno Canard**, pour avoir cru en moi et pour m'avoir permis de réaliser mes travaux de recherche sur un projet quasiment sur mesure et qui me tenait à cœur. Grâce à eux, j'ai pu réaliser cette thèse dans d'excellentes conditions. Ils ont été présents quand j'en avais besoin, que ce soit pour m'aider dans mes réflexions, me recentrer quand c'était nécessaire et m'encourager également.

Je souhaite également remercier les personnes avec qui j'ai pu collaborer sur les différents projets, tant en microscopie électronique avec **Bruno Klaholz**, **Clara Ottilie Loeffelholz Von Colberg** et **Charles Barchet**, qu'en cristallographie avec **Mikael Feracci**, **Véronique Fattorini**, **Adrien Delpal**, **Ashleigh Shannon**, **François Ferron** et **Nicolas Papageorgiou** (dit Nikos).

Je me suis découvert une réelle passion pour la biologie structurale et notamment pour la cristallographie. C'est notamment grâce à vous, **Véronique Zamboni**, **Claire Zimberger** et **Mikael Feracci**, grâce à vos petits conseils lors de la préparation de mes échantillons et des essais de cristallogenèse que j'ai réalisés. Merci également à **François Ferron** et **Nicolas Papageorgiou** pour votre formation « à la dure » sur la théorie de la cristallographie et sur la résolution des structures cristallographiques. Un grand merci à **Gerlind Sulzenbacher** et **Marc Zamboni** pour votre aide et votre bienveillance au quotidien lors de la préparation des dizaines de missions que j'ai pu réaliser.

Je voudrais également remercier les différentes personnes que j'ai pu croiser au synchrotron Soleil et à l'ESRF, qui ont pris de leur temps pour partager leurs expériences sur ces fabuleux outils, certains de leurs secrets de cristallographe, et pour les discussions passionnantes sur la science et leurs métiers. Je pense notamment à **David Flot**, **Pierre Legrand**, **Andrew Thompson** et **Pierre Montaville**.

Je voudrais aussi remercier **Barbara Selisko** et **Adrien Delpal** pour leur grande aide (et leur patience) sur la mise en place des différents tests enzymatiques que j'ai pu réaliser au cours de ma thèse.

Je voudrais remercier **« le meilleur bureau »** avec qui j'ai vraiment passé d'excellents moments, de l'envoi d'un papier à la soumission digne d'un décollage de fusée, en passant par les « quarts d'heure de folie » pour décompresser de la journée, et tout un tas de discussions et débats très sérieux en science ou sur la vie.

Plus largement, je remercie sincèrement chaque membre de l'équipe BC pour tous ces merveilleux moments passés ensemble. Bien évidemment, je remercie les deux autres camarades toulousains, **Adrien** et **Camille**. Ce fut un vrai plaisir de partager cette heureuse aventure avec vous depuis notre arrivée, de nos galères d'appartement en plein confinement à nos réussites mutuelles. Nous avons grandi scientifiquement ensemble, nous nous sommes soutenus, et avons appris les uns des autres sans pudeur et avec bienveillance, et ce fut un plaisir.

Adrien, je suis heureux d'avoir pu partager tous ces moments avec toi, notamment sur le projet nsp14 où nous avons pu mutuellement nous apporter notre expertise, ainsi que sur les choix stratégiques à la cantine du synchrotron.

Camille, je me souviendrai toujours de ces litres de nsp12 que nous avons pu purifier ensemble dès notre arrivée. Nous nous sommes ensuite spécialisés chacun sur une partie du projet avant que tu prennes la meilleure décision de ta vie, celle de ne pas faire de thèse. On en a souvent rigolé de ce point-là. Cela n'empêche pas de s'être vus évoluer, j'ai vu la fierté dans tes yeux à chaque fois que tu réussissais, et je suis content d'avoir pu partager ces moments avec toi et de nous être soutenus tout au long de nos parcours respectifs. Merci pour tout cela.

Ashleigh, je ne suis pas sûr que tu aies conscience de toute l'aide et l'inspiration que tu as pu m'apporter au cours des années. Je t'ai vue évoluer, d'abord en tant que jeune post-doc, nous avons travaillé ensemble sans compter sur différents projets. Je t'ai vue grandir au fur et à mesure, tout en restant toujours présente pour répondre à mes questionnements scientifiques, en trouvant toujours les mots justes, même en français. J'ai grandement apprécié ce soutien mutuel, et je t'en serai toujours reconnaissant. Maintenant que tu es Chargée de Recherche, venant de décrocher ton ERC, la suite sera radieuse et je sais que tu encadreras très bien les prochaines personnes avec qui tu travailleras. Continue à faire de belles choses.

François, nous nous sommes découverts au tout début autour des ÄKTA, puis tu m'as fait découvrir la cristallographie, la beauté de cette discipline et toutes les subtilités qu'elle demande, comme ses besoins de « quelques grains de poussière pour que la magie opère ». En plus de découvrir une discipline qui fait aujourd'hui partie de moi, j'ai aussi découvert un encadrant de thèse avenant, avec beaucoup d'humour, et réellement passionné par ce qu'il fait. Cela nous a amenés à de nombreuses discussions sur différentes thématiques, en les abordant sous différents angles. En plus d'être admiratif, cela a aussi nourri ma curiosité et m'a appris à cultiver ma rigueur scientifique, et je t'en remercie.

Alice, merci pour ce que tu es. Tu respires la joie de vivre et la passion pour la science. Cela passe notamment par les fameuses purifications du vendredi, alors que moi j'étais plus adepte de celles du lundi. Malgré cette différence presque notable de point de vue, c'était toujours un plaisir de partager nos autres points de vue lors d'échanges avec toi, ou plus simplement de simples conseils en couture, travaux manuels (j'ai toujours ton porte-clés) et sur la vie. Merci.

Vincent, j'ai adoré toutes nos discussions de geek et linuxien que nous avons pu avoir ensemble et avec François. Je dois avouer que j'ai parfois été un peu dépassé quand tu m'expliquais, par exemple, ton choix de type de base de données pour les logiciels que je créais, mais cela nous permettait d'échanger respectivement sur nos points de vue avec bienveillance, ce qui était toujours plaisant et instructif.

Barbara, tu es un vrai repère dans notre laboratoire. Toujours de bons conseils, tant au niveau scientifique qu'humain. Ta bienveillance et ta bonne humeur sont importantes dans le labo, et c'est un vrai plaisir de travailler avec toi.

Oney, le latin lover. On en a appris des choses avec toi. Des éclats de rire aux « ehh non mais... ». On a souvent eu des points de vue différents sur comment aborder la vie, mais ce fut toujours un plaisir de partager nos points de vue et tous ces moments passés ensemble.

Bhawna, I still sometimes hear your laughter when Oney was being silly, or the "Oneyyyyy" when he was again up to something naughty. You're a true ray of sunshine in the lab, and I wish you all the best and to keep shining in your next lab.

Jean, en plus d'avoir pu libérer le plein potentiel humoristique de notre ami Oney (on ne t'en voudra pas), j'ai aussi découvert un chic type avec de belles valeurs et plein d'énergie. J'espère que tu utiliseras toute cette énergie pour suivre ton projet de vie qui te tient à cœur.

Véro, merci pour tout ce que tu es. Merci pour ta bonne humeur, pour toute l'organisation et l'aide que tu nous apportes au quotidien et qui aident grandement cette équipe. C'est un plaisir de travailler avec toi.

Aux « nouveaux » doctorants : Lucile, ne change rien à ton humour et à toute l'énergie que tu as. Hugo, tu resteras toujours de la team BC dans mon cœur. Jérémie, j'espère que tu continueras à prendre soin des chats. Blague à part, vous êtes tous les trois de superbes personnes et de bons scientifiques. Prenez soin de vous et je vous souhaite le meilleur dans vos thèses respectives et pour la suite.

Aux « Priscila » El-Kazzi & Sutto, **Cécilia**, **Insaf**, c'était toujours un plaisir de discuter avec vous. De partager sur vos projets ou vos points de vue, et d'avoir votre expertise et votre bonne humeur au quotidien.

Léa, je suis sûr que ton rire résonne encore dans le bureau et dans le cœur de plein de personnes ici. Merci pour tout et pour les visites de ton chat.

Tom, j'espère que tu as pu trouver tes marques dans notre équipe, qui est ton équipe maintenant. Tu as beaucoup évolué depuis ton premier jour, et je suis sûr que tu iras loin dans tes projets.

Aux structuralistes de l'équipe **Claire.Z**, **Candice**, **Mika**, **Sergio**, **Nikos**, j'ai adoré travailler avec vous, discuter technique et de plein d'autres choses en fonction de nos projets. Ne changez rien.

Aux chimistes de l'équipe, **Aurélie**, **Laura**, **Méderic**, **Alexandre**, j'avoue que j'ai eu du mal à me familiariser avec vos domaines au début, mais au fur et à mesure de vos présentations, j'ai vraiment découvert un monde fascinant, et surtout une belle complémentarité avec « nous » les biologistes et structuralistes.

Florian, on s'est découverts sur le tard, mais j'ai vraiment apprécié le gars que tu es, en science et humainement. Je te souhaite sincèrement le meilleur.

Fabienne, la seconde maman du labo avec Barbara, un grand merci pour tout ce que tu fais au labo et toutes tes petites attentions au quotidien.

Amélie & **Axel**, on s'est vus quelque temps avant que vous vous réorientiez et trouviez votre voie. Il faut une grande dose de courage pour cela, et en même temps, quand on a pu voir vos qualités humaines, on ne pouvait pas douter que vous réussiriez. Merci pour tout ce que vous avez apporté au labo, et je vous souhaite le meilleur pour la suite.

Jean-Claude, tu as toujours un œil très avisé et de bons conseils. Un mot pour chacun et c'est toujours plaisant de discuter avec toi.

Étienne, merci pour tes points de vue lors des réunions d'équipe et pour tout ce que tu fais lors de tes communications publiques pour une science plus ouverte et consciencieuse.

Nadia, tu as souvent été de bons conseils et d'une grande aide lors de différentes étapes de ma vie de doctorant. Merci pour tout.

Karine, la grande chimiste que tu es a toujours un œil avisé sur ce qu'on fait, même pour nous, les biologistes. J'ai grandement apprécié avoir tes avis et nos discussions au cours de ma thèse.

Bruno, le Toulousain de cœur que je suis te remercie infiniment d'avoir donné leur chance à trois jeunes biologistes à peine sortis du master. À l'époque, on ne savait pas encore dans quoi on s'embarquait réellement, tant au niveau recherche qu'humain. J'ai déjà entendu dire « qu'on a l'équipe qu'on mérite » et quand je vois la diversité et les qualités humaines et scientifiques de l'équipe que tu as créée, je dois dire que je suis assez d'accord. Je te serai toujours reconnaissant pour tout cela, et ce que cela m'a permis de vivre et d'être aujourd'hui.

L'ambiance de cette équipe est juste géniale ! Et je vous en remercie sincèrement tous.

Je ne pensais pas en arriver là en partant de Toulouse avec deux de mes collègues du master, quinze jours à peine après le premier confinement dû à notre « ami » le SARS-CoV-2, sans même avoir pu rendre mon appartement dû toujours à notre ami viral, avec juste un sac à dos pour un CDD de six mois « pour commencer ».

Cette période particulière m'a également permis de passer certains de mes meilleurs moments dans cette équipe. Plus que des collègues, ce travail intense des « Virale-Avengers » saupoudré de moments de détente et de partages nichés au cœur des calanques, nous a permis de créer des liens uniques, et je suis fier de pouvoir vous compter parmi mes amis aujourd'hui.

Plus que des amis, c'est vraiment un esprit de famille qui règne au sein de l'**AFMB**. Je voudrais ainsi remercier plus largement chacun d'entre vous. J'ai une pensée particulière pour **Denis**, **Béa** et **Pat** pour leur aide et leur présence au quotidien. Mais aussi à **Véro**, **Gerlind**, **Anaïs** et **Maria** pour votre enthousiasme ainsi que votre aide en cristallographie, en bio-physique et les projets que nous avions en commun, et je pense notamment à **Claire.D** & **Renaud** pour ce dernier point également. Je voudrais remercier également plein d'autres personnes comme Vincent pour toutes les courbatures que j'ai eues grâce à lui. Plus simplement, d'autres belles personnes que j'ai eu plaisir à croiser dans ce laboratoire, notamment **Sonia**, **Jaime**, **Rhian**, **Walid**, **Julie**, **Nicolas**, **Élodie**, **Juliette**, **Ahmed**, **Andrei**, **Maëva**, **Yves**, **Juan**, et bien sûr aussi la gang **Polo**, **Marc** & **Lana**. J'ai certainement oublié des personnes, et si c'est le cas, j'en suis désolé.

Bref, l'**AFMB**, vous êtes géniaux. Ne changez rien. Je vous souhaite le meilleur pour la suite. Que ce laboratoire continue à grandir, accueillir de belles personnes et faire de la belle science.

Je tenais à renouveler mes sincères remerciements aux différents **membres de mon jury de thèse** qui ont accepté de prendre de leur temps pour évaluer mes travaux, discuter avec bienveillance et m'apporter leur expertise.

Je voulais remercier également **tous mes professeurs** que j'ai eus au cours de ma scolarité. Dernièrement, mes professeurs de la fac **Samuel Tranier**, **Fabienne Bardou**, **Annaïk Quémard**, **Laurent Paquereau**, **Vincent Ecochard**, **Pascal Demange** et bien d'autres. Vous avez été essentiels dans ma culture de la recherche scientifique. Merci également aux personnes qui m'ont donné ma chance pour faire un stage avec eux, **Grégory Arna**l & **Harry Brumer**, ainsi que **Thomas Lautier** & **Gilles Truan**.

À mes professeurs de lycée, **Mme & M. Valérino**, **Mme Terrazzoni**, **Mr Sarda**, **Mme Jalabert**, **Mr Andrieu**, **Mme Lahaille**, **Mr Thiolier**, **Mme Bertrand**, sans oublier **Yves Casaux**, **Pierre Bartés**, **Yves Guallar**, **Hervé Lamiaud** et bien d'autres encore. **Saint Jo** a été réellement formateur pour moi. Cela m'a également énormément appris sur le travail d'équipe et l'entraide, ce qui m'a, à coup sûr, grandement aidé dans la poursuite de mes études.

De façon générale, je suis vraiment reconnaissant à chacun de mes professeurs. Avec le recul, je pense réellement que vous êtes des cultivateurs de la conscience, là pour nous aider à aiguiller notre esprit critique et notre curiosité pour le restant de notre vie. J'essaierai humblement de faire de même, à mon échelle, et essayer de perpétuer cet art qui m'a tant inspiré.

Durant toutes ces années, je me suis fait des amis que je considère aujourd'hui comme ma famille, je pense notamment à vous, **Seb & Diane**, et **Julia & Hugo**, **Christelle**, et tous mes amis du lycée, de la fac, de mes voyages et de l'AFMB.

Je voudrais finir par remercier chaque membre de ma famille. **Mes Tantes et Oncles** qui ont toujours été là pour moi. **Mes Cousins et Cousines** avec qui j'ai pu faire les 400 coups durant mon enfance. Bien sûr, **ma Mère**. **Ma Sœur** pour nous être mutuellement soutenus en tout temps. Et **mes deux Grands-Mères** qui m'ont aussi élevé et trouvaient toujours les mots justes. Je voudrais finir par remercier **mon Grand-Père** et **Mon Père**, qui m'ont grandement inspiré depuis petit, et épaulé tout au long de ma vie, ce qui m'a permis aussi d'en être là aujourd'hui.

Sur ces dernières années, **Ma Belle-Famille** y est aussi pour beaucoup. Ils ont pu voir quasiment chaque étape de ma thèse, et m'ont aussi énormément soutenu tout du long de ce parcours. Un grand merci pour tout.

Merci du fond du cœur à tous.

Pierre

À mes professeurs, ma famille, mes amis et mes pairs.

Une tache de diffraction n’est pas le résultat de l’ondulation d’une seule onde,
mais d’un travail d’équipe

Mariusz Jaskólski, 2024

La Science,
C’est faire des approximations pour espérer s’approcher de la vérité

Table des matières

Affidavit 2
Liste de publications et/ou brevets et participation aux conférences 3
Résumé 6
Abstract 8
Remerciements 9
Table des matières 18
Abréviations 22
Chapitre 1 : Introduction Générale 24
1. Introduction 25
1.1. Les virus 25
1.1.1. La découverte des Virus 25
1.1.2. Les origines présumées des virus 30
1.2. La structure des virus & leurs classifications 32
1.2.1. Les capsides et les enveloppes 33
1.2.2. Les génomes 34
1.2.3. Classification par type de génomes (Classification de Baltimore) 34
1.2.4. Diversité des supports génétiques 36
1.2.5. Classification du Comité international de taxonomie des virus (Classification ICTV) 43
1.3. L'Ordre des *Nidovirales* 46
1.3.1. Organisation génomique 46
1.3.2. Cycle de vie des Nidovirus 50
1.3.3. Événements d'insertions et conséquences 51
1.4. La famille des Coronaviridae 53
1.4.1. Structure des virions 55
1.4.2. Description génomique des Coronavirus 57
1.4.3. La réplication des Coronavirus 63
1.4.4. Le système de coiffage des ARN 64
1.5. Les polymérases virales des virus à ARN(+) 70
1.5.1. Les ARN polymérases dépendantes de l'ARN (RdRp) 70
1.5.2. L'ARN polymérase des CoVs et son complexe de réplication 75
1.6. Les exoribonucléases virales 83
1.6.1. La nsp14 des Coronavirus 83
1.7. Les méthyltransférases virales 86
1.7.1. La N7-méthyltransferase nsp14 des Coronavirus 86
1.7.2. La 2'O-méthyltransferase nsp16 des Coronavirus 87

1.8. Les antiviraux 88
1.8.1. Le RTC, une cible thérapeutique 89
1.8.2. Les nucléosides, nucléotides et leurs analogues 89
1.8.3. Les nucléotides analogues (NAs) 90
1.8.4. Les chélateurs d'ions 94
1.8.5. Les analogues du SAM 95
1.8.6. Différence entre le RTC des SARS et du MERS 95
Objectif de la thèse 97
Chapitre 2 : Matériel & Méthode 99
2. Matériel & Méthode 100
2.1. Généralités 100
2.2. Clonage des plasmides 100
2.2.1. Constructions plasmidiques 100
2.2.2. Amorces et amplifications par PCR 101
2.2.3. Mutagenèse dirigée par PCR 102
2.3. Transformation de plasmides 102
2.4. Expression et purifications des protéines 103
2.4.1. nsp12 de SARS-CoV 103
2.4.2. nsp12 de SARS-CoV-2 104
2.4.3. nsp7, nsp8 ou nsp8L7 de SARS-CoV et SARS-CoV-2 105
2.4.4. nsp14 et nsp10 de SARS-CoV-2 106
2.4.5. VHH 107
2.4.6. nsp12 de MERS 108
2.4.7. nsp8, nsp7, de MERS 109
2.5. Essais d'expression à haut débit 109
2.6. Essais de purification à haut débit 110
2.7. Isolation ADNc des VHHs 110
2.8. Phage display 111
2.9. Test Elisa 111
2.10. Assemblage & purification des complexes avec nsp14 111
2.11. BLI (Bio-Layer Interferometry) 111
2.12. Caractérisation en double référent des VHH1 et VHH2 112
2.13. SEC-MALS 112
2.14. SEC-SAXS 113
2.15. Cristallographie 113
2.16. Test d'activité ExoN 114
2.17. Test d'activité MTase by filter binding assay (FBA) 115
Chapitre 3 : Étude structurale et fonctionnelle des nsp12 des Coronavirus 116
3. Étude structurale et fonctionnelle des nsp12 des Coronavirus 117
3.1. Études du RTC de SARS-Cov-2 117
3.1.1. Présentation de l'article 1 « A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase » 118

3.1.2. Présentation des mes contributions 123
3.2. Comparaison structurale et fonctionnelle SARS & MERS 124
3.3. Étude du RTC du virus MERS 127
3.3.1. Introduction 127
3.3.2. Essais de production et de purification en système bactérien 127
3.3.3. Test à haut débit des contions de production en système bactérien 129
3.3.4. Conclusions & Perspectives 135
Chapitre 4 : Étude structurale de méthyltransférases virales 136
4. Étude structurale de méthyltransférases virales 137
4.1. Introduction 137
4.2. Étude structurale de la 2'-O-méthyltransférase de NS5 de la Dengue 137
4.2.1. Présentation de l'article « AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5 » 138
4.2.2. Présentation de mes contributions 142
4.3. Étude structurale d'une nouvelle N7-méthyltransférase 143
4.3.1. Présentation de l'article « A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome » 143
4.3.2. Présentation de mes contributions 147
4.4. Structure de nsp14 du SARS-CoV-2 148
4.4.1. Historique sur les structure de NSP14 chez SARS-CoV-1 et 2 149
4.4.2. Utilisation des nanobodies comme protéine chaperonne de la cristallogenèse 149
4.4.3. Résultats nsp14 SARS-CoV-2 150
4.4.4. Conclusions 178
4.4.5. Perspectives 179
5. Conclusion générale 180
6. Bibliographie 182
Sommaires des figures 197
Sommaires des tableaux 199
ANNEXES 200
a. Publications 201
A1 : A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase 202
A2 : AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5 212
A3 : A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome 249
b. Financement du projet de thèse 264
c. Figures supplémentaires 265
C1 : Tableaux des Essais d'expression à haut débit 266

C2 : Gel SDS-PAGE 4-20% des échantillons élution effectuées en haut débit, après purifications IMAC-Ni 272
C3 : Compléments des données de SEC-SAXS 275
284

Abréviations

ADN : Acide DésoxyriboNucléique
ADNc : ADN complémentaire
ARN : Acide RiboNucléique
ARNds : ARN double brin (double-stranded (ds)RNA)
ARNg : ARN génomique
ARNm sg : ARNm sous-génomiques
ARNm : ARN messager
AZT : Azidothymidine
Coronaviridaea : Coronavirus
CDR : Complementary-Determining Regions
DMV : Vésicules à Double Membrane (double-membrane vesicles)
E : Enveloppe
eIF4E : facteur 4E d'initiation de la traduction chez les eucaryotes (eukaryotic translation initiation factor 4E)
EndoU : Endoribonucléase
ExoN : ExoNucléase
FDA : Food & Drug Administration
HE : Hémagglutinine-Estérase
ICTV : La classification ICTV pour Comité international de taxonomie des virus (International Committee on Taxonomy of Viruses)
IDP : protéines désordonnées intrinsèquement (intrinsically disordered proteins)
IFN : interféron
Kpb : Kilo paire de bases, soit mille paire de bases
LUCA : Dernier ancêtre cellulaire universel (Last Universal Common Ancestor)
MALS : Diffusion de la lumière à multiples angles (Multi-Angle Light Scattering)
M : Membrane
MDA5 : gène 5 associé à la différenciation du mélanome (melanoma-differentiation-associated gene 5)
MERS : Syndrome respiratoire du Moyen-Orient (Middle East Respiratory Syndrome)
MHV : Virus hépatite murine (murine hepatitis virus)
miARN : microARN
Mpb : Méga parie de bases, soit un million de paires de bases
MTase : MétylTransférase
N : Nucléocapside
N7-MTase : N7-MétylTransférase
NA : Nucléotide Analogue
NCLDV : Virus nucléocytoplasmiques à grand génome d'ADN (nucleocytoplasmic large DNA viruses)

Nidovirales : Nidovirus
NiRAN : Nucléotidyltransférase associée au Nidovirus RdRp
NRTI : Nucléosidiques de la RétroTranscriptase Inverse
NSP : Protéines Non-Structurale (Non-Structural Proteins)
NTP : Ribonucléotides TriPhosphates
OMS : Organisation Mondiale de la Santé
ORF : Cadres de Lecture Ouverts (Open Reading Frame)
Pb : Paires de bases
Pi : Phosphate inorganique
PPi : PyroPhosphate
pp1a : PolyProteine 1a
pp1ab : PolyProteine 1ab
PrEP : Prophylaxie Pré-Exposition
RBP : Protéine de liaison à l'ARN (RNA-binding proteins)
RdRp : ARN polymérase ARN-dépendante (RNA-dependent RNA-polymerase)
RE : réticulum endoplasmique
RIG-1 : protéine du gène 1 inductible à l'acide rétinoïque (retinoic acid inducible gene 1 protein)
RMP : GS-441524 mono-phosphate (Remdesivir MonoPhosphate)
RTC : Complexe de réplication et Transcription (Replication/Transcription Complex)
RTP : GS-441524 tri-phosphate
TRS : Séquence Régulatrice de Transcription (trancription regulatory sequence)
TRS_B : TRS body
TRS_L : TRS leader
S&S : Forme et Taille (Shape and Size)
SAM : S-AdénosylMéthionine
SAH : S-Adénosyl-L-Homocystéine
SARS-1 ou SARS-CoV-1 : Coronavirus du Syndrome Respiratoire Aigu Sévère 1 (Severe Acute Respiratory Syndrome Coronavirus 1)
SARS-2 ou SARS-CoV-2 : Coronavirus du Syndrome Respiratoire Aigu Sévère 2 (Severe Acute Respiratory Syndrome Coronavirus 2)
SAXS : Diffusion des rayons X aux petits angles (Small Angle X-ray Scattering)
SEC : Chromatographie d'Exclusion Stérique (Size Exclusion Chromatography)
S : Spike
TMV : Mosaïque du tabac (Tobacco mosaic virus)
VHB : Virus de l'Hépatite B
VHC : Virus de l'Hépatite C
VIH : Virus de l'Immunodéficience Humaine
VYR : Réduction de la Croissance Viral (Virus Yield Reduction)
ZBS : Doigts de zinc (Zinc Binding Site)

Chapitre 1 : Introduction Générale

1. Introduction

1.1. Les virus

Les virus sont les entités biologiques les plus abondantes sur notre planète. Le mot virus vient du latin *vīrus*, qui signifie venin, poison ou substance toxique, et qui pouvait rendre malade. Ils ont donc un impact important sur l'évolution des biosphères et sur l'écologie mondiale.

1.1.1. La découverte des Virus

En 50 après J.-C., Cornelius Aulus Celsus disait que « La rage est causée par un virus ». Même si à l'époque les maladies infectieuses étaient encore un concept métaphysique, cela ouvrait la porte à la découverte d'agents infectieux. Plus proche de nous, vient le concept de microbe (du grec mikrós « petit » et bíos « vie ») introduit par le chirurgien français Charles-Emmanuel Sédillot en 1878 pour parler des êtres vivants infiniment petits.

Deux ans plus tard, les travaux de Pasteur et de Roux ont permis de prouver que la rage était une maladie causée par un agent infectieux qu'ils ne pouvaient voir, un tout petit microbe (Lwoff 1957). Alors que certains microbes commençaient à pouvoir être observés et classés, ils prirent le nom de bactérie. Les premières techniques étaient des techniques de coloration comme celle qui a permis la découverte des Mycobactéries par Robert Koch (améliorée par Ziehl et Neelsen), puis la coloration inventée par Hans Christian Gram qui permet de classer les différentes bactéries en fonction des propriétés de leurs parois. Mais d'autres microbes restaient invisibles sous les microscopes optiques des chercheurs.

1.1.1.1. Le virus modèle, la mosaïque du tabac

En 1876, Adolf Mayer, chimiste et microbiologiste, devient directeur d'une station agricole expérimentale à Wageningen (Pays-Bas). Durant cette période, il étudie une maladie qui affecte les plants de tabac des régions alentour. Cette dernière a pour symptôme la coloration anormale des feuilles, avec un enroulement sur elles-mêmes des feuilles les plus âgées, ainsi que le ralentissement de la croissance des plants. Cette maladie rend impropres les feuilles pour la fabrication des cigares à partir des plants de tabac contaminés. Mayer est le premier tenter d'identifier la cause, et propose même quelques moyens de lutte contre cette maladie après différentes expérimentations. Devant la diversité de noms que les producteurs ont donnés à cette

maladie et pour éviter les confusions avec d'autres, Mayer propose de donner un nom international à cette maladie qu'il appelle la mosaïque du tabac (Mayer A., 1886).

Durant ses travaux, le microbiologiste démontre également que lorsque des feuilles malades sont broyées, le jus qui en résulte contient une substance infectieuse capable de rendre malades des plants sains. Il observe également que ce liquide infectieux reste efficace même après dilutions dans de l'eau, et entreprend d'étudier les effets des concentrations d'inoculum sur le taux d'infection.

Quelques années plus tard, Dimitri Ivanovski reprend les travaux de Mayer et constate que « la maladie de la mosaïque conserve ses qualités infectieuses, même après filtration à travers un filtre-chandelle de Chamberland. » (Chamberland C., 1884). Cette technique de filtration sur céramique était connue à l'époque pour retenir les bactéries, et le filtrat était considéré comme stérile. Après ce constat, Ivanovski conclut en disant que la cause de la mosaïque du tabac est : soit une toxine de bactérie ; soit une très petite bactérie capable de passer à travers les pores des filtres de Chamberland (Ivanovski D., *et al*, 1892).

Il faudra attendre un autre microbiologiste pour avoir une début de réponse, Martinus Beijerinck. Dans un article publié en 1898, il explique qu'après avoir tenté de cultiver dans des conditions aérobies ou anaérobies une potentielle bactérie de la mosaïque du tabac extraite des feuilles et passée sur un filtre de Chamberland, en vain. Il constate néanmoins que cet extrait de plante malade est capable d'infecter « un nombre infini » de plantes saines. Il est le premier à utiliser le mot virus pour parler de l'agent contenu dans le virus. D'autres observations sur le « comportement du virus » sur les plantes et leurs environnements. Cela permet au chercheur de conclure que pour se multiplier, ce virus a une préférence pour les plants jeunes, et de ce fait se conserve dans le sol où des débris de plantes infectées sont capables d'en infecter de nouveaux. Beijerinck a certainement été le premier à entrevoir la nature très particulière des virus, ces microbes « incapables de se multiplier indépendamment » et qui « sont amenés dans les jeunes cellules en division où ils se multiplient en grande quantité sans perdre, d'une façon quelconque, leur propre individualité » (Beijerinck M.W., *et al*, 1942).

D'autres virus furent découverts par la suite comme la fièvre aphteuse chez les animaux en 1898, ou encore la fièvre jaune chez l'homme, en 1901 (Lustig et Levine 1992), la virologie est née.

Au cours du 20ème siècle, cette discipline continue de se développer, permettant ainsi de mieux comprendre ce qu'est un virus, et étant le premier, la mosaïque du tabac (Tobacco mosaic virus (TMV)) est devenue un modèle d'étude de ce nouveau monde.

1.1.1.2. La virologie moderne

Aujourd'hui, nous découvrons encore l'immense diversité des virus. Celle-ci s'étend maintenant des petits virus à ARN comme les *Picornavirus* (environ 7,5 kb) aux grands virus à ADN comme les *Poxvirus* (jusqu'à 375 kb). Et plus récemment encore, des virus dit géants comme les *Mimivirus* (de plus de 1 Mb) ont été découvert, de même que les virophages, ces parasites de virus comme le virus Sputnik qui infecte les *Mimivirus* pour se répliquer.

Virus	Type de génome	Taille du génome
Picornavirus	ARN	Environ 7,5 kb
Virus VIH	ARN	Environ 9,1 kb
Virus Zika	ARN	Environ 10 kb
Influenza	ARN	Environ 13,5 kb
Virus Ebola	ARN	Environ 19 kb
Virophages	ADN/ARN	De 18-30 kb
Coronavirus	ARN	De 27-32 kb
Adénovirus	ADN	Environ 36 kb
Herpèsvirus	ADN	De 150-250 kb
Poxvirus	ADN	De 130-375 kb
Mimivirus	ADN	Plus de 1 Mb
Pandoravirus	ADN	De 1-2,5 Mb

Tableau 1 : Diversité des tailles de génomes viraux

Ce tableau présente un échantillon de virus, classés en fonction de leur type de génome (ARN ou ADN) et de la taille approximative de leur génome (en kilobases, kb, ou mégabases, Mb). Les virus répertoriés incluent des pathogènes humains et des virus géants.

Cette diversité s'accompagne d'une grande variété de structures, de cycles de réplication et d'hôtes. Les activités de l'homme ont également favorisé les zoonoses, ces infections virales transmises des animaux à l'homme, et sont devenues un sujet central de la recherche en virologie. On peut notamment citer l'émergence de virus comme Zika, Ebola, mais surtout le virus de l'immunodéficience humaine (VIH) qui a énormément stimulé la recherche après sa découverte dans les années 80, et plus récemment, le SARS-CoV-2.

Les avancées en matière de séquençage à haut débit ont aussi révolutionné notre capacité à identifier de nouveaux virus et à comprendre leur évolution. Cela nous permet ainsi de mieux comprendre les familles de virus et notamment d'analyser le comportement des maladies vétérinaires et de leurs réservoirs naturels, et ainsi

prévenir l'émergence de potentiels futurs agents pathogènes humains, ou du moins y être mieux préparés.

Les Coronavirus, en particulier, illustrent parfaitement ce phénomène. Initialement considérés comme des pathogènes d'intérêt vétérinaire principalement , plusieurs d'entre eux ont déjà émergé comme le SARS-CoV en 2003, le MERS-CoV en 2012, et plus récemment le SARS-CoV-2 en 2019.

1.1.1.3. Le cas des virus géant et de leurs parasites

Lwoff est le premier à établir une série de critères pour différencier le monde des virus de celui des cellules (Lwoff 1957). Puis, en 1966, il affine ces critères avec Tournier pour mieux classer ces virus (André Lwoff et Tournier 1966). Pour eux, un virus est une entité composée d'un génome (ARN ou ADN), mais ne peut pas se diviser pour se reproduire, contrairement à une cellule ou un parasite intracellulaire. Il ne code pas pour une machinerie de traduction des protéines ou de production d'ATP (ou d'autres nucléotides) et préfère utiliser et consommer ceux de la cellule infectée.

D'après ces critères, Lwoff et Tournier ne mentionnent pas la taille des virions. Mais c'est pourtant cette propriété physique de la taille des virions qui a permis d'isoler les virus de l'environnement, de la découverte du TMV jusqu'à aujourd'hui, où des filtres de différentes tailles sont utilisés.

C'est peut-être à cause de cette technique d'isolation des virus à partir d'un échantillon environnemental qui a retardé la découverte de ce qu'on appelle les virus géants. En 1992, lors d'une collecte d'un échantillon d'amibes lors d'une épidémie de pneumonie en Angleterre, un autre micro-organisme parasitaire obligatoire qui se développait parmi elles a été découvert. Ce dernier ressemblait à une bactérie de type coque, entouré de fibres, et réagissait à la coloration de Gram comme gram-positif (Birtles *et al*. 1997).

Malgré cette caractéristique, il n'a pas été possible de le séquencer via des amorces universellement utilisées pour séquencer l'ARNr 16S (Birtles *et al*. 1997). Il aura fallu une dizaine d'années pour établir que ce parasite était en réalité d'origine virale. Du fait de sa réaction à la coloration de Gram, il fut nommé *Mimivirus* pour le fait qu'il « imite un microbe » (mimicking microbe). Ces virus se distinguent par la taille de leurs génomes, la diversité des gènes qui les composent, et bien sûr par les dimensions de leurs virions de forme icosaédrique de 400 nm de diamètre (Scola *et al*. 2003). Pour comparaison, les plus gros virus connus de l'époque étaient les *Poxvirus*, en forme de cylindre, mesurant généralement environ 200 nm de large et 300 nm de long. Ce premier Mimivirus pouvait avoir un génome d'ADN atteignant la taille d'environ 1,2 Mb (Raoult *et al*. 2004).

Lors des observations de ces *Mimivirus* en microscopie électronique, il a également été découvert la présence de petites taches blanches à l'intérieur de ces virus. On découvrira plus tard qu'il s'agissait d'un autre virus à l'intérieur de ce virus, et celui-ci fut nommé Sputnik (Voir Figure 1 ; La Scola *et al.* 2008).

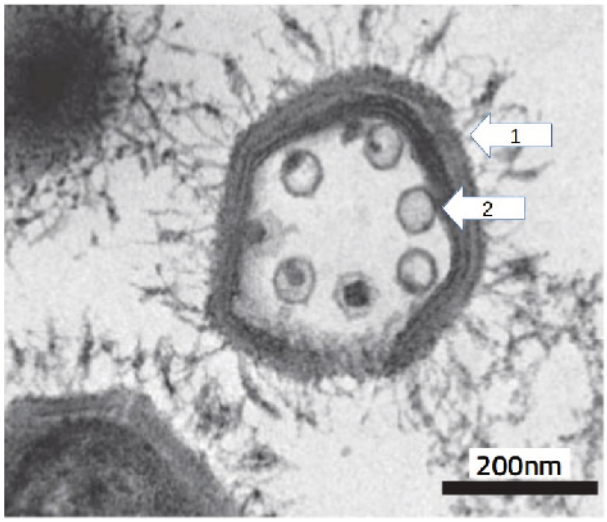


Figure 1 : Cas d'infection d'une capside de Mimivirus par le virus Sputnik

Observations par microscopie électronique à transmission en coloration négative d'un des cas possible d'infection de *Mimivirus* par le virus Sputnik. 1) Une capside de *Mimivirus* 2) Une particule virale du virus Sputnik. Image extraite de (La Scola *et al.* 2008).

Au cours des 10 années suivantes, différents autres virus géants ont été identifiés et caractérisés, gagnant à chaque fois en taille et en nombre de gènes. En 2013 fut publié un nouveau travail de l'équipe de Chantal Abergel et Jean-Michel Claverie, et après l'analyse de plusieurs collectes d'échantillons marins, un nouveau type de virus géant avait été découvert. Il s'agit d'une nouvelle famille de virus. L'un de ses membres découverts, P. salinus, a un génome pouvant atteindre le double des précédents *Mimivirus*, avec près de 2,5 Mb (Philippe *et al.* 2013).

Ce type de virus, de par leurs dimensions et leur diversité génomique (> 1000 gènes), peut ainsi être plus « complexe » que des bactéries pathogènes (Colson *et al.* 2011) ce qui questionne leur origine potentielle.

Pour l'heure, les premiers virus ont été capables d'infecter les amibes, les insectes, les algues, les poissons et certains invertébrés, mais pas l'homme (Barthelemy, Goto, et Faure 2019).

1.1.2. Les origines présumées des virus

La découverte des virus puis des virus géants a profondément redéfini notre perception des virus et à alimenter les théories sur leurs origines, soulevant de nouvelles questions fondamentales sur le dernier ancêtre cellulaire universel (LUCA). Les virus, aujourd'hui considérés comme des parasites intracellulaires obligatoires, continuent de susciter des débats quant à leur origine, et trois hypothèses ont été explorées au fil des ans (Forterre 2006).

Une des premières théories avancées est celle de la coévolution des virus, suggérant que les virus pourraient représenter des reliques d'un monde pré-cellulaire (Koonin et Dolja 2013). Ainsi les virus seraient à l'interface entre la chimie et la biologie, et évoluant parallèlement aux premières formes de vie cellulaire jusqu'à nous. Ce scénario est soutenu par la diversité observée dans les systèmes d'expression et de réplication des génomes viraux, ce qui suggère une longue histoire évolutive antérieure aux cellules elles-mêmes (Koonin et Dolja 2013).

Une autre théorie importante est celle de la régression cellulaire, qui postule que les virus auraient évolué à partir de cellules complexes ayant perdu certaines de leurs fonctions au fil du temps. Privés de mécanismes essentiels comme la traduction des protéines, ces cellules simplifiées se seraient adaptées à un mode de vie parasitaire, nécessitant l'infecter d'autres cellules pour se répliquer. Cette hypothèse a pris de l'ampleur avec la découverte des virus géants (Abergel, Legendre, et Claverie 2015 ; Legendre *et al*. 2012).

L'hypothèse de l'évasion propose que les virus seraient issus de plasmides ou de transposons ayant acquis au cours de l'évolution la capacité de se répliquer et de se propager d'une cellule à l'autre (Forterre et Krupovic 2012 ; Claverie 2006). Selon cette théorie, les virus se seraient formés à partir de ces éléments génétiques mobiles et seraient devenus infectieux par l'acquisition de gènes spécifiques au cours du temps. Ce scénario est renforcé par des observations sur les rétrovirus, dont le mode de réplication évoque cette origine par évasion de gènes cellulaires.

Malgré ces trois hypothèse, il n'est pas possible de trancher pour l'une ou l'autre. La découverte des virus géants a joué un rôle crucial dans l'évolution de notre compréhension du monde viral en offrant des indices sur la continuité potentielle du vivant entre les cellules et les virus. En effet, des éléments comme les gènes, tels que ceux codant pour des enzymes de synthèse d'ARNt, rappellent des fonctions que l'on croyait réservées aux cellules. Pourtant les *Pandoravirus* et les *Megaviridae*, par exemple, possèdent des gènes codant pour des acide-aminé-ARNt ligases permettant la synthèse ARNtPro, ARNtMet, ARNtTrp (mais pas des autres), bien que ces virus soient dépourvus des composants clés du ribosome et des enzymes du cycle de Krebs, essentiels à la vie cellulaire (Philippe *et al*. 2013 ; Colson *et al*. 2011).

D’autres recherches récentes, notamment en 2018 sur une analyse métagénomique du groupe des *Klosneuvirus*, ont révélé que certains virus codent pour des éléments de la machinerie de traduction, possédant des ligases pour l'ensemble des 20 acides aminés. Pourtant, une analyse phylogénétique complète suggère que ces gènes ont probablement été acquis au cours de l'évolution à partir des cellules hôtes infectées, et non hérités d’un ancêtre cellulaire commun (Schulz *et al*. 2017). En 2019, la découverte de *Megaklothovirus* horridgei, un virus de plus de 4 µm de long, a encore repoussé les limites de ce que l’on considérait comme un virus, renforçant l’idée que les virus géants pourraient être capables de générer de nouveaux gènes, une notion qui alimente la spéculation sur leur rôle potentiel dans l’évolution de nouvelles fonctions biologiques (Barthelemy, Goto, et Faure 2019).

Pour le moment, l'absence de certaines caractéristiques fondamentales, comme des homologues protéiques impliqués dans la division cellulaire, ne permet pas de considérer ce type de virus comme dérivé de cellules. Pour exemple, même les plus petits parasites intracellulaires avec un génome simplifié possèdent encore des mécanismes de division cellulaire que les virus géants n’ont pas (Claverie et Abergel 2013 ; Claverie, Abergel, et Legendre 2018).

Ainsi, les virus géants sont loin d’être « de simple virus » qui se contentent d'infecter leurs hôtes. Les nouvelles fonctions génétiques découvertes dans ces derniers brouillent encore plus la frontière entre le monde viral et cellulaire. Les origines des virus ne sont toujours pas clairement définies, mais ce qui est certain c’est que la classification du monde des virus devra tenir compte de ces découvertes récentes, notamment avec une réévaluation de la classification ICTV qui sera vue dans la partie « Classification du Comité international de taxonomie des virus (Classification ICTV) ».

1.2. La structure des virus & leurs classifications

L'étude des premières structures de virus a été réalisée dès 1935 par cristallisation des premiers virus découverts. Cela a permis d'observer qu'un virus est un assemblage de "protéines autocatalytiques" qui a besoin de cellules vivantes pour se multiplier (Stanley 1935). Cette découverte a été récompensée par le prix Nobel de chimie en 1946. Grâce à la diffraction des rayons X, pour le cas de la mosaïque du tabac, il a été possible de déterminer que le virus est « en forme de bâtonnets » et semble être « identique en section transversale » (Bawden *et al.* 1936). Cette description de la structure du virus a été confirmée par d'autres observations en 1939, cette fois au microscope électronique (Kausche, Pfankuch, et Ruska 1939 ; Voir Figure 2).

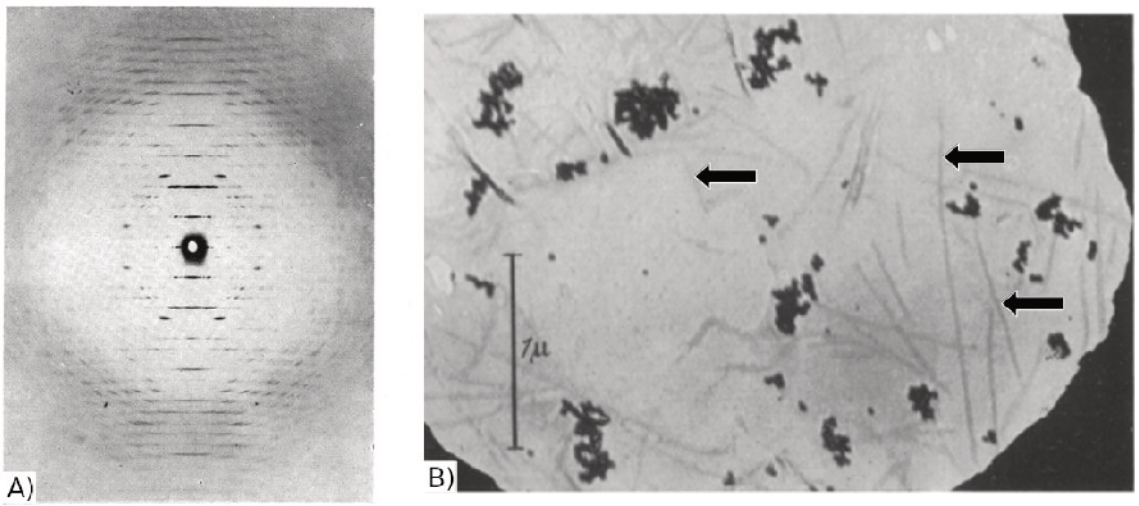


Figure 2 : Clichés de TMV

A) Diagramme de diffraction des rayons X par un gel de cristaux du TMV B) Image de TMV capturé en microscopie électronique. Clichés respectivement extraits de (Witz 1969) et de (Kausche, Pfankuch, et Ruska 1939).

Plus tard, l'auto-assemblage de virus fonctionnels à partir de protéines et d'ARN isolé a été étudié, et il a été démontré que seul l'ARN était capable d'avoir le pouvoir infectieux du virus, et que l'ARN est ainsi le support de l'information génétique (Gierer et Schramm 1956).

1.2.1. Les capsides et les enveloppes

La capside virale a deux fonctions principales : compacter et contenir le génome viral pour le protéger de l'environnement. Les capsides sont composées de sous-unités appelées capsomères. Un virus peut avoir plusieurs types de capsomères, certains formant la paroi du virion, d'autres des spicules. Ces capsomères ont la propriété de s'auto-assembler pour former la capside. Il est généralement admis qu'il existe quatre formes principales de capsides :

- Icosaédrique, pour former une sphère à 20 faces, comme certains Coronavirus
- Hélicoïdale, pour former un tube sphérique, comme le TMV
- Conique, comme pour le VIH
- Mixte, c'est notamment le cas des bactériophages

Les virus qui n'ont que leur capside remplie de leur génome dans leur forme virion sont appelés des virus nus. Ils infectent généralement les cellules hôtes par endocytose (médiée via la reconnaissance de sites sur la spicule du virus), ou via pinocytose (absorption du virus dans une petite vésicule, sans nécessairement de liaison spécifique à un récepteur).

D'autres virus possèdent une enveloppe lipidique en plus de leur capside. Ces virus enveloppés acquièrent cette membrane lorsqu'ils bourgeonnent à partir de la membrane cellulaire de l'hôte. L'enveloppe est généralement dérivée de la membrane plasmique, du réticulum endoplasmique ou de l'appareil de Golgi de la cellule hôte. Les spicules et autres protéines de surface du virus sont alors imbriquées directement dans cette enveloppe pour leur permettre de jouer leur rôle crucial dans l'attachement et l'entrée du virus dans la cellule hôte. Cette membrane permet au virus d'adapter plus facilement et rapidement la composition de ses protéines de surface, mais le rend aussi plus fragile et sensible à son environnement. Les familles de virus enveloppés sont les plus nombreuses, et les Coronavirus en font partie.

Plus récemment, grâce notamment aux avancés en microscopie électronique, il a été observé qu'en réalité la diversité des capsides virales était bien plus grande que ce que l'on pensait auparavant. Par exemple, rien que pour les capsides icosaédriques, il existe plus d'une vingtaine de variantes et évolutions de capsomères, en réalité d'origine très diverses (Krupovic et Koonin 2017). Cela suggère qu'il pourrait exister d'autres structures encore plus complexes dans la nature.

On commence à s'apercevoir de cela entre autres, avec l'étude des virus géants. En effet, ces derniers comme les *Mimivirus*, *Pandoravirus*, et *Pithovirus*, possèdent des capsides d'une taille et d'une complexité sans précédent.

Par exemple, la capside du *Mimivirus* a un diamètre d'environ 500 nm et présente une structure icosaédrique modifiée avec des fibres protéiques à sa surface. Ces

fibres jouent un rôle crucial dans l'interaction initiale avec les cellules hôtes (Abergel, Legendre, et Claverie 2015 ; Claverie et Abergel 2009).

Les *Pandoravirus*, quant à eux, possèdent une capside ovoïde unique en son genre, d'environ 1 µm de longueur, qui défie les classifications traditionnelles des formes de capsides. Cette structure est recouverte d'une couche dense de glycoprotéines, formant une "forêt" de protubérances à sa surface (Philippe *et al*. 2013 ; Abergel, Legendre, et Claverie 2015 ; Legendre *et al*. 2018).

Le *Pithovirus* se distingue pour sa part avec sa forme en amphore et sa taille impressionnante de 1,5 µm de long, ce qui en fait l'un des plus grands virus connus. Sa capside est composée d'une couche protéique striée, surmontée d'un "bouchon" spécialisé à une extrémité, qui s'ouvre pour libérer le contenu viral dans la cellule hôte (Abergel, Legendre, et Claverie 2015).

Leurs capsides uniques de virus géant soulèvent de nouvelles questions sur l'évolution des structures virales et les mécanismes d'assemblage de capsides aussi complexes (Abergel, Legendre, et Claverie 2015).

1.2.2. Les génomes

Les génomes sont la structure qui sert de support à l'information génétique. Ils sont composés d'ADN ou d'ARN et protégés par la capside ou la nucléocapside, et possèdent bien d'autres caractéristiques qui seront abordé dans le chapitre suivant « 1.2.4. Diversité des supports génétiques ». Le génome permet au virus, après son entrée dans un nouvel hôte, d'avoir les informations nécessaires à sa réplication, et ainsi de former de nouveaux virions.

1.2.3. Classification par type de génomes (Classification de Baltimore)

À mesure que les différents types de génomes et la façon dont ceux-ci sont répliqués sont étudiés, un système de classification regroupant ces caractéristiques a été créé : la classification de Baltimore proposée par David Baltimore en 1971 (Baltimore 1971). Le principe est de trier les virus en 7 groupes comme suit (Voir Figure 7).

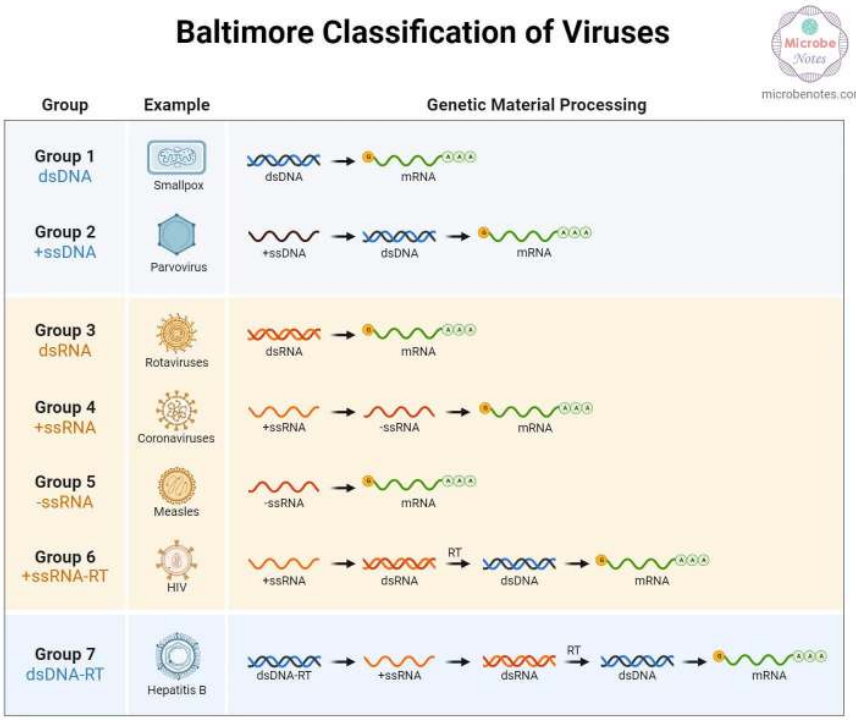


Figure 3 : Classification de Baltimore

Cette classification regroupe les virus en sept groupes selon la nature de leur génome (ADN ou ARN, simple brin ou double brin) et leur stratégie de réplication et de transcription. Les exemples de virus sont indiqués pour chaque groupe, avec un schéma illustrant le processus de conversion du matériel génétique viral en ARNm. Cette classification met en lumière la diversité des mécanismes utilisés par les virus pour produire les protéines nécessaires à leur cycle de vie dans les cellules hôtes. Classification de Baltimore extraite de microbenotes.

Cette classification a l'avantage de synthétiser les différentes stratégies de réplication des virus ainsi que de l'expression de leurs protéines. Cependant, cette classification ne tient pas compte des processus évolutifs, notamment le gain ou la perte de fonction.

Les caractéristiques et la diversité des supports génétiques regroupés dans cette classification vont maintenant être approfondies.

1.2.4. Diversité des supports génétiques

1.2.4.1. Les génomes à ARN ou ADN

Au début des années 1950, une série de trois articles historiques a été publiée dans le journal Nature (Watson et Crick 1953 ; Wilkins, Stokes, et Wilson 1953 ; Franklin et Gosling 1953) sur l'acide désoxyribonucléique (ADN). En s'appuyant sur ces données et leurs travaux, James Watson et Francis Crick proposent une structure détaillée de l'ADN en 1953 (Watson et Crick 1953b).

Aujourd'hui nous connaissons la structure exacte de cette molécule d'ADN et sa variante, l'acide ribonucléique (ARN). Les « briques » de ces molécules sont des nucléotides et seront présenté plus en détail dans la partie « 1.7.1. Les nucléosides, nucléotides et leurs analogues ». Mais ce qu'il faut retenir ici, c'est que cette chaîne de nucléotides est composée de trois sous-parties. Un ribose dont seul une légère variation en 2' du ribose, permet d'avoir soit de l'ADN si cette position est 2'-désoxy, ou de l'ARN si cette dernière est 2'-hydroxyl. Une base azotée, qui peut être soit de type purine, soit de type pyrimidine. Les bases puriques sont les résidus adénine (A) et guanine (G), et les bases pyrimidiques sont les résidus thymine (T) et cytosine (C) pour l'ADN, ou uracile (U) pour l'ARN. Le pentose et sa base azotée sont appelés nucléoside, puis quand en 5' du pentose, 1, 2 ou 3 groupements phosphate peuvent être ajoutés, on parlera alors de nucléotide, qui est un nucléoside 5'-mono, di ou tri-phosphate.

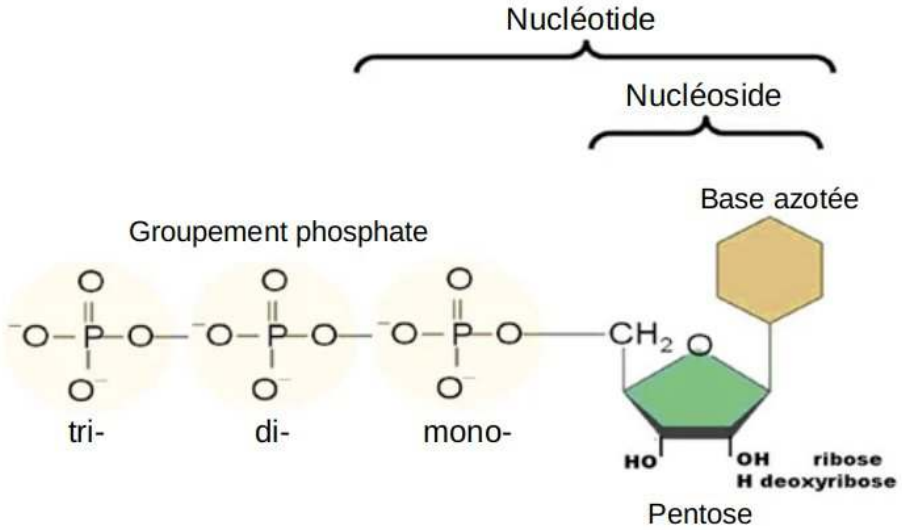


Figure 4 : Schéma du Nucléotide

Ce schéma représente la structure d'un nucléotide, constitué d'un nucléoside (composé d'une base azotée et d'un pentose) et d'un ou plusieurs groupements phosphate. Le nucléoside, lui-même composé d'un pentose (ribose ou désoxyribose) et d'une base azotée, et devient un nucléotide lorsqu'un ou plusieurs groupements phosphate sont ajoutés. Les formes tri-, di- et monophosphate du nucléotide

sont représentées à gauche du schéma, illustrant l'importance du nombre de groupements phosphate dans la fonction et l'énergie du nucléotide.

Les nucléotides ainsi formés peuvent s'apparier les uns aux autres, toujours dans le sens 5' vers 3' du pentose. Ainsi, le nucléoside triphosphate sera ajouté via une réaction de condensation sur le premier phosphate en 5', avec le 3'-OH du nucléotide précédent. Cette nouvelle liaison phosphodiester ainsi formée entraînera la libération d'une molécule d'eau et des deux phosphates liées qu'on appelle pyrophosphate inorganique.

1.2.4.2. L'information génétique

L'enchaînement des nucléotides et la combinaison des bases azotées (A, T, C, G) utilisées comme code génétique permettent « d'écrire » sur ce support génétique. Ce dernier permet de contenir toutes les informations nécessaires à la vie et à la reproduction de chaque organisme, quel qu'il soit.

Cette information peut être répliquée à l'identique par les polymérases grâce à un système de correspondance permis par les 2 liaisons hydrogène qu'il est possible de créer lorsqu'on positionne un A en face d'un T (ou U), ou 3 liaisons hydrogène lorsqu'on place un C en face d'un G.

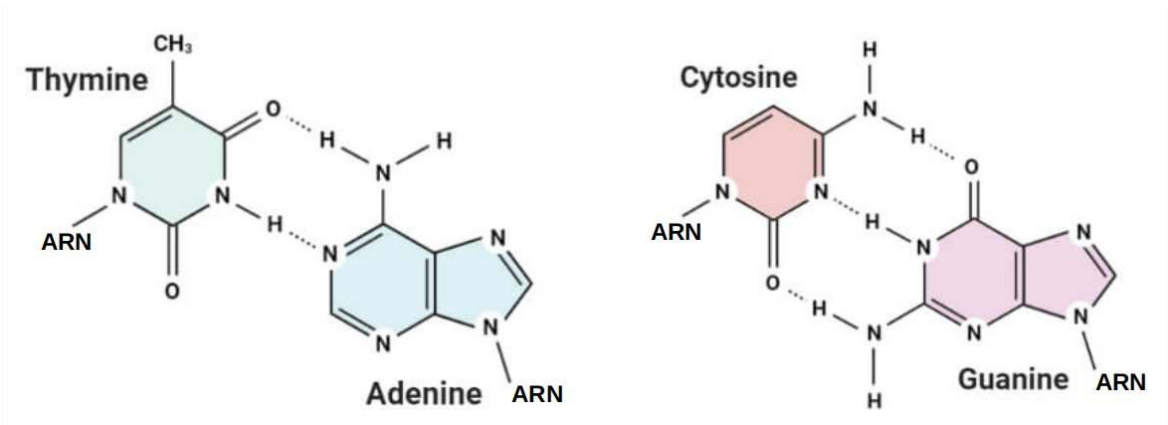


Figure 5 : Schéma des bases azotées et de leurs liaisons hydrogènes

Cette figure représente les structures chimiques des quatre bases azotées de l'ADN (adénine, thymine, cytosine, et guanine) et leurs liaisons hydrogène spécifiques qui permettent l'appariement des bases. La mention « ADN » représente le radicale de chaque base d'ADN ou ARN. Dans l'ARN, la thymine est remplacée par l'uracile. Ces liaisons hydrogène (respectivement deux et trois), stabilisent la double hélice d'ADN ou d'ARN et sont essentielles pour la complémentarité des brins lors de la réplication et de la transcription.

Ce système très robuste permet de gérer l'information génétique inscrite dans ces molécules indifféremment du type de support, ADN ou ARN. Cependant, un génome d'ARN sera généralement moins stable chimiquement, et plus sensible à l'environnement cellulaire de son hôte, qui peut avoir des RNases pour lutter contre ce type de virus. Le support ARN permet également d'avoir une traduction directe du génome en protéines s'il s'agit d'ARN(+), à l'inverse d'un génome ARN(-) (comme nous le verrons dans le chapitre « 1.2.4.5. Polarité des génomes »).

Un génome d'ADN sera considéré comme plus stable et moins sujet à la dégradation dans la cellule hôte, et de par le fait qu'il peut utiliser des mécanismes de réparation, qui le rendent moins susceptible de muter. Cependant, ce type il aura besoin d'ajouter une étape de transcription avant de pouvoir faire traduire ses protéines.

Le nombre de copies du génome joue également un rôle capital dans la préservation de l'information génétique au cours des réplications du virus, en fonction qu'il s'agisse d'un génome à 1 ou 2 brins, composé d'ARN ou d'ADN, comme nous allons le voir dans la paragraphe suivant.

1.2.4.3. Doubles brins, simple brin

Les brins d'ARN ou d'ADN peuvent être sous forme d'un seul brin (monocaténaire) ou sous forme d'un appariement de deux brins complémentaires (bicaténaire).

Une molécule monocaténaire aura ainsi son génome plus compact, et ainsi une particule virale potentiellement plus compacte, ce qui peut conférer des avantages pour l'infection. Cela peut également permettre une réplication plus rapide du génome, mais peut aussi être plus fragile et engendrer un taux de mutations plus important. Ce dernier point peut être un avantage pour s'adapter à un nouvel hôte ou environnement. Les brins monocaténaires sont aussi plus facilement reconnus comme du « non-soi » par l'immunité cellulaire. Ils doivent ainsi trouver des systèmes de contournement de cette dernière pour s'en prémunir. Par exemple, les virus à ARN monocaténaire, comme le virus de la grippe ou les Coronavirus, utilisent des mécanismes pour échapper à la détection immunitaire (Park *et al*. 2022).

Un génome bicaténaire, quant à lui, est généralement plus stable (structurellement) et moins susceptible aux mutations, car l'information génétique est portée deux fois (Duffy, Shackelton, et Holmes 2008). Cependant, l'accès aux gènes durant la transcription est plus complexe, car il faut dans un premier temps ouvrir le double brin, ce qui, *in fine*, permet une régulation plus fine de l'expression des gènes viraux. Un génome bicaténaire sera aussi plus résistant aux mécanismes de l'immunité cellulaire, ce qui peut être un avantage dans la phase d'infection du virus. Par exemple, les virus à ADN bicaténaire, comme le virus de l'herpès, sont capables de persister dans l'organisme hôte en échappant à la détection immunitaire. Les virus à ARN bicaténaire existent également, comme les Birnaviridae, qui peuvent infecter certains animaux

d’élevage, ou les Reoviridae, qui peuvent causer des gastro-entérites et des symptômes respiratoires chez l’Homme.

1.2.4.4. Les génomes segmentés

Certains génomes peuvent être répartis en plusieurs morceaux, on dit alors qu’ils ont un génome segmenté. Cette caractéristique peut permettre au virus d’échanger certains segments avec d’autres souches en cas d’infections multiples, ce qui peut favoriser son adaptation et l’émergence d’une nouvelle souche. Cette segmentation peut également permettre une réplication plus rapide et un contrôle de l’expression des gènes plus aisé.

Cette structure peut également rendre plus complexe l’assemblage de tous les segments pour l’assemblage des virions, mais peut aussi permettre une forme plus compacte de ce dernier.

Un exemple notable de ce type de virus est le virus de la grippe (Influenza), qui possède un génome d'ARN segmenté en 8 segments. Cette segmentation permet au virus de la grippe de subir des recombinaisons génétiques, un processus connu sous le nom de "réassortiment", lorsque deux souches différentes infectent la même cellule. Ce mécanisme est à l'origine de l'émergence de nouvelles souches de grippe, parfois avec des propriétés antigéniques différentes, ce qui peut entraîner des épidémies ou des pandémies.

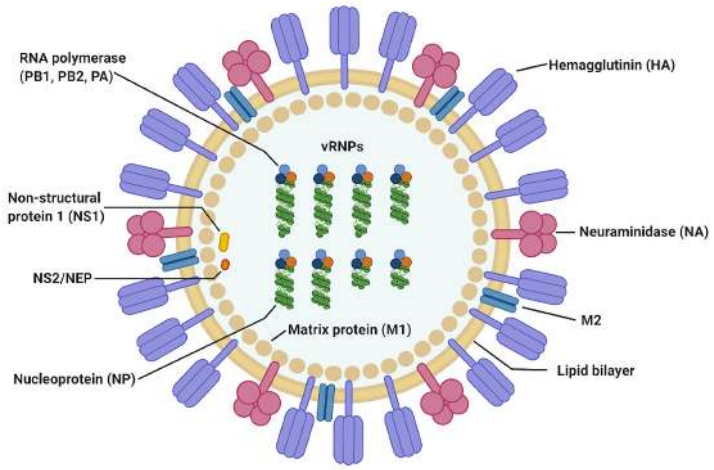


Figure 6 : Schéma du virus de la grippe (Influenza)

Ce schéma illustre la structure du virus de la grippe, qui possède un génome segmenté en 8 segments d'ARN. Chaque segments du génome est encapsidé par des protéines nucléaires (NP) et associé à des complexes enzymatiques de polymérases virales (PB1, PB2, PA). Les glycoprotéines de surface, comme l'hémagglutinine (HA) et la neuraminidase (NA), jouent un rôle clé dans l'entrée et la libération du virus des cellules hôtes. Ce génome segmenté permet au virus de la grippe d'échanger des

segments entre souches lors de la co-infection, augmentant ainsi la variabilité génétique. (Extrait de Wikipédia).

Les *Bunyaviricetes*, possèdent un génome segmentés. Cette segmentation leur permet de s'adapter rapidement à différents hôtes et environnements, en redistribuant les segments du génome lors de co-infections par plusieurs souches d'un même virus, ce qui est crucial pour leur survie, évolution et leur propagation. Les Arenavirus, quant à eux, ont un génome segmenté en 2 segments d'ARN, ce qui leur confère une flexibilité génétique similaire.

La segmentation des génomes est plus courante chez les virus à ARN, mais il existe aussi des virus à ADN segmenté. Par exemple, les poxvirus, comme celui de la variole, possèdent un génome d'ADN bicaténaire segmenté.

Ainsi, la segmentation des génomes viraux peut permet d'offrir des avantages significatifs en termes de flexibilité génétique, de réplication rapide et de contrôle de l'expression des gènes. Cependant, cette caractéristique peut également ajouter des contrainte en termes d'assemblage des virions et de stabilité génétique.

1.2.4.5. Polarité des génomes

Une des particularités des génomes monocaténaires est qu'ils peuvent également être caractérisés selon le sens (ou polarité) de leurs génomes.

Un génome d'ARN de sens positif (+) a l'avantage de pouvoir être utilisé comme ARN messager (ARNm). Les ribosomes de l'hôte pourront ainsi traduire directement les protéines virales à l'entrée du virus. Cela permet une réplication rapide et efficace, car le génome peut immédiatement servir de matrice pour la synthèse des protéines nécessaires à la réplication virale. Des exemples de virus à ARN de sens positif incluent les Picornavirus, les Flavivirus et les Coronavirus. Un virus à ADN, lui, n'a qu'une étape supplémentaire à ajouter, celle de la transcription en ARNm, avant que les protéines virales puissent être traduites.

L'autre avantage des génomes à sens positif est que, pour leur réplication, une seule copie complémentaire négative sert de matrice pour la création de nombreuses nouvelles copies du génome (Gorbalenya *et al*. 2006). Cette efficacité de réplication est un atout majeur pour la propagation rapide du virus.

Les virus à génome d'ARN de sens négatif (-) nécessitent un complexe de réplication (composé d'un polymère de nucléoprotéines autour des segments et d'une polymérase), qui permet une étape de transcription d'ARNsg pour la traduction des premières protéines virales. En conséquence, les virus à ARN(-) emportent avec eux leur propre ARN polymérase ARN-dépendante (RdRp) directement dans le virion. Parmi les virus à ARN de sens négatif, on peut citer les virus d'Ebola, de la rage et de la grippe. Cette caractéristique est également observée chez certains virus particuliers à ARN(+), les rétrovirus (Bedwell et Engelman 2021).

Bien qu'un génome à sens négatif (-) puisse sembler, à première vue, moins avantageux qu'un génome à sens positif (+), il se distingue par une meilleure protection contre l'immunité cellulaire. En effet, le fait que les génomes ARN(-) soient mieux protégés par leurs nucléocapsides et qu'ils transportent leurs polymérases permet, d'une part, de ne pas avoir à produire cette dernière (transcription et traduction) par l'hôte, et d'autre part, de produire séquentiellement les ARNsg puis les protéines nécessaires à la réplication de ce type de virus. Ainsi, les génomes à sens négatif sont moins facilement reconnus par les mécanismes de défense de l'hôte, ce qui peut conférer un avantage en termes de survie et de propagation du virus (Walpita et Flick 2005).

Il existe aussi des génomes de polarité négative particuliers, dits ambisens. Ces derniers représentent une catégorie spécifique de génomes viraux qui combinent leurs gènes sur un même brin d'ARN (répartis en plusieurs segments), mais dont l'expression est décalée dans le temps. Ainsi, l'antigénome permet la transcription et l'accès à d'autres gènes du génome durant une étape ultérieure du cycle viral. Cette configuration permet une flexibilité accrue dans la régulation de l'expression des gènes et la réplication virale. On peut citer comme exemple les Arenavirus, qui utilisent cette stratégie (Papageorgiou *et al*. 2021). Chez les Arenavirus, le génome est réparti sur deux segments (les segments S et L). Chaque segment subit une transcription primaire, permettant la production d'ARNm ensuite traduit en protéines, respectivement la nucléoprotéine (NP) pour le segment S et l'ARN polymérase (L) pour le segment L. Une fois suffisamment de protéines accumulées, la réplication commence par la synthèse des antigénomes qui serviront de matrice pour générer de nouveaux génomes. Ces antigénomes serviront également de matrice pour la transcription « tardive » des ARNm codant pour la glycoprotéine (GP) et la protéine Z du virus (Martínez-Sobrido, Cheng, et De La Torre 2016 ; Figure 7).

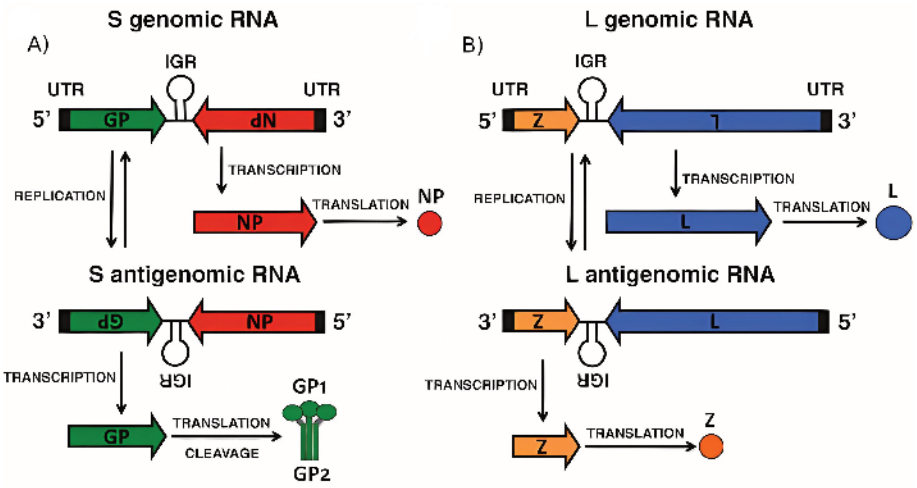


Figure 7 : Schéma du génome ambisens d'Arenavirus.

Ce schéma illustre les segments S (panel a) et L (panel b) du génome d'un Arenavirus. Le segment S code pour la nucléoprotéine (NP) et les glycoprotéines (GP), tandis que le segment L code pour l'ARN polymérase (L) et la protéine Z. Chaque segment subit une transcription primaire, produisant l'ARN messager nécessaire pour la traduction de leurs premières protéines pour lesquels ils codent (respectives la protéine NP pour le segment S, et la protéine L pour le segment L). Par la suite, une réplication supplémentaire des brins d'ARN est nécessaire pour générer des brins antigénomiques qui permettent la transcription et la traduction complémentaires des protéines GP et Z. Extrait de (Martinez-Sobrido, Cheng, et De La Torre 2016).

Ainsi, cette dualité offerte aux virus ambisens leur confère une capacité d'adaptation unique. Cette organisation génomique leur permet de contrôler temporellement l'expression de leurs gènes et d'adopter une réponse plus fine aux conditions spécifiques lors de l'infection de l'hôte, améliorant ainsi leur capacité à échapper aux défenses immunitaires. Cette flexibilité génétique leur permet de s'adapter rapidement aux variations environnementales et aux différentes pressions sélectives rencontrées dans divers hôtes.

1.2.4.6. Des polymérases à acides nucléiques synthétisent les génomes

Les enzymes capables de répliquer ces génomes sont les polymérases.

Nous avons commencé à aborder les ARN polymérases ARN-dépendantes (RdRp) présentes chez les virus à ARN. Les RdRp synthétisent de l'ARN à partir d'une matrice d'ARN. Leur équivalent pour l'ADN est l'ADN polymérase ADN-dépendante.

Viennent ensuite les polymérases qui permettent de changer de type de produit par rapport à la matrice d'origine. C'est le cas des ARN polymérases ADN-dépendantes, capables de synthétiser de l'ARN à partir d'une matrice d'ADN. C'est notamment ce type d'enzyme qui permet la production d'ARNm. Le dernier type de polymérase est l'ADN polymérase ARN-dépendante. Elle synthétise un brin d'ADN complémentaire (ADNc) en utilisant l'ARN viral comme matrice. C'est notamment le cas des transcriptases inverses que l'on retrouve chez les Rétrovirus comme le VIH.

Toutes ces polymérases, en plus des caractéristiques sur le type de matrice et de produits pris en charge, se distinguent par bien d'autres parcours évolutifs. Certaines ont évolué pour répliquer leurs génomes avec une haute fidélité, avec même des systèmes de correction d'erreur intégrés à l'enzyme. Cela leur permet d'avoir un taux de mutation plus faible, au prix d'un génome répliqué plus lentement. A l'opposé, dans le cas des Coronavirus, une polymérase extrêmement rapide « sous traite » la vérification et la correction d'erreur d'incorporation à d'autres de leurs enzymes. Il s'agit alors de complexes de réplication multiprotéiques, nous y reviendrons plus tard en détail.

1.2.5. Classification du Comité international de taxonomie des virus (Classification ICTV)

La classification de Baltimore commençait à manquer de détail à mesure que les techniques de séquençage s'amélioraient. Les théories sur l'évolution de certains virus ont évolué à mesure que de nouveaux génomes étaient connus, et des virus que l'on pensait génétiquement distants étaient en fait bien plus proches qu'anticipé (Koonin *et al*. 2020).

La classification des virus a évolué avec le Comité International de Taxonomie des Virus (ICTV). Ce comité scientifique a pour objectif d'aller encore plus loin que la classification de Baltimore en classant hiérarchiquement les virus. Pour cela, tout comme ce qui est utilisé aujourd'hui pour les autres organismes vivants, une classification en 8 rangs permettant d'avoir plus de profondeur dans la classification a été choisie (Voir Tableau 2 ; Siddell *et al*. 2023)

Rang taxonomique	Nombre de taxons	Suffixe pour les taxons
Domaine	6	-viria
Règne	10	-virae
Phylums	17	-viricota
Classe	40	-viricetes
Ordre	72	-virales
Famille	264	-viridae
Sous-famille	182	-virinae
Genre	2818	-virus
Espèce	11273	
Rang taxonomique	**Nomenclature**	**Exemple**
Domaine	Majuscule initiale et italique	*Viruses*
Règne	Majuscule initiale et italique	*Orthornavirae*
Phylums	Majuscule initiale et italique	*Riboviria*
Classe	Majuscule initiale et italique	*Pisoniviricetes*
Ordre	Majuscule initiale et italique	*Nidovirales*
Famille	Majuscule initiale et italique	*Coronaviridae*
Sous-famille	Majuscule initiale et italique	*Orthocoronavirinae*
Genre	Majuscule initiale et italique	*Betacoronavirus*
Espèce	en italique (si en Latin)	Betacoronavirus pandemicum

Tableau 2 : Classification du Comité international de taxonomie des virus (ICTV) de 2023, ainsi qu'un exemple de nomenclature

La partie du haut de ce tableau présente les différents rangs taxonomiques utilisés pour classer les virus, ainsi que le nombre de taxons associés à chaque rang proposé par le Comité international de taxonomie des virus (ICTV) en 2023 (Zerbini *et al*. 2023). Les suffixes utilisés pour nommer les groupes de virus à chaque niveau taxonomique sont également indiqués. La classification comprend 6 domaines, 10 règnes, 17 phylums, et va jusqu'à plus de 11 000 espèces. La partie basse de ce tableau, un exemple de nomenclature (A.E. Gorbalenya, Lauber, et Siddell 2019) pour le SARS-CoV-2.

Puis, comme l'ARN ribosomal est utilisé chez les eucaryotes comme gène universellement conservés et de référence pour construire leurs arbres phylogéniques, des gènes de référence sont également utilisés dans cette classification. On peut par exemple citer l'utilisation du gène de la RdRp, dont le degré de dérive génétique fournit une horloge moléculaire adéquate pour la classification des virus à ARN. D'autres enzymes ainsi que la structure du génome viral sont également utilisés pour construire cette classification (Kuhn *et al*. 2019).

La découverte récente des virus géants et notamment les pandoravirus, a demandée à la communauté scientifique de revoir la classification pour intégrer ces virus particuliers, et un nouvel embranchement a été créée. Ce Phylum est nommé *Nucleocytoviricota* pour ce qu'on appelle désormais les virus nucléocytoplasmiques à grand génome d'ADN ou NCLDV (nucleocytoplasmic large DNA viruses ; Rodrigues *et al*. 2021). Cet embranchement comme composé de deux classes, dont celle correspondant aux virus géants est celle des *Megaviricetes*, composée aujourd'hui de 3 ordres, 12 familles, 5 sous-familles, 35 genres, 75 espèces au total, jusqu'aux prochaines découvertes et mises à jour de cette classification. L'autre classe de ce Phylum correspond au *Pokkesviricetes* dont les Poxvirus font partie.

ICTV répond donc à un besoin de classer plus finement les virus comparé à la classification de Baltimore qui regroupe simplement les virus par type de réplication de leurs génomes.

En 2023, l'ICTV avait permis la classification au total de 11 273 espèces de virus, réparties en 264 familles (Zerbini *et al*. 2023).

1.3. L'Ordre des *Nidovirales*

L'ordre des *Nidovirales* est composé de virus à enveloppe, à ARN(+) non segmenté, dont la taille varie de 13 Kb pour les Arteriviridae à 42 Kb pour les Mononiviridae (Vabret *et al*. 2009). L'une des particularités de cet ordre est qu'il possède les virus (appelé Nidovirus) ayant parmi les plus grands génomes d'ARN de sens positif connus à ce jour. Cet ordre est constitué de 130 espèces de virus réparties en 14 familles selon la mise à jour de l'ICTV de 2023. Ces virus ont pour hôtes principaux les animaux, tels que les animaux d'élevage, les poissons, les oiseaux et les humains (Le Poder 2021). Ce sont des virus zoonotiques présentant un enjeu vétérinaire et de santé publique.

Il y a une grande diversité génétique dans cet ordre, notamment dans la composition de la machinerie pour la réplication virale (réplication du génome, production des virions, systèmes de défense contre l'immunité cellulaire). Cette diversité a certainement contribué à la variété de la taille des génomes à ARN(+) que nous observons aujourd'hui.

Le nom des *Nidovirales* vient du fait que lors de la production des ARNm sg, ces derniers ont leurs extrémités 3' regroupées ensemble lors de leur synthèse (Knoops *et al*. 2008 ; Neuman, Angelini, et Buchmeier 2014). Cette caractéristique peut faire penser à un « nid » (à se représenter comme un point fixe) duquel s'élancerait la production d'ARN, nidus en latin, ce qui a donné le nom de cet ordre (Cavanagh 1997).

1.3.1. Organisation génomique

Les génomes des Nidovirus sont monocaténaires de sens positif (+). Ils possèdent une coiffe en 5', ce qui permet la reconnaissance des ARN par le ribosome, et en 3' se trouve une séquence polyadénylée (Lai et Cavanagh 1997). Ces caractéristiques font qu'un ARN viral peut être directement pris en charge par l'appareil de traduction de la cellule hôte infectée dès sa libération dans le cytoplasme.

Cela permet de produire le complexe de réplication et de transcription virale (RTC ; qui sera décrit dans la partie "1.4.3.1. Le Complexe de Réplication et Transcription (RTC) des Coronavirus") contenu sur les cadres de lecture ouverts (ORF) 1a et 1b, et ces deux ORF codent pour ce qu'on appelle les protéines non-structurales (nsp). Les ORF suivants codant pour les protéines structurales et accessoires (Voir partie "1.4.1.1. Description des Protéines Structurales").

Dans la figure 8, nous pouvons apercevoir différentes familles de virus appartenant à l'ordre des *Nidovirales* qui ont cette même organisation génomique.

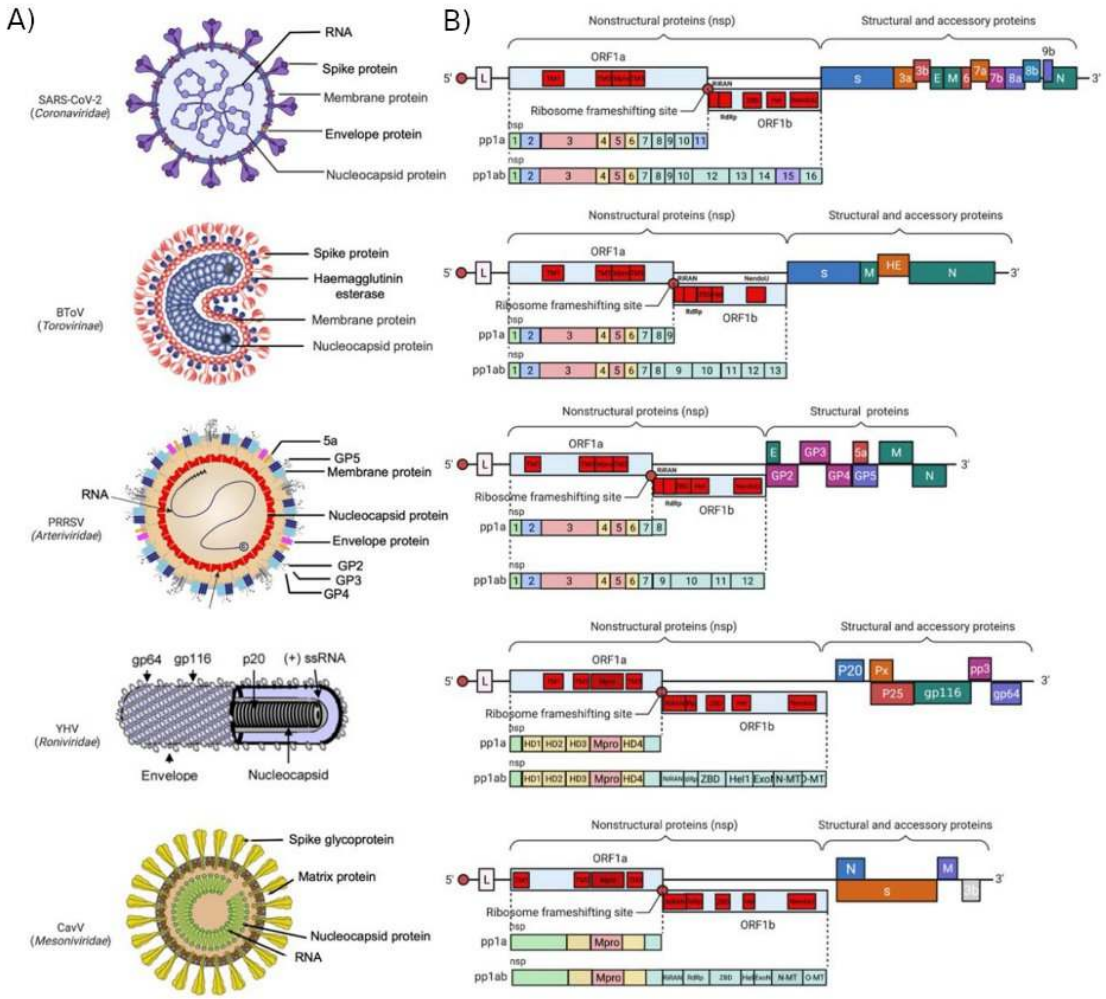


Figure 8 : Diversité génomique et structurel de familles et sous familles de l'ordre des *Nidovirales*

Ce schéma représente les différences structurelles et organisationnelles des génomes de diverses familles de l'ordre des *Nidovirales*, comme les *Coronaviridae*, *Torovirinae*, *Arteriviridae*, *Roniviridae*, et *Mesoniviridae*. Le panel A illustre la morphologie des virus avec leurs protéines de surface et de capside, tandis que le panel B présente l'organisation des génomes, mettant en évidence les cadres de lecture ouverts (ORF), les protéines non structurales (nsp) et les protéines structurales et accessoires, reflétant la diversité des génomes, tant au niveau de leurs tailles qu'au niveau fonctionnelle pour ces virus. Figure extrait de (Liao *et al*. 2024).

Les deux premiers ORF ainsi transcrits sont les ORF1a et ORF1b, qui codent pour environ les deux tiers du génome viral. Ces deux ORF codent respectivement pour la polyprotéine 1a (pp1a) et la polyprotéine 1ab (pp1ab). et seront abordé plus en détail dans la partie « 1.4.2.1. Description des ORF1a et ORF1b ». Selon les virus de cet ordre, il peut y avoir jusqu'à 16 nsp sur la polyprotéine la plus grande, la pp1ab, qui regroupe la traduction des ORF1a et de l'ORF1b.

Ces polyprotéines sont clivées par une protéase principale appelée Mpro ou 3CLpro, et il peut y avoir jusqu'à 3 protéases en tout selon les familles (Mielech *et al*. 2014 ; Roe *et al*. 2021).

Une fois les protéines nsp libérées, la RdRp et ses partenaires vont pouvoir former le complexe de réplication et de transcription (RTC). Ce dernier va permettre la synthèse des ARN génomiques (ARNg) et la transcription d'ARNm sous-génomiques (ARNm sg). Cela nécessite le passage par une étape intermédiaire via la production d'ARN de sens négatif pour servir de matrice.

La production d'ARNm sg permet *in fine* aux virus de cet ordre de pouvoir avoir une gestion plus fine de la production des protéines virales. En effet, les ARN des Nidovirus possèdent deux régions non codantes régulatrices, appelées TRS (transcription regulatory sequence ; Di *et al*. 2017).

Ce mécanisme de régulation utilise la similarité de séquence (une séquence de type ACGAAC pour SARS-CoV et SARS-CoV-2 par exemple) entre plusieurs zones du génome, pour transcrire des séquences d'ARNsg(-) de plus ou moins grande longueur. En effet, lors de la transcription des ARNsg(-), le RTC peut faire une pause en s'arrêtant sur l'une de ces séquences qu'on appelle TRS_B (pour TRS "body" ou corps) avant de reprendre la transcription sur une autre séquence TRS plus proche de la séquence 5' qu'on appelle TRS "leader" (TRS_L ; Wang *et al*. 2021). Ce changement de matrice permettra la production d'ARN sg(-) de tailles différentes qui permettront à leur tour la synthèse d'ARNm sg(+ ; Brant *et al*. 2021 ; Liao *et al*. 2024).

Comme la transcription complète du génome (transcription continue), ce mécanisme de transcription dite discontinue permet à chaque ARN sg d'avoir la même séquence 5' et la même séquence 3'. Cette transcription TRS_L-TRS_B représente 90 % des ARNsg(-) transcrits (Brant *et al*. 2021). La fréquence du saut du TRS_B vers le TRS_L détermine l'abondance respective des ARNsg(-).

Les TRS_B sont très présents sur le dernier tiers 3' du génome et moins présents au début du génome, du fait de la taille des ORF1a et 1ab (Di *et al*. 2017). La première région TRS_B en 3' se trouve devant la région codant pour la protéine N, ce qui entraîne une surreprésentation de cette région comparée aux autres régions codantes de plus de 50 % (Brant *et al*. 2021).

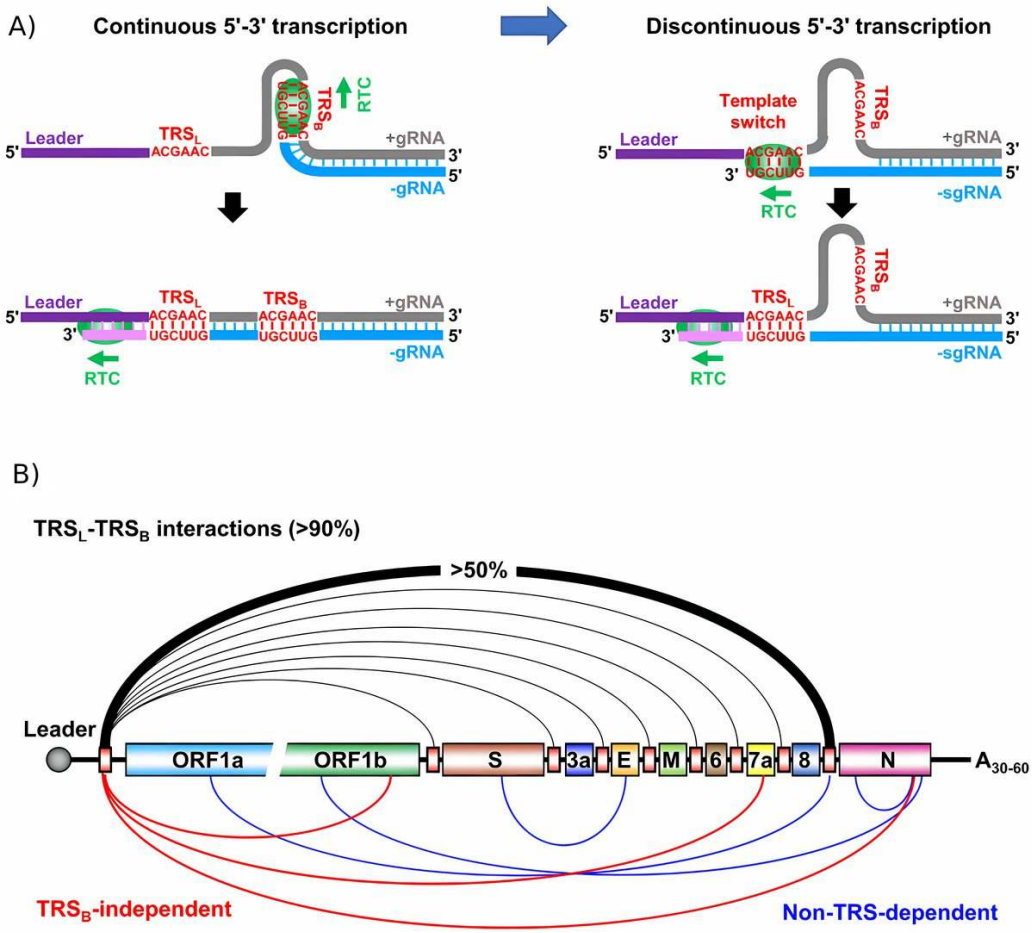


Figure 9 : Schéma de la transcription des ARNm sg

Ce schéma représente les mécanismes de transcription continue et discontinue. A) Ce schéma présente les deux mécanismes de transcription, continue et discontinue, des ARNm sous-génomiques, impliquant les séquences TRS (Transcription Regulatory Sequence). Dans la transcription discontinue, un basculement de matrice se produit au niveau de la séquence TRS_B vers la séquence TRS_L, facilitant ainsi la formation des ARNm sg de différentes longueurs. B) Représentation des interactions TRS_L-TRS_B (en noir) qui correspondent à >90% des ARNm sg. Viennent ensuite (en rouge) les interactions TRS_B indépendantes, et (en bleu) les interactions qui ne nécessites pas de TRS noté « Non-TRS-dependent » pour la transcription d'ARNm sg. Ce processus est essentiel pour la régulation de l'expression des protéines structurales et accessoires. Schéma extrait de (Brant *et al*. 2021)

Il existe également deux autres mécanismes de transcription. Le premier est dit TRSB-indépendant et dans ce cas-là, le RTC n'a pas besoin de TRS_B pour interrompre sa transcription avant de rejoindre le TRS_L et finir son transcrit. Ce déplacement du RTC serait dû à des structures secondaires spécifiques, différentes de celles permises par les TRS_B (Chrisman *et al*. 2021). Enfin, il existe aussi une transcription dite non-TRS-dépendante, qui n'utilise pas les séquences TRS mais des sites "cryptiques" ou non canoniques pour l'initiation de la transcription. Ces transcrits supplémentaires découverts récemment peuvent être des petits ORF codant pour des protéines tronquées ou fusionnées. Cette particularité et les rôles potentiels dans la réplication

ou la pathogenèse chez les Nidovirus restent encore à approfondir (Malone *et al*. 2022).

Au final, ces mécanismes permettent l'obtention d'au moins 14 ARN sg(-) différents, qui permettront à leur tour la transcription d'ARNm sg(+), ce qui va permettre de moduler le niveau d'expression des protéines virales finales dans la cellule de l'hôte.

1.3.2. Cycle de vie des Nidovirus

Les Nidovirus partagent un cycle de vie similaire. Après fixation du virus à la surface d'une cellule hôte, celui-ci entrera principalement par endocytose dépendante de la clathrine (Li *et al*. 2023) ou de la cavéoline (Trevino *et al*. 2024), fusionnant sa membrane avec la cellule pour former un endosome. L'entrée par macropinocytose est également possible, entraînant alors la formation de macropinosomes (Freeman *et al*. 2014). Ces trois types de vésicules contenant le virus sont ensuite maturées et acidifiées au cours du temps, ce qui aura pour effet de déstructurer les protéines de la capside, libérant ainsi le génome du virus de cette dernière. La vésicule finira par libérer le génome viral dans le cytoplasme de la cellule hôte.

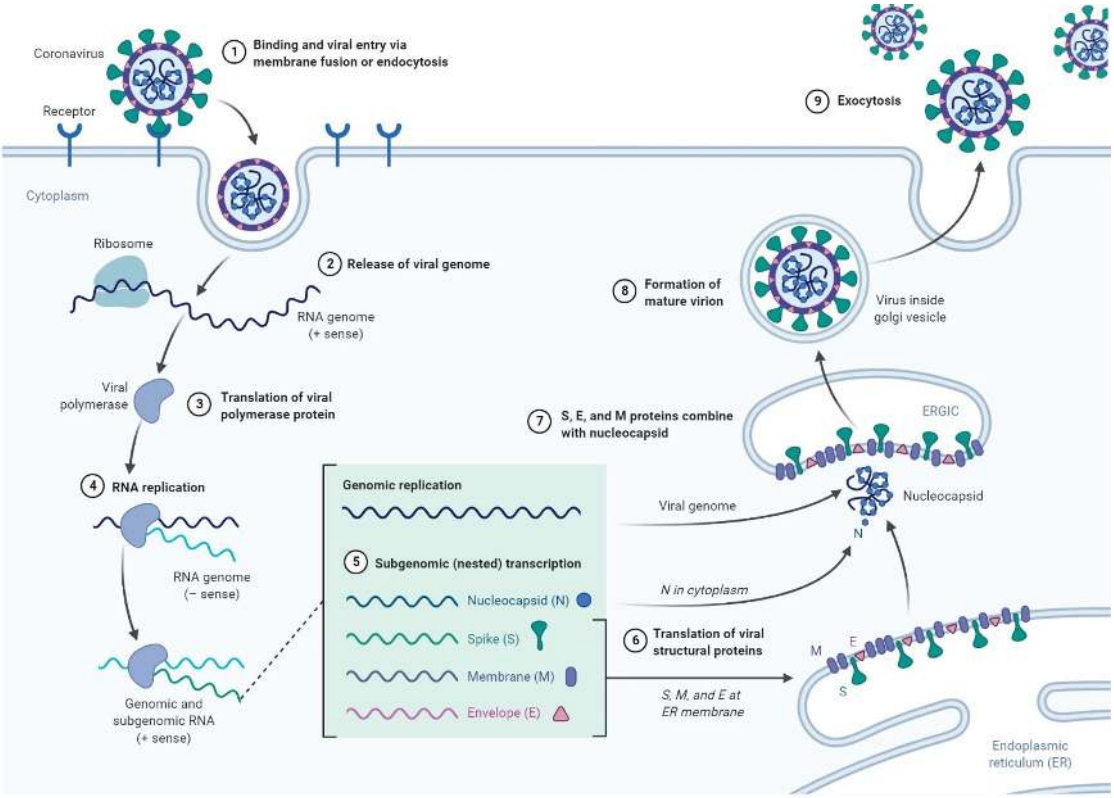


Figure 10 : Cycle viral type d'un Coronavirus

Ce schéma décrit les principales étapes du cycle de réplication des Coronavirus. 1) Le virus se lie à son récepteur et pénètre dans la cellule via fusion membranaire ou endocytose. 2) Le génome viral à ARN est libéré dans le cytoplasme. 3) La traduction des protéines pour la réplication du génome virales a

lieu. 4) L'ARN(-) viral est répliqué. 5) Transcription d'ARN sous-génomique pour permettre la production des autres protéines virales, dont les protéines structurales. 6) Les protéines structurales sont traduites et s'assemble au niveau de la membrane du RE. 7) Les protéines S, E et M commencent à se combiner autour du génome compacté avec la protéine de nucléocapside (N). 8) Le virion mature formé, 9) va être libéré par exocytose. Cycle viral type d'un Coronavirus extraite de (Liao *et al*. 2024).

Comme précédemment présenté, les premières protéines à être produites sont les protéines non-structurales (nsp). Une fois les protéines nsp libérées, les "usines à virions" pourront se mettre en place avec la formation de vésicules à double membrane (DMV) à partir des membranes du réticulum endoplasmique (RE) grâce aux protéines nsp3, nsp4 et nsp6, si l'on prend l'exemple des Coronavirus (Knoops *et al*. 2008 ; Neuman, Angelini, et Buchmeier 2014). Ainsi, les DMV nouvellement créées permettront de débuter la réplication du génome viral (ARNg) et de différents ARNm sous-génomiques (ARNm sg) via des matrices intermédiaires d'ARNsg(-) comme décrit précédemment. Cette étape de la réplication virale dans les DMV permet de préparer le matériel viral nécessaire à la production des prochains virions, et cela dans un micro-environnement qui limite l'exposition à l'immunité cellulaire (Brant *et al*. 2021 ; V'kovski *et al*. 2015 ; Santerre *et al*. 2021).

Certaines des dernières nsp vont également circuler librement dans le cytoplasme et jouer différents rôles de protection contre l'immunité cellulaire de l'hôte (Otter *et al*. 2024).

Après la formation des DMV, différentes protéines structurales impliquées dans la formation des virions pourront ainsi être traduites à partir des ARNm sg(+). Ces protéines structurales (qui seront décrites dans la partie « 1.4.1.1. Description des Protéines Structurales ») vont alors créer des bourgeonnements à partir du RE, et avec la condensation de l'ARNg par la nucléoproteine, la formation de nouveaux virions pourra avoir lieu.

Le rôle exact des protéines nsp sera décrit dans la partie « 1.4.2.2. Description des Protéines Non-Structurales (NSP) ».

1.3.3. Événements d'insertions et conséquences

Les Nidovirus se classent parmi les virus à ARN les plus complexes et leur long génome les distingue des autres groupes de virus à ARN. Tout en maintenant la structure de base de leur organisation génomique, les génomes des Nidovirus se seraient progressivement étendus par la duplication de gènes, l'acquisition de nouveaux gènes et probablement aussi par des recombinaisons de leurs ARNg.

L'expansion de la taille et la stabilisation de ces grands génomes d'ARN chez les virus à ARN (+) seraient dues à plusieurs facteurs. D'une part, l'évolution a probablement favorisé chez ces virus une spécialisation de la RdRp dans son activité de réplication du génome, alors qu'une autre enzyme spécialisée dans l'activité ExoN pouvait prendre le relais de cette fonction nécessaire au maintien de la stabilité du génome (Voir partie « 1.6. Les exoribonucléases virales » ; Gorbalenya *et al*. 2006 ; Ferron *et al*. 2021).

La stabilisation globale de l'ARNg via des méthylations, comme celle apportée par la coiffe des ARN (qui sera abordée dans la partie « 1.4.4. Le système de coiffage des ARN ») médiée par des MTase acquises au cours du temps, semble également être l'un des événements majeurs dans l'accroissement de la taille des génomes des Nidovirus (Gorbalenya *et al*. 2006 ; Ferron *et al*. 2021).

1.4. La famille des Coronaviridae

Parmi les virus de la famille des Coronaviridae, les coronavirus sont principalement responsables chez l'Homme de maladies respiratoires légères, de type rhume (Van Der Hoek, Pyrc, et Berkhout 2006).

Cependant, l'identification du Coronavirus responsable du syndrome respiratoire aigu sévère (SARS-CoV) comme cause de l'épidémie de SRAS en 2003 a mis en évidence le risque de transmission zoonotique de Coronavirus hautement pathogènes pour les humains. Plus tard, la découverte du coronavirus appelé syndrome respiratoire du Moyen-Orient (MERS-CoV), responsable du MERS, une maladie respiratoire virale signalée pour la première fois en Arabie Saoudite en 2012, a renforcé cette crainte. Fin 2019, l'émergence d'un nouveau type de syndrome respiratoire aigu sévère (SARS-CoV-2), provoquant cette fois-ci une pandémie mondiale de COVID-19 en moins de 6 mois. Ces événements soulignent le potentiel de ces virus à causer des maladies graves dans la population humaine et l'importance de la recherche pour s'en prémunir dans les années à venir.

Selon l'ICTV, les Coronavirus contiennent 3 sous-familles : les *Letovirinae*, les *Orthocoronavirinae* et les *Pitovirinae*. Seule la sous-famille des *Orthocoronavirinae* contient plusieurs genres, à savoir les *Alpha*-, *Beta*-, *Gamma*- et *Deltacoronavirus*. Quand on parle « des Coronavirus », en parle d'ailleurs souvent cette sous famille en particulier. Les trois Coronavirus humains cités plus haut, qui sont les plus étudiés, font tous partie du genre des *Betacoronavirus* (Voir Figure 11).

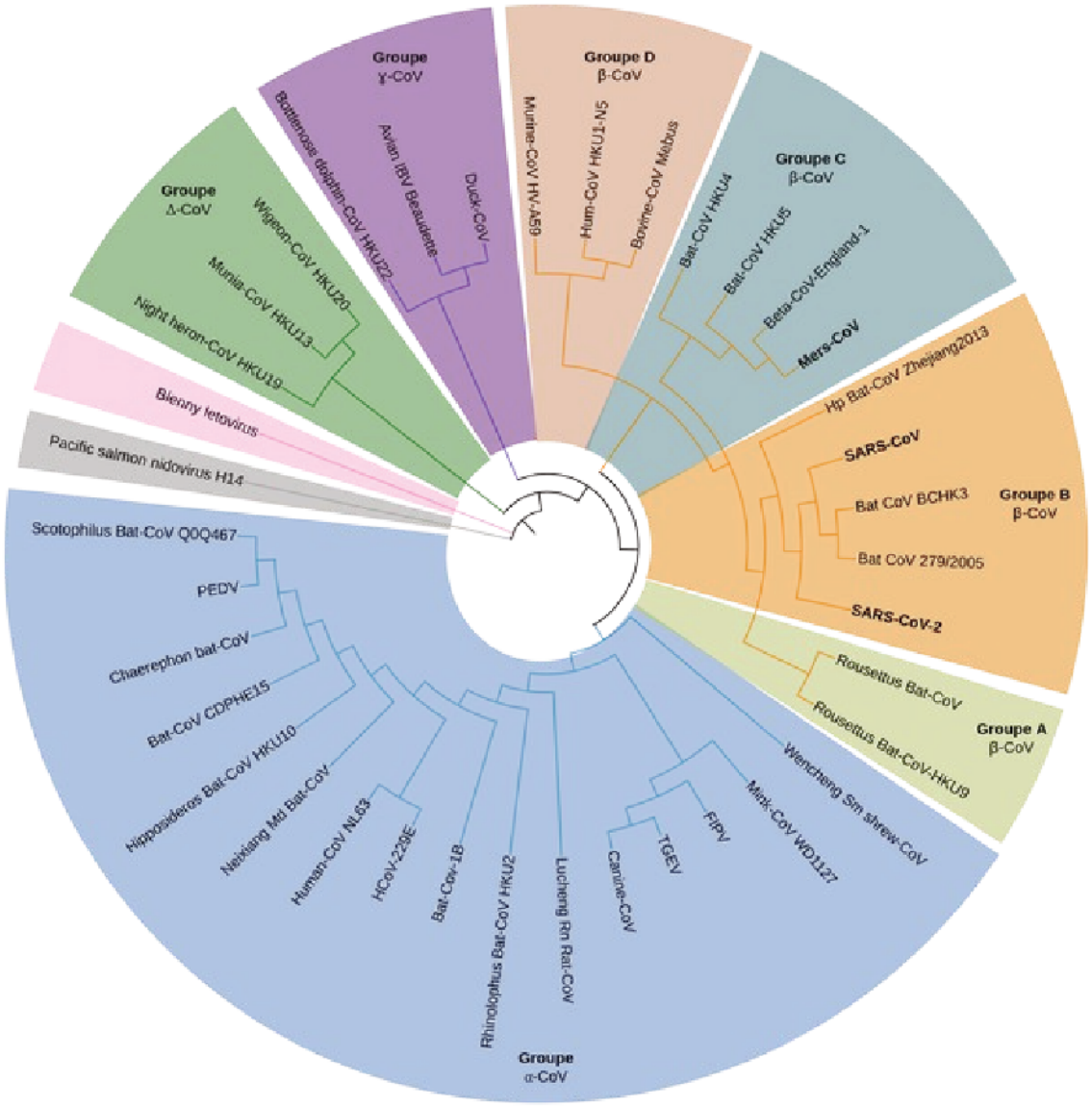


Figure 11 : Arbre phylogénétique de la famille des Coronavirus

Arbre phylogénétique des Coronavirus créer à partir d'alignement de séquence de différentes RdRp. En rose est retrouvé un représentant de la sous-famille des *Letovirinae*, en gris un représentant de la sous-famille des *Pitovirinae*, et les autres taxons représentes différents virus de la sous-famille des *Orthocoronavirinae*. Faite en Colaboration avec Vincent Wilde.

Les Coronavirus ont une très grande taille de génome, notamment grâce à deux protéines. La nsp12 et son activité RdRp, permet la réplication rapide de ces grands génomes. Et la protéine nsp14, avec son activité exonucléase (ExoN), permet la relecture et la réduction du nombre d'erreurs ajoutées à chaque copie du génome. Cette dernière enzyme évite ainsi l'accumulation de mutations délétères et permet le maintien de la stabilité génétique du grand génome des Coronavirus.

1.4.1. Structure des virions

Le dernier tiers du génome des Coronavirus, situé en 3', comporte les gènes codant pour les protéines de structure. Il s'agit des protéines Hémagglutinine-Estérase (HE ; mais qui n'est pas présente chez tous les Coronavirus), Spike (S), Enveloppe (E), Membrane (M) et Nucléocapside (N). Les ORFs de ces protéines suivent toujours le même ordre, mais elles sont séparées par d'autres ORFs putatifs ou démontrés, codant pour des protéines non structurales accessoires, en nombre et en position variables selon les espèces de CoV. Ces ORFs accessoires sont le plus souvent localisées dans les régions situées entre les gènes S et E et aussi les gènes M et N. La fonction d'un grand nombre de protéines non structurales accessoires est encore mal connue et certaines chevauchent les gènes codant pour les protéines de structure (Brant *et al*. 2021).

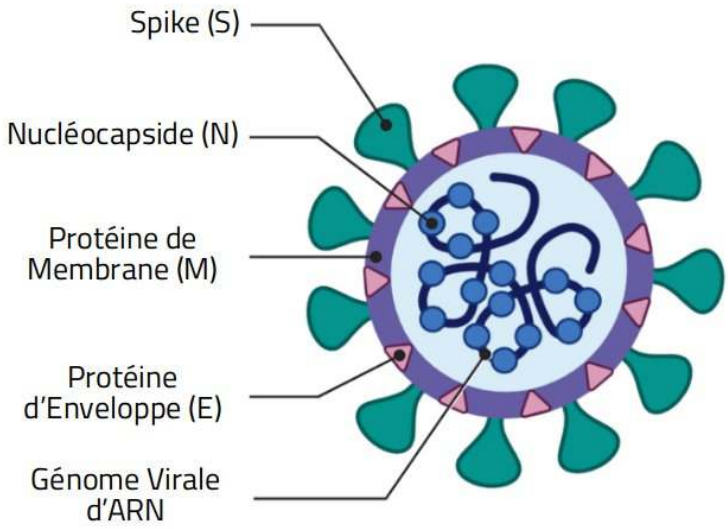


Figure 12 : Structure schéma du virions de SARS-CoV-2

Ce schéma représentant les différentes composantes structurales d'un virion de SARS-CoV-2. La protéine Spike (S) responsable de la liaison avec le récepteur hôte, la protéine de membrane (M) qui donne la forme au virus, et la protéine d'enveloppe (E) impliquée dans l'assemblage et la libération des virions, et la nucléocapside (N) qui encapside le génome viral d'ARN simple brin.

Les virions s'assemblent au niveau de la membrane plasmique où se trouvent les protéines S, E, M. La protéine N, après avoir compacté le génome viral, va se rapprocher des autres protéines structurales en attente, puis un bourgeonnement émergera, créant ainsi un nouveau virion (Voir Figure 12 ; Brant *et al*. 2021). Ces virions sont donc des virus enveloppés et ont un diamètre d'environ 20 nm (Lai et Cavanagh 1997). La protéine S, étant proéminente et formant des « pics » sur la particule virale, donne l'impression d'une tête couronnée lors de son observation en microscopie électronique, ce qui a conduit à l'appellation de Coronavirus de cette famille.

1.4.1.1. Description des Protéines Structurales

(a) La protéine Hémagglutinine-Estérase (HE)

La Hémagglutinine-Estérase (HE) est présente uniquement chez certains *Betacoronavirus* de lignée A (les SARS sont de lignée B et le MERS-CoV de lignée C ; Forni *et al*. 2017). C'est un homodimère relié par un pont disulfure et d'un poids moléculaire total de 120 kDa. Il a été observé que la HE du virus de l'hépatite murine (MHV) a environ 30 % d'homologie de séquence avec la HE du virus de la grippe du groupe C (Luytjes *et al*. 1988). Cela suggère que la HE a été obtenue par recombinaison non homologue de l'ARN et que cet événement aurait permis la séparation de cette lignée (Snijder *et al*. 2003).

Elle forme une seconde pointe à la surface des virions et utilise les acides sialiques comme récepteur. Bien que non essentielle, l'hypothèse sur son rôle est que la HE pourrait améliorer la fixation générale du virus via sa liaison aux acides sialiques (Peng *et al*. 2011).

(b) La protéine Spike (S)

La protéine Spike (S) est la protéine caractéristique qui donne aux virions leur « couronne ». Présente en grand nombre à la surface du virion, elle s'assemble en homotrimères et pèse environ 150 à 200 kDa. Chez certains Coronavirus comme le SARS-CoV-2 (Örd, Faustova, et Loog 2020) et le MERS-CoV (Millet et Whittaker 2014), la protéine S peut être clivée en deux sous-unités (appelées S1 et S2) pendant leur synthèse dans les cellules infectées hôtes. Ces derniers ont en effet un site de clivage spécifique des proprotéines convertases, en l'occurrence la furine, dont le motif peut varier selon les virus et les variants (Örd, Faustova, et Loog 2020). Chez le SARS-CoV-1, cette caractéristique de la protéine S est présente au résidu 797 (Belouzard, Chu, et Whittaker 2009), avec un site de clivage de moindre efficacité. Le clivage de S1/S2 est donc très important pour l'infectiosité de ces virus.

À l'inverse de cette stratégie, pour d'autres Coronavirus, la protéine S n'est clivée que lorsqu'elle atteint la future cellule hôte. Les sous-unités S1 et S2 vont par la suite médier une cascade réactionnelle pour permettre l'entrée du virion dans la cellule (Jackson *et al*. 2022).

(c) La protéine d'Enveloppe (E)

La protéine d'Enveloppe (E) est une petite protéine de 8 à 12 kDa. Abondamment exprimée, elle est impliquée dans la morphogenèse et l'assemblage de la future particule virale (Santos-Mendoza 2023). La protéine E, aidée de la protéine M, est également impliquée dans la libération des particules virales (Neuman *et al*. 2011). La création de l'enveloppe des futures particules virales passe aussi par des interactions

de la protéine E avec la protéine N dans l'assemblage de cette dernière (Ravi, Saxena, et Panda 2022).

Il a été montré que sans cette protéine, le titre viral est extrêmement diminué et les virions produits ne sont plus tous complets (Ravi, Saxena, et Panda 2022).

(d) La protéine de Membrane (M)

La protéine de Membrane (M) est la protéine la plus abondante. D'une taille de 25-30 kDa, elle interagit avec la protéine S, la protéine N avec l'ARNg et la protéine E, aidant ainsi à la formation de l'enveloppe du virion (Neuman *et al*. 2011).

L'interaction des protéines S et M est essentielle pour la rétention de S dans le complexe intermédiaire au niveau du réticulum endoplasmique (RE) puis du Golgi, et ensuite lors de la formation du virion. Concernant la liaison aux protéines N, la protéine M stabilise la nucléocapside en complexe avec son ARNg. La protéine M aide également, avec la protéine E, à la création, la formation et à la libération des particules virales (Fehr et Perlman 2015 ; Ravi, Saxena, et Panda 2022 ; Neuman *et al*. 2011).

(e) La protéine de la Nucléocapside (N)

La protéine de la Nucléocapside (N) est une nucléoprotéine de 43 à 50 kDa qui protège et compacte le génome d'ARN viral sous forme de chapelet. Elle est constituée de deux domaines, chacun utilisant un mécanisme différent pour se replier sur l'ARN (Hurst, Koetzner, et Masters 2009). Différentes études suggèrent qu'il y aurait de 14 à 40 nucléotides par nucléocapside (Neuman *et al*. 2011). La protéine N interagit également avec la protéine M lors de la formation des particules virales (Fehr et Perlman 2015).

De plus, il a également été montré que la protéine N interagit avec la nsp3, l'utilisant comme ancrage lors de la phase de réplication du génome viral dans les vésicules. Cela permettrait ainsi d'aider à localiser et stabiliser l'ARN viral dans les vésicules formées grâce à nsp3, nsp4 et nsp6 (Hurst, Koetzner, et Masters 2009 ; Angelini *et al*. 2013).

1.4.2. Description génomique des Coronavirus

Les Coronavirus possèdent des génomes parmi les plus grands (entre 26 et 32 Kb) chez les virus à ARN (Subissi *et al*. 2014). Ils possèdent la même organisation génomique décrite dans la partie « 1.3.1. Organisation génomique des Nidovirus ».

Cette particularité d'un grand génome d'ARN est due à différents "chemins évolutifs". Alors que d'autres virus à ARN comme celui de la poliomyelite ont évolué pour avoir une polymérase relativement fidèle mais lente pour un génome de 7,5 Kb,

les Coronavirus, eux, ont évolué pour avoir une polymérase rapide mais avec une fidélité de réplication moins importante.

Pourtant, une telle taille de génome nécessite une réplication à la fois fidèle et rapide chez les Coronavirus pour éviter un taux d'erreur trop élevé, qui pourrait être fatal pour le virus (Ferron *et al*. 2021). Face à ces contraintes, les Coronavirus ont acquis dans l'évolution une machinerie de réplication sophistiquée potentiellement composée de 15 à 16 protéines non-structurales (nsp ; Ogando *et al*. 2019). Les protéines impliquées dans la synthèse et l'édition de ces ARN viraux sont codées dans l'ORF1b (E.J. Snijder, Decroly, et Ziebuhr 2016).

Ainsi, comprendre la structure et le rôle de cette machinerie particulière des Coronavirus est essentiel, car elle représente une cible thérapeutique spécifique à ces derniers.

1.4.2.1. Description des ORF1a et ORF1b

Comme pour tous les Nidovirus , les deux tiers du génome du Coronavirus (CoV) sont occupés par deux grands cadres de lecture ouverts (ORF) que sont l'ORF1a et l'ORF1b (Voir Figure 14).

Les gènes de ces ORF vont permettre la synthèse des protéines impliquées dans la réplication virale. Ces ORF sont tous deux traduits à partir du génome viral et se chevauchent sur quelques dizaines de nucléotides. En effet, l'expression d'ORF1b nécessite un décalage du cadre de lecture du ribosomal de -1 base, appelé *frameshifting*, et cela se produit juste en amont du codon de terminaison d'ORF1a (Bhatt *et al*. 2021 ; Atkins *et al*. 2016).

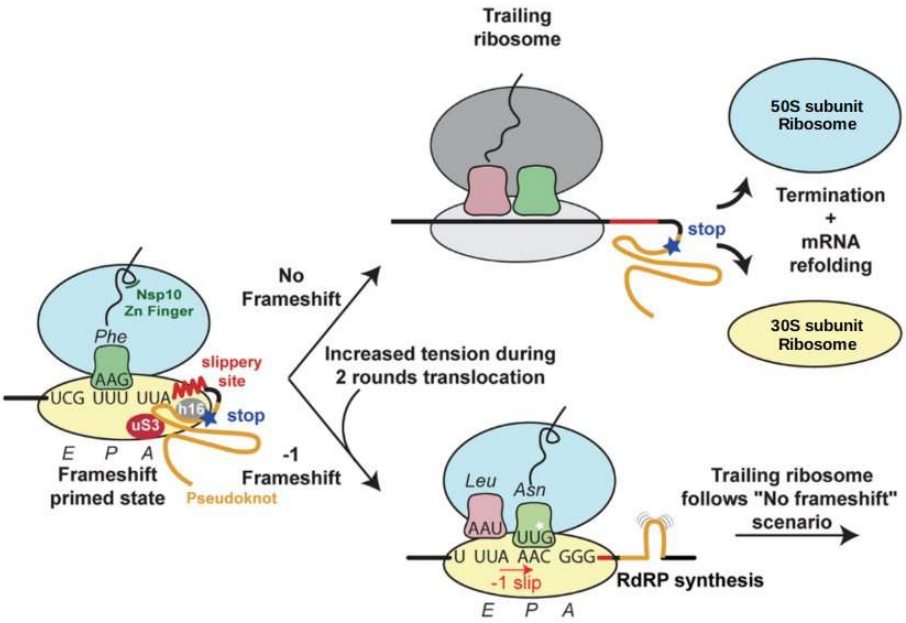


Figure 13 : Description du *frameshifting* chez les Coronavirus

Ce schéma montre le mécanisme de changement de cadre (-1 frameshift ; en bas de la figure) ou l'arrêt de la traduction (en haut de la figure) au niveau d'un site particulier sur l'ARNm viral, en fonction de la structure en pseudo-nœud de l'ARN dû à la tension exercé sur l'ensemble de l'ARN. Ce processus permet la production des protéines virales pp1a ou de pp1ab à partir d'un seul ARN codant pour l'ORF1a ou 1b. Lorsqu'il y a cette tension sur l'ARN qui dénature le pseudo-nœud de l'ARN, le ribosome pourra alors traduire la seconde partie de l'ARN codant pour l'ORF1ab, et donc les protéines correspondant au protéines impliqué dans la réplication du génome virale. Figure adapté de (Bhatt *et al.* 2021).

Cela est dû à des structures d'ARN primaires et secondaires dans cette zone. L'efficacité de cet événement de décalage de cadre, qui peut atteindre 50 % dans le cas des CoV (Irigoyen *et al.* 2016), favorise la traduction des protéines codées par ORF1a par rapport aux protéines codées par ORF1b.

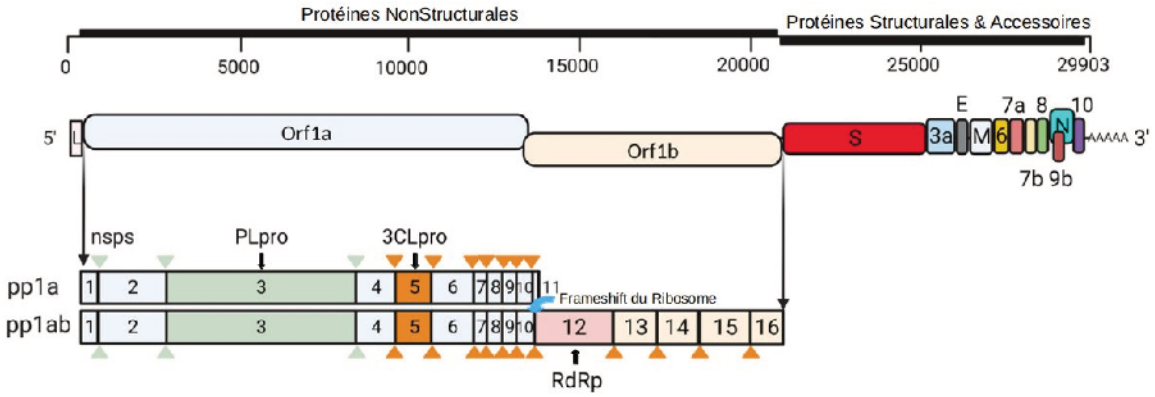


Figure 14 : Description des ORF codant pour les Protéines Non Structurales (NSP).

Ce schéma montre l'organisation des cadres de lecture ouverts (ORF) type des génomes des Coronavirus comme celui du SARS-CoV-2. L'ORF1a et l'ORF1b codent pour les poly-protéines pp1a et pp1ab, qui sont clivées en protéines non structurales (nsp). Le site de changement de cadre (-1 frameshift) permet la traduction de pp1ab, incluant l'ARN polymérase dépendante de l'ARN (RdRp) et protéines partenaires pour la réplication du génome virale. Les protéines structurales et accessoires, comme la protéine Spike (S), l'enveloppe (E), la membrane (M), la nucléocapside (N) et accessoires, sont codées dans les régions en aval. Figure adapté de (Sun, Chan, et Yuan 2023).

Une fois les pp1a et les pp1ab traduites, deux protéases (Hilgenfeld 2014) codées par ORF1a vont commencer à cliver la poly-protéine produite pour séparer les nsp (Voir Figure 14). Ce point a commencé à être abordé dans la partie "1.3.2. Organisation génomique" chez les Nidovirus et va maintenant être approfondi chez les Coronavirus.

La plupart des Coronavirus, y compris le SARS-CoV-2, possèdent deux protéases principales. La première protéase à intervenir lors de la protéolyse de pp1a et pp1ab est une protéase de type papaine (PLpro, pour « papaïn-like protease »), dont l'activité est portée par la grande sous-unité de nsp3 (Angelini *et al*. 2013). Cette dernière intervient sur 3 sites de clivage, de nsp1 au début de nsp4. Les sites de clivage de cette protéase semblent être composés du motif LXGG ou similaire (M. Yang *et al*. 2023).

La seconde, hautement conservée, est une protéase de type 3C-like (3CLpro) dont l'activité est portée par nsp5 et est la protéase principale. Elle est aussi appelée Main protease (Mpro ; Zhang *et al*. 2020). La nsp5 clive de nsp4 à nsp11 pour la pp1a et de nsp4 à nsp16 pour la pp1ab. Cela fait respectivement 7 et 11 sites de clivage conservés dont le motif est Leu-Gln-(Ser/Ala/Gly).

1.4.2.2. Description des Protéines Non-Structurales (NSP)

(a) La protéine nsp1

Elle n'est pas présente chez tous les Coronavirus. Nsp1 intervient au niveau du ribosome et bloque la traduction de l'ARNm de l'hôte, inhibant ainsi les réponses immunitaires innées et facilitant l'infection (Narayanan *et al*. 2015 ; Thoms *et al*. 2020 ; Simeoni *et al*. 2021).

(b) La protéine nsp2

Elle perturbe la liaison des microARN (miARN) au ribosome, diminuant ainsi la traduction des protéines impliquées dans l'immunité cellulaire (Naeli *et al*. 2023).

(c) La protéine nsp3

Elle est composée de deux domaines. Le grand domaine est une protéase de type papaine (PLpro) et clive de la nsp1 au début de nsp4 (Angelini *et al*. 2013). Le second domaine de nsp3, situé en C-terminal, participe avec nsp4 et nsp6 à la production de vésicules à double membrane (DMV) pour la production des futurs virions. Cette partie de nsp3 semble nécessaire à la formation de ces vésicules (Ji *et al*. 2022).

(d) La protéine nsp4

Cette protéine participe avec la nsp3 à la formation des vésicules à double membrane (DMV ; Ji *et al*. 2022). Nsp3 et nsp4 sont donc nécessaires à la formation minimale de DMV (Zimmermann *et al*. 2023).

(e) La protéine nsp5

C'est la seconde et principale protéase des Coronavirus. La nsp5 est hautement conservée, et comme nous l'avons vu précédemment, de type 3C-like (3CLpro), aussi appelée *Main protease* (Mpro ; Zhang *et al*. 2020). Cette protéase clive de nsp4 à nsp11 pour la pp1a et de nsp4 à nsp16 pour la pp1ab (Roe *et al*. 2021).

(f) La protéine nsp6

Elle intervient dans la formation des vésicules à double membrane (DMV) avec nsp3 et nsp4. La nsp6 s'oligomérise pour aider au bourgeonnement des DMV à partir du réticulum endoplasmique (RE), mais permet aussi de sélectionner la composition lipidique et d'exclure les protéines de l'hôte (Ricciardi *et al*. 2022).

(g) La protéine nsp7

Il s'agit d'une des co-enzymes de nsp12 (Shannon *et al*. 2022), mais pourrait aussi intervenir dans la lutte contre l'immunité cellulaire (Deng *et al*. 2023).

(h) La protéine nsp8

C'est une des co-enzymes de nsp12 lors de l'activité ARN polymérase ARN-dépendante (RdRp) de cette dernière. Cependant, le rôle complet de nsp8 n'est pas encore complètement connu. Il a été rapporté qu'elle posséderait une activité adénosine-transférase (Tvarogová *et al*. 2019), qu'elle présenterait des activités d'extension d'amorce (Te Velthuis, Van Den Worm, et Snijder 2012), et qu'après la synthèse de ces courts oligonucléotides d'amorçage (Te Velthuis *et al*. 2010 ; Lehmann *et al*. 2015), elle aurait une activité ARN polymérase ARN-dépendante, non canonique (Imbert *et al*. 2006).

(i) La protéine nsp9

Cette protéine est une protéine de liaison à l'ARN (RBP), est intervient dans deux processus. Le premier permet d'amorcer la synthèse de l'ARN pour initier la synthèse des ARNv après que la nsp9 ait été covalemment NMPylée au niveau de l'asparagine N-terminale conservée. Ce processus est appelé NMPylation de nsp9 est médié par nsp12 et son activité nucléotidyltransférase associée au RdRp des Nidovirus (NiRAN ; Slanina *et al*. 2021 ; Schmidt *et al*. 2023).

Le second processus dans lequel intervient nsp9, est celui de la formation de coiffage de l'ARN, précisément dans les deux premières étape de ce processus. Lors

de ce processus de formation, le domaine NiRAN de la nsp12 et son activité kinase vont transférer l'ARN à l'extrémité N-terminale de nsp9, formant un intermédiaire covalent nsp9-pN-ARN dans un processus appelé RNAylation. Par la suite, le domaine NiRAN le GDP sur le phosphate du nsp9-pN-ARN, formant ainsi la structure de base de la coiffe, le GpppA-ARN (Schmidt *et al*. 2023 ; Park *et al*. 2022).

(j) La protéine nsp10

Nsp10 est une des co-enzymes pour l'activité ExoN de la nsp14, aidant à structurer une partie de cette dernière pour la rendre pleinement fonctionnelle (Ferron *et al*. 2018). Elle joue également un rôle de co-enzyme pour la nsp16, en particulier pour son activité 2'O-MTase (Hausdorff *et al*. 2023).

(k) La protéine nsp11

Le cas particulier de nsp11 est encore mal déterminé. Avec sa longueur de 13 acides aminés, nsp11 pourrait être une protéine désordonnée intrinsèquement (IDP), mais pourrait aussi adopter une conformation en hélice lorsqu'elle est mise en présence de micelles de SDS (Gadhave *et al*. 2021). La région codant pour nsp11 se situe après l'ORF1a, dans la zone où se produit le décalage du cadre de lecture du ribosome (frameshifting) pour continuer la traduction et donner la pp1ab (Bhatt *et al*. 2021). Nsp11 pourrait ainsi jouer un rôle dans les structures secondaires que l'ARN adopte dans cette région pour permettre le changement de cadre de lecture (*frameshifting*) du ribosome.

(l) La protéine nsp12

Nsp12 est la RdRp principale, mais également une enzyme bifonctionnelle. Elle possède l'activité nucléotidyltransférase associée au RdRp des Nidovirus (NiRAN), qui est structurallement liée à la famille des pseudo-kinases (Sreelatha *et al*. 2018). Ce domaine, situé en N-terminal de nsp12, transfère de manière covalente des monophosphates nucléosidiques (NMP) sur ses cofacteurs nsp7, nsp8 et nsp9. Ce processus est appelé NMPylation, mais le rôle de cette activité dans le cycle viral reste encore mal connu (Slanina *et al*. 2021 ; Shannon *et al*. 2021). Il a également été démontré que le NiRAN participe à l'étape de guanylyltransferase lors de la synthèse de la coiffe des ARN (Yan *et al*. 2021). Ce qui est certain, c'est que la mutation du domaine NiRAN est létale pour le virus (Lehmann *et al*. 2015).

L'activité ARN polymérase ARN-dépendante (RdRp) est localisée en C-terminal, et nsp7 et nsp8 en sont ses co-enzymes. Nsp12 est l'enzyme centrale du complexe de réplication et de transcription (RTC ; Prentice *et al*. 2004 ; Te Velthuis *et al*. 2010), et

sera décrite plus en détail dans la partie « 1.5.1. Les ARN polymérases dépendante de l'ARN (RdRp) ».

(m) La protéine nsp13

Elle intervient dans le complexe de réplication et de transcription (RTC) comme RNA hélicase de type I, déroulant l'ARN de 5' vers le 3' et possède par conséquent une activité ATPase (Yue *et al*. 2022). Plus récemment, des études ont suggéré que nsp13 interfère également dans la réponse de l'immunité cellulaire et limite la production d'interféron (IFN ; Feng *et al*. 2023 ; A. Li *et al*. 2023).

(n) La protéine nsp14

C'est une enzyme bifonctionnelle. En N-terminal est portée l'activité exonucléase (ExoN) avec sa co-enzyme nsp10. Cette activité est impliquée dans la correction des erreurs potentielles lors de la polymérisation par le domaine RdRp de nsp12. En C-terminal est portée l'activité N7-guanine méthyltransférase (N7-MTase). Impliquée dans la voie de création de la coiffe des ARN viraux, elle permet la N7-méthylation de cette dernière (Ferron *et al*. 2018 ; Hausdorff *et al*. 2023 ; Park *et al*. 2022).

(o) La protéine nsp15

C'est une endoribonucléase (EndoU) qui limite l'ARN double brin (ARNds) au sein de la cellule hôte afin de diminuer la réponse de l'immunité cellulaire, et joue également un rôle dans la maturation des ARN viraux (Otter *et al*. 2024).

(p) La protéine nsp16

Elle est impliquée dans la 2'-O-méthylation de la coiffe des ARN viraux lors de la formation cette dernière (Park *et al*. 2022). Pour réaliser cette activité, elle a besoin de sa co-enzyme nsp10 (Vithani *et al*. 2021).

1.4.3. La réplication des Coronavirus

1.4.3.1. Le Complexe de Réplication et Transcription (RTC) des Coronavirus

Le Complexe de Réplication et Transcription (RTC) des Coronavirus est composé de plusieurs enzymes. L'enzyme principale est nsp12, qui possède une activité ARN polymérase ARN-dépendante (RdRp), aidée de ses deux co-enzymes, nsp7 et nsp8.

Ce complexe constitue le minimum nécessaire pour assurer l'activité d'extension de la transcription. La nsp12 possède également une activité nucléotidyltransférase associée aux Nidovirus RdRp (NiRAN), située en N-terminal (Prentice *et al*. 2004 ; Te Velthuis *et al*. 2010). Cette partie de nsp12 permet de NMPyler nsp9 de manière covalente au niveau de l'asparagine N-terminale conservée. Ces deux enzymes interviennent bien sûr dans la formation de la coiffe, qui sera développée en « 1.4.4.3 La coiffe des Coronavirus », mais cette modification permet ici d'utiliser nsp9 NMPylé comme amorce pour débuter la synthèse de l'ARN du SARS-CoV-2 (Slanina *et al*. 2021)).

Une fois la polymérase en cours de transcription, le duo nsp14 et sa co-enzyme nsp10 permettent la relecture et la correction des erreurs d'incorporation grâce à l'activité ExoN. En effet, avec l'expansion des génomes des Coronavirus, qui peuvent atteindre des tailles de plus de 30 kb, ces derniers se sont adaptés pour exprimer une activité de relecture et de correction des erreurs d'incorporation à nsp14/10, permettant ainsi à nsp12 d'augmenter, au cours de l'évolution, sa vitesse de transcription (Ferron *et al*. 2021). La nsp13, avec son activité hélicase, facilite la réplication du génome en retirant les structures secondaires et tertiaires de l'ARN servant de matrice (Brant *et al*. 2021).

1.4.4. Le système de coiffage des ARN

1.4.4.1. Les fonctions de coiffe canonique des ARNm eucaryotes

La coiffe $^{7Me}GpppA_{2'\text{-}O\text{-}Me}$ canonique est formée par plusieurs étapes enzymatiques (Voir Figure 15). Si l'on entre dans le détail, les N7-guanine méthylations permettent la reconnaissance de l'ARNm par les complexes de traduction de l'hôte, en particulier par le facteur d'initiation eIF4E (eukaryotic translation initiation factor 4E). Ce dernier se lie à la coiffe via la N7-méthylation de la guanosine (^{7Me}G), ce qui est essentiel pour l'initiation de la traduction en protéines (Quiocho, Hu, et Gershon 2000 ; Walker *et al*. 2021).

En plus de la N7-méthylation, la 2'-O-méthylation du ribose de la première base (ou parfois la deuxième) est également importante. Cette modification semble favoriser la reconnaissance de l'ARN par les ribosomes, en particulier dans des conditions où le facteur eIF4E est limitant (Züst *et al*. 2011). Par ailleurs, ces méthylations empêchent la détection des ARNm par les capteurs de l'immunité innée tels que RIG-I (retinoic acid inducible gene 1 protein) et MDA5 (melanoma-differentiation-associated gene 5), qui sont des protéines responsables de la reconnaissance des ARN viraux dans les cellules (Bouvet *et al*. 2012 ; Walker *et al*. 2021).

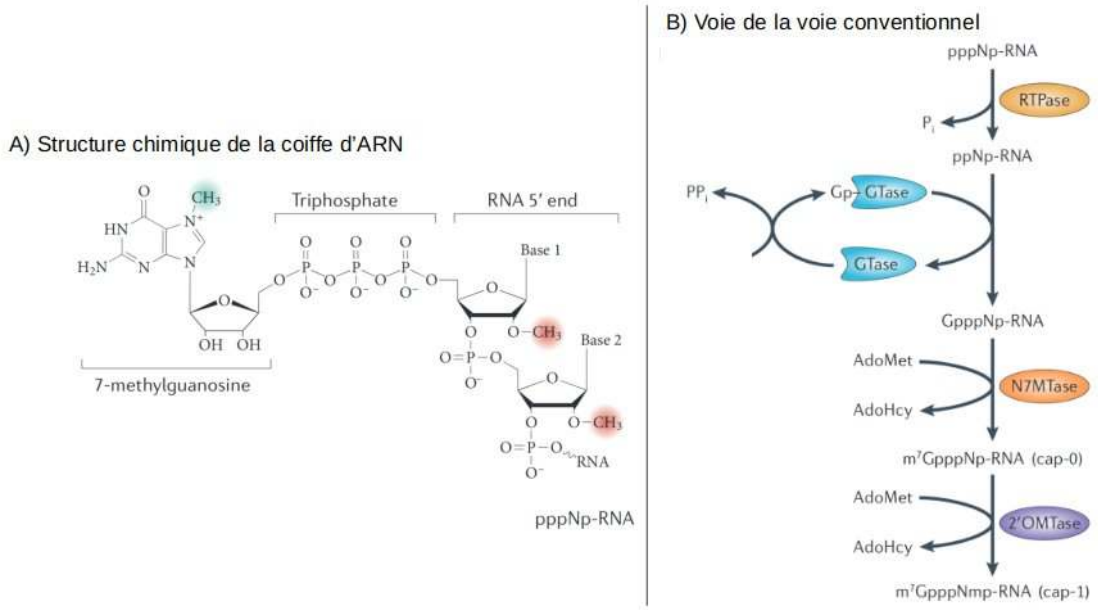


Figure 15 : Schéma de la coiffe canonique de l'ARN

A) Structure chimique de la coiffe en 5' de l'ARN, comprenant une guanosine méthylé en N7 (en bleu) liée par un triphosphate à l'extrémité 5' de l'ARN. Les premiers nucléotide de l'ARN sont aussi marqué, cette fois ci par des 2'O-méthylations. Ces modification protège l'ARN de la dégradation et facilite la traduction. B) Voie de formation de la coiffe par la voie conventionnelle, dite canonique, retrouvé chez les eucaryotes, incluant plusieurs étapes enzymatiques : l'activité RTPase, GTase, N7-MTase, et 2'O-MTase, qui introduisent les modifications d écrites précédemment successivement sur l'extrémité 5' de l'ARN. Figure adapté de (Decroly *et al*. 2012).

1.4.4.2. Les coiffes non-canoniques

En plus de la coiffe dont le processus de formation est classiquement observée chez les eucaryotes, certain virus ont aussi adoptés la coiffe de leurs ARN mais dans un processus de formation différent (Voir Figure 16).

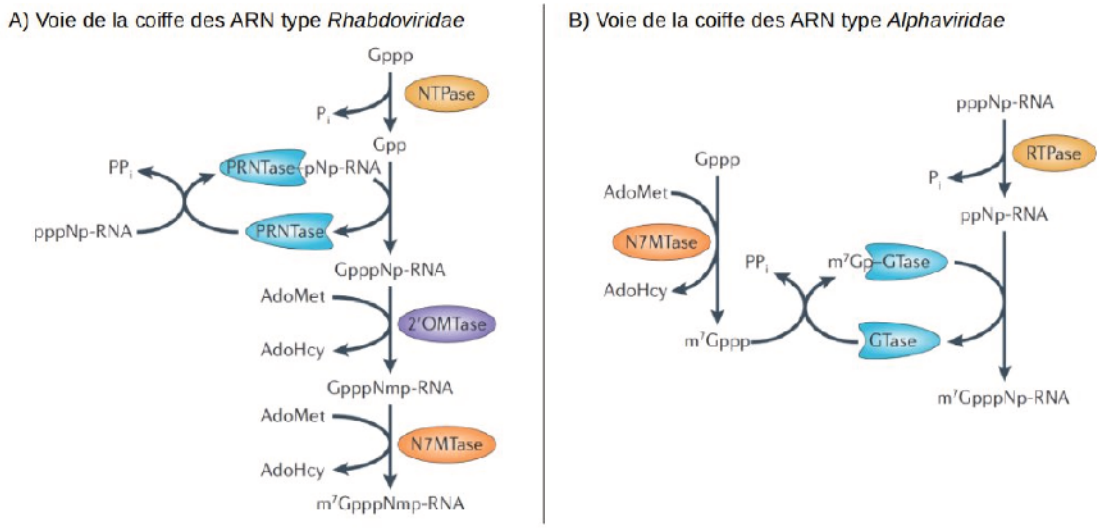


Figure 16 : Schéma d'exemples de voies non canoniques de coiffage de l'ARN

A) Voie de coiffage de l'ARN chez les Rhabdoviridae, où les deux premières étapes de la formation de la coiffe diverge comparé à la formation de la coiffe canonique. Dans cette voie, l'enzyme PRNTase joue un rôle central dans l'ajout de la coiffe sur le Gpp, s'en suivent alors les modifications supplémentaires par les enzymes 2'OMTase et N7MTase connue de la coiffe canonique. B) Voie de coiffage de l'ARN chez les Alphaviridae, illustrant une voie différente impliquant les enzymes RTPase, GTase, et N7-MTase, mais dans un ordre réactionnel différent de celui connue dans la formation de la coiffe canonique, et sans l'étape d'ajout de 2'-O-méthylations. Ces deux exemples de voies montrent la diversité des mécanismes de coiffage de l'ARN entre différents groupes viraux. Figure adapté de (Decroly *et al.* 2012).

Il existe également ce qu'on appelle « Le vol de coiffe (ou cap-snatching) chez certains virus, comme les Arenavirus ou les Orthomyxovirus (dont les virus de la grippe). Dans ce mécanisme, le virus clive les extrémités 5' coiffées des ARNm de l'hôte pour les utiliser comme amorce de synthèse leur propre ARN viral (Walker *et al.* 2021 ; Park *et al.* 2022). Cela leur permet de détourner la machinerie de traduction cellulaire en se servant directement des coiffes d'ARNm cellulaires, évitant ainsi la détection par les défenses immunitaires de l'hôte.

1.4.4.3. La coiffe des Coronavirus

Les mécanismes pour la formation de la coiffe ne sont pas tous entièrement connus pour tous les virus. Celui des *Coronavirus* utilisent leurs propres enzymes pour former

leurs coiffes, dans un processus différent du mécanisme conventionnel (Decroly *et al*. 2012 ; Park *et al*. 2022).

Les ARN du SARS-CoV-2, par exemple, contiennent une coiffe en 5′ ($^{7Me}GpppA_{2'\text{-}O\text{-}Me}$) dont le processus de formation proposé récemment est le suivant (Voir Figure 17) :

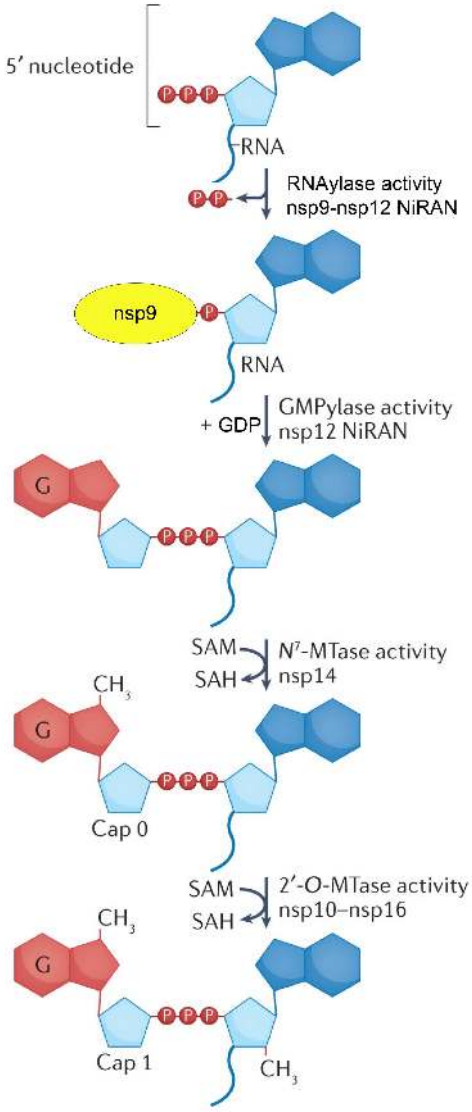


Figure 17 : Schéma de la coiffe chez les Coronavirus

Ce schéma présente les étapes successives de la formation de la coiffe en 5' de l'ARN viral chez les Coronavirus. Les activités enzymatiques incluent la RNAylase (associée à nsp9 et au NiRAN de nsp12), la GMPylase (associée à nsp12 NiRAN), la N7-MTase (nsp14) et la 2'-O-MTase (nsp10-nsp16), qui introduisent les modifications nécessaires pour produire la coiffe Cap-0 puis Cap-1.

La première étape, consiste à retirer deux phosphates de l'extrémité 5′ de l'ARN (pppN-ARN → pN-ARN) grâce au transfert de l'ARN triphosphorylé sur la protéine virale nsp9, et libérant un pyrophosphate (PPi ; Park *et al*. 2022 ; Voir Figure 18). Ce étape nécessite l'intervention du domaine NiRAN de nsp12 qui assure le transfert de

manière covalente de l'ARN sur l'extrémité N-terminale de la protéine virale nsp9. Ce processus est appelé RNAylation de nsp9, et formant l'intermédiaire covalent NSP9-pN-ARN (Yan *et al*. 2022).

La seconde étape nécessite à nouveau l'intervention du domaine NiRAN de la nsp12 pour transférer l'ARN-pN sur un GDP, libérant nsp9 et formant ainsi la structure de base de la coiffe, le GpppA-ARN (Schmidt *et al*. 2023 ; Park *et al*. 2022). Durant cette étape, le GDP incorporé pourrait venir de l'activité GTPase de nsp13 qui transformerait le GTP environnant en GDP (Park *et al*. 2022) bien que cela reste en discussion. Cette étape et processus de guanylylation de la coiffe chez les Coronavirus est ainsi complètement différent du système de coiffe canonique habituellement retrouvé chez les eucaryotes (Ferron *et al*. 2020).

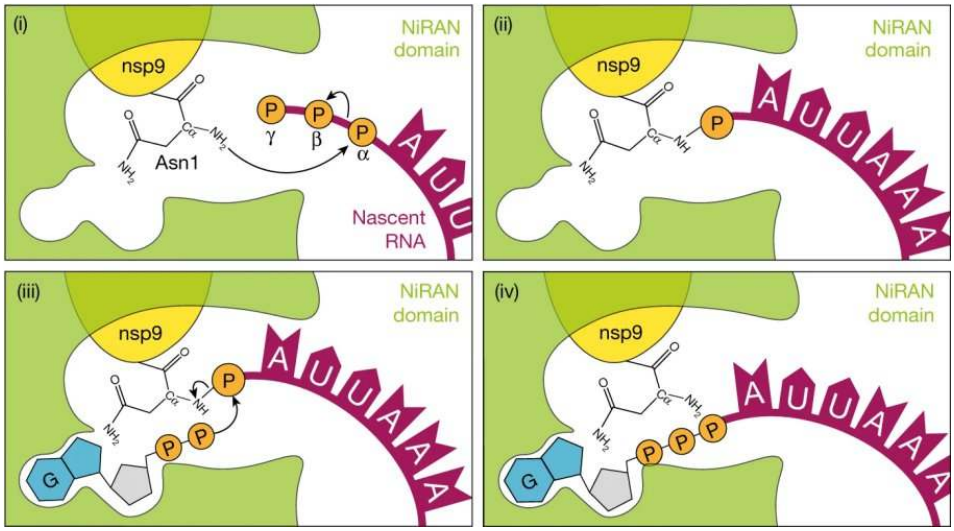


Figure 18 : Schéma de l'étape de RNAylation de nsp9 lors de la seconde étape (GMPlysation) de la formation de la coiffe pour former l'intermédiaire GpppA-ARN

Ce schéma illustre le mécanisme dans lequel nsp9, après avoir sa RNAylation, permet l'addition d'une guanosine di-phosphate (GDP) sur le phosphate de l'extrémité 5' d'un ARN naissant via l'aide du domaine NiRAN de nsp12. Cette seconde étape de la formation de la coiffe 5' de l'ARNm viral. Figure extraite de (Park *et al*. 2022).

Ainsi, ces deux étapes distingues de la formation de la coiffe conventionnel ou de celle décrite précédemment chez d'autres virus. La première étape de RNAylation est distincte de l'activité GMPylase de la seconde étape de la formation de la coiffe, toute deux associées au NiRAN. Cette dernière activité ressemble d'ailleurs davantage à une activité polyribonucleotidyltransferase (PRNTase) comme décrit chez certains virus à ARN à sens négatif de l'ordre des *Mononegavirales* (Ogino et Banerjee 2007 ; Ogino et Green 2019).

La troisième étape permet le premier type de méthylation de la future coiffe, la m7-guanine méthylation. Médiée par le domaine C-terminal de la nsp14 et son activité N7-

méthyltransférase (N7-Met), cette dernière permet le transfert d'un groupement méthyl de son cofacteur, la S-adénosylméthionine (SAM), en position 7 sur la base guanine du GpppA-ARN. Après transfert, le cofacteur libéré sera la S-adénosyl-L-homocystéine (SAH) et le produit 7MeGpppA-ARN, aussi appelé Cap-0, permettra de passer à la dernière étape (Ferron *et al*. 2018 ; Hausdorff *et al*. 2023 ; Park *et al*. 2022).

La quatrième et dernière étape consiste à méthyler en 2′-O le ribose des deux premières bases de l'ARN du 7MeGpppA-ARN. Cette étape, qui se renouvellera deux fois, fait intervenir la 2′-O-méthyltransférase nsp16 avec sa co-enzyme nsp10. Cette réaction fera également intervenir le cofacteur SAM qui sera alors transformé en SAH, permettant ainsi la formation de la coiffe finale 7MeGpppA2′-O-Me, aussi appelée Cap-1 (Park *et al*. 2022 ; Vithani *et al*. 2021).

L'aspect structurale des deux dernières enzymes de cette voie sera développé dans les parties « 1.7.1. La N7-méthyltransférase nsp14 des Coronavirus » et « 1.7.2. La 2'-O-méthyltransférase nsp16 des Coronavirus ».

1.5. Les polymérases virales des virus à ARN(+)

Nous avons vu que la réplication du génome viral est un processus complexe impliquant plusieurs facteurs viraux et cellulaires de l'hôte. Il a précédemment été abordé les différentes étapes dans le processus de la réplication du virus dont l'une des étapes clé et la réplication du génome virale.

Ainsi, la majorité des virus à ARN codent pour une ARN polymérase dépendante de l'ARN (RdRp), à l'exception des Retroviridae, comme le VIH, qui utilise sa transcriptase inverse (Fanales-Belasio *et al*. 2010).

Dans les parties suivante, nous allons discuter de la structure générale des RdRP, avant de nous attarder sur la composition du Complexe de Transcription et Réplication (RTC) des Coronavirus.

1.5.1. Les ARN polymérases dépendantes de l'ARN (RdRp)

Les ARN polymérases ARN-dépendantes (RdRp) jouent un rôle crucial et sont très conservées parmi les virus à ARN, bien que leur structure spécifique puisse varier selon les familles virales et être associée à d'autres domaines enzymatiques en fonction des virus.

La RdRp adopte généralement une forme de "main droite fermée", structurée en sous-domaines nommés "paume", "pouce" et "doigts" (Shannon *et al*. 2022). Certains RdRp possèdent également des "doigts de zinc" pour stabiliser leur structure. Cette organisation crée deux canaux convergeant vers le site actif de l'enzyme : le canal principal guide la matrice d'ARN, tandis que le canal secondaire permet l'ajout des ribonucléotides triphosphates (NTP). Lors de la catalyse, le groupe 3'-hydroxyle du brin naissant d'ARN attaque le α-phosphate du NTP, libérant une molécule de pyrophosphate (PPi), un processus facilité par deux ions Mg^{2+} coordonnés par les phosphates du NTP et des résidus aspartate conservés (Černý *et al*. 2014).

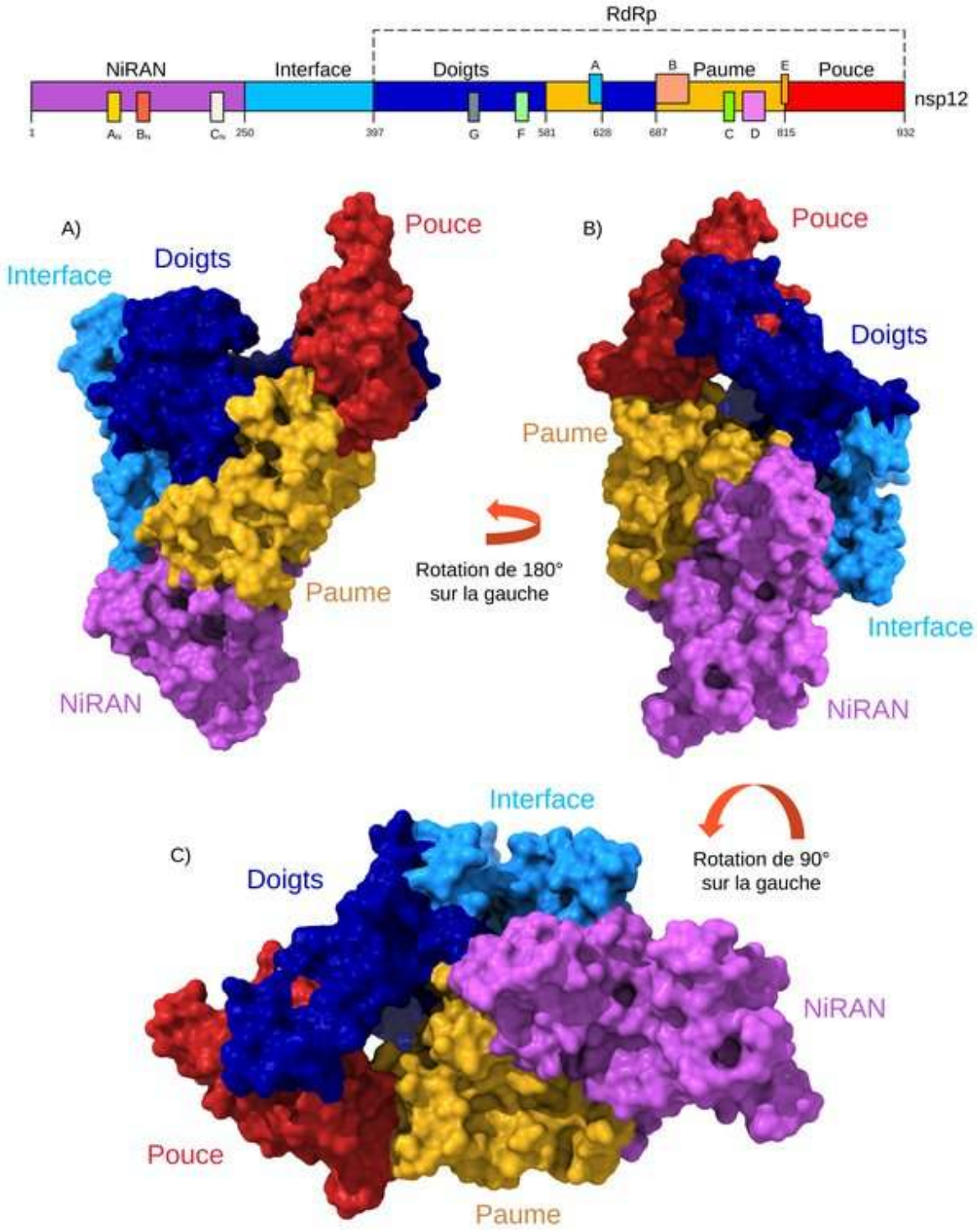


Figure 19 : Schéma des différents sous domaines de nsp12

Cette figure représente les différents sous domaines de nsp12 (PDB : 7ED5). Sur le panel du haut, les différents domaines sont représenté schématiquement sur la séquence. Sur le panel du bas, une vue globale de la structure avec les sous-domaines colorés est présenté : le domaine NiRAN en violet, l'interface en cyan, les "doigts" en bleu, la "paume" en jaune et le "pouce" en rouge. A) la position de la structure est présenté de face, la « main droite » vers soi. Une rotation de 180° vers la gauche (B), puis la vue latérale obtenue après une rotation de 90° sur la gauche (C), permet de mieux visualiser

l’organisation spatiale et l’entrée des nucléotide dans la poche catalytiques au centre (paume, doigts, pouce). L’agencement du domaine NiRAN, ainsi que l’interface sont également visible sur cette vue.

Les RdRp partagent également plusieurs motifs de séquences aminées conservées, cruciaux pour leur fonction. Ces motifs, désignés de A à G, sont définis à la fois structurallement et fonctionnellement. Des mutations dans ces motifs sont liées à l'évolution du génome viral et à des variations des taux de mutation. Chez les Coronavirus, les motifs A, B, C, D et F sont particulièrement conservés, cette conservation étant partagée avec les familles *Picornaviridae* et *Flaviviridae* (Campagnola *et al*. 2015).

Les motifs A et C, situés au cœur du site actif, encadrent la réaction catalytique où un nouveau nucléotide est ajouté au brin d’ARN en formation. Les résidus aspartate conservés, mentionnés précédemment, appartiennent à ces motifs. Le motif A est composé d’acides aminés DxxxxD, tandis que le motif C contient les acides aminés S/GDD (Shannon *et al*. 2020).

Le motif B une grande boucle considérée comme un autre de point de contrôle pour la fidélité des RdRp. La sérine de ce motif, via la liaison hydrogène au groupe 2'-OH du ribose incorporé, améliore la fidélité de la polymérase, contrairement aux RdRp où cette sérine est remplacée par une méthionine (Shannon *et al*. 2020 ; (Selisko *et al*. 2018).

Le motif F joue également un rôle dans la fidélité et forme le "plafond" du tunnel par là où les nucléotides entrent pour atteindre le site actif. Ce motif positionne les groupes phosphates du NTP entrant et du pyrophosphate sortant via des interactions avec des résidus lysine et arginine (Selisko *et al*. 2018 ; Yang *et al*. 2017).

A l’entrée du tunnel de la matrice d’ARN, se trouve le motif G. Proche du motif F, il pourrait aider au repliement de de ce dernier, mais aussi influencer l'orientation des résidus de la matrices en - 2 et - 1 et ainsi altérer la reconnaissance NTP, influençant ainsi la fidélité globale de la RdRp (Griesemer *et al*. 2017 ; Meng et Kwang 2014).

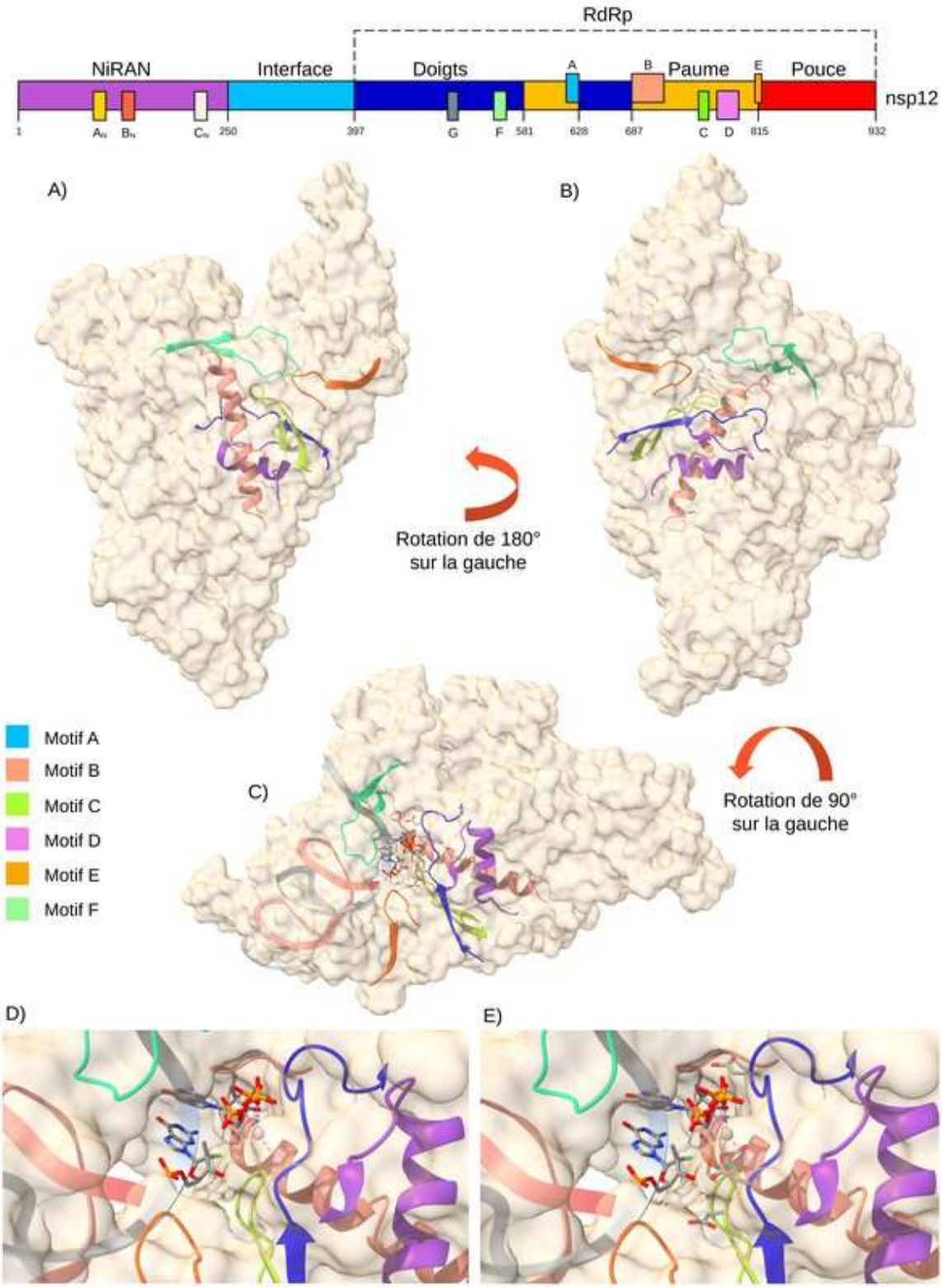


Figure 20 : Schéma des motifs du site catalytique de la RdRp de nsp12

Sur le panel du haut sont présenté sur la séquence, les différents motifs du site catalytique de la RdRp de nsp12 du SARS-CoV-2. Sur le panel du bas, une vue globale des motifs colorés, présenté sur la structure PDB : 7ED5. Plusieurs vues de l'enzyme sont représenté : A) Vue de face de la « main droite » C) Vue latérale de l'entrée de des nucléotides dans le site catalytique et de ses principaux motifs impliqués dans la liaison du nucléotide et la catalyse de l'incorporation d'ARN : motif A (bleu), motif B

(saumon), motif C (vert), motif D (violet), motif E (orange) et motif F (vert menthe). En D) un zoom sur le site actif est réalisé avec l'analogue de nucléotidique (AT9010) lié à l'ARN en cours de synthèse, et un autre prêt à être incorporé.Tous deux montrant les interactions entre les motifs et le substrat. En E) les acides aminés impliqué dans la réaction sont exposés.

Proche du motif A et du site catalytique, se trouve également le motif D. Au cours de la réaction de catalyse, le 3'-OH de l'ARN attaque le NTP et transfert des protons vers le pyrophosphate sortant doit se produire pour assurer la neutralité globale de la réaction. Un acide aminé basique, le plus souvent une lysine, est impliqué dans ce transfert de protons en plus de servir à extraire vers l'extérieur le pyrophosphate. D'autres études proposent que la boucle du motif D avec une autre lysine se déplace pendant la catalyse, permettant ce mouvement de bascule pour extraire le pyrophosphate. Ainsi en neutralisant et facilitant l'extraction du pyrophosphate, le motif D est directement impliqué dans la vitesse globale de polymérisation de la RdRp. La composition et la structure de ce motif permettrait un taux augmentation de 50 fois à 2000 fois la vitesse de polymérisation, malgré qu'il ne soit pas le motif le plus conservé de tous (Castro *et al*. 2009).

Le motif E, en forme d'épingle à cheveux, est situé entre les sous-domaines "paume" et "pouce". Il est essentiel pour l'initiation *de novo* et pour l'élongation, notamment en alignant correctement le 3'-OH du nucléotide d'amorçage (Hastie *et al*. 2011 ; Selisko *et al*. 2018).

Il est généralement observé que les virus à ARN ont un taux de mutation élevé, ce qui a conduit à l'évolution de mécanismes de relecture chez ces virus. Certaines RdRp peuvent corriger elles-mêmes les mésappariements, interrompe l'élongation en remontant le long de la matrice d'ARN et en clivant les derniers nucléotides incorporés avant de reprendre la polymérisation, ce qui ralentit le processus. Chez les Coronavirus et certains autres virus, cette fonction de relecture a été déléguée à une exonucléase 3'-5' (NSP14) pour corriger les erreurs pendant la synthèse de l'ARN et qui est hautement conservée dans la famille des Coronavirus (Hastie *et al*. 2011). La NSP14 a été déterminante dans l'évolution des grands génomes d'ARN comme ceux des CoV en prévenant l'excès de mutations létales et permettant à la RdRp de se spécialiser graduellement dans la vitesse de synthèse (Dolja et Koonin 2018 ; Ferron *et al*. 2021).

Les Coronavirus ont donc évolué au cours du temps, en se spécialisant et adaptant, développant ainsi des caractéristiques uniques a ces virus, notamment dans leur complexe de réplication et de transcription (RTC). Plusieurs protéines virales composent le RTC, et leurs fonctions individuel ont été abordées dans le chapitre « 1.3.3.1 Le complexe de réplication et de transcription (RTC) des Coronavirus ». Le chapitre suivant explorera l'organisation structurale des protéines du RTC.

1.5.2. L'ARN polymérase des CoVs et son complexe de réplication

Le complexe de réplication et de transcription (RTC) est organisé autour de la nsp12, qui porte l'activité principale du complexe, l'ARN polymérase ARN-dépendante (RdRp). Cette dernière catalyse la synthèse de l'ARN en s'appuyant sur les cofacteurs nsp7 et nsp8. Le système de réplication à l'échelle cellulaire a été abordé dans la partie "1.3.2. Cycle de vie" et nous allons maintenant nous attarder sur la conformation adoptée par ces trois protéines avant de développer les interactions possibles des autres partenaires du RTC, à savoir nsp9, nsp10, nsp14 et nsp13.

Ainsi, le domaine RdRp de la nsp12 est accompagné de ces cofacteurs nsp7 et nsp8, formant le complexe central du RTC, et le plus étudié. Deux nsp8 sont nécessaires pour servir de glissière à l'ARN double brin sortant. Nsp7 est essentiel à la formation d'un complexe RdRp actif en interagissant avec l'une des nsp8, ce qui contribue à la bonne conformation de nsp12 (voir Figure 19). Des expériences de mutagenèse dirigée ont permis d'identifier plusieurs résidus clés chez nsp8 (K58A, D99A, P116A, P183A et R190A) pour la formation du complexe nsp12/nsp7/nsp8 (Subissi *et al*. 2014). D'autres résidus sur nsp7 (C8G et V11A) semblent également nécessaires au pré-assemblage de nsp7/nsp8 sur la grande zone d'interactions qu'ils ont en commun (Biswal *et al*. 2021). De même, trois résidus pour nsp7 (K7, H36 et N37) et un résidu nsp8 (K58) semblent être importants pour l'interaction du RTC avec l'ARN (Subissi *et al*. 2014).

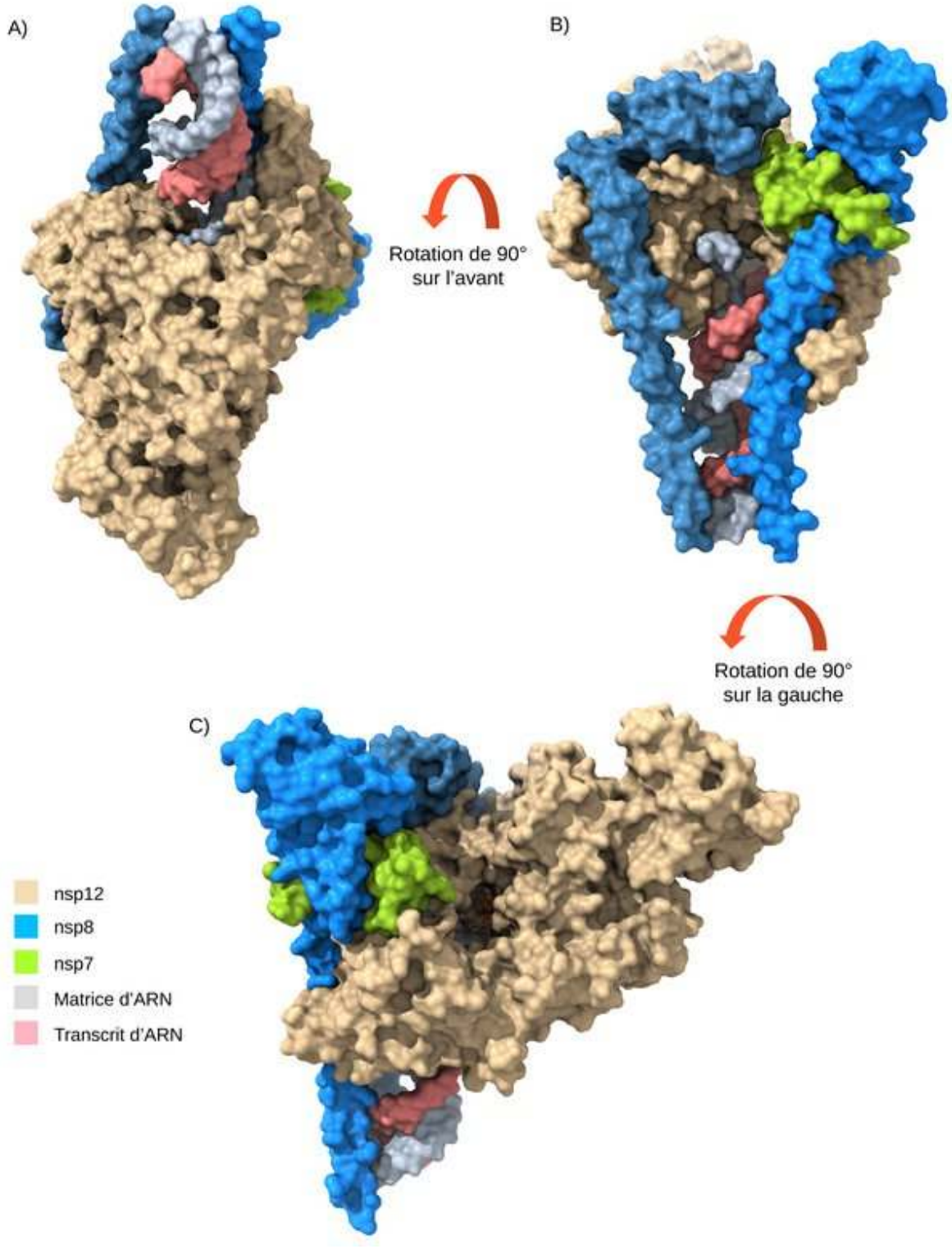


Figure 21 : Structure du Complexe de réplication et transcription (RTC) nsp12/7/8

Présentation de la structure (PDB : 7ED5) du Complexe de Réplication et de Transcription (RTC) formé par nsp12, nsp7 et nsp8 du SARS-CoV-2. En A) Vue globale du complexe RTC, toujours « de face » montrant les sous-unités : nsp12 (en beige), les deux nsp8 (en bleu), nsp7 (en vert), la matrice d'ARN (en gris) et le neo brin d'ARN (en rose). En B) La même structure après une rotation de 90° vers l'avant,

offrant une vue plus détaillée du positionnement des sous-unités par rapport au canal d'entrée de l'ARN, où la matrice et le transcrit d'ARN sont visibles descendant le long des hélices-α des deux nsp8. En C) Vue latérale après une rotation de 90° vers la gauche, mettant en évidence à nouveau le site actif du domaine RdRp de nsp12, avec une nucléoside analogue (AT9010) à l'intérieur. Cette vue permet également de mieux apercevoir la liaison entre nsp7 et nsp8.

En poursuivant l'assemblage du RTC, cette fois du côté du domaine Nidovirus RdRp-associated nucleotidyltransferase (NiRAN) de la nsp12, on peut ajouter la protéine nsp9. Cette petite protéine interviendrait dans le RTC en se fixant près d'un site catalytique du NiRAN (Small *et al*. 2023).

Plusieurs structures du complexe nsp12/7/8/9 existent, notamment certaines qui montrent les étapes intermédiaires de NMPylation par NiRAN de nsp9 (PDB : 8SQ9), et également lors de la RNAylation de nsp9 toujours par le NiRAN lors de la formation de la coiffe (PDB : 8SQK, 8SQJ ; Small *et al*. 2023 ; Voir Figure 22 A et B).

On peut également l'observer dans des structures du RTC au complet (PDB : 7EGQ et 7CYQ) dans un contexte général de réplication et les premières étapes de la formation de la coiffe (Voir Figure 23 ; Yan, Yang, *et al*. 2021).

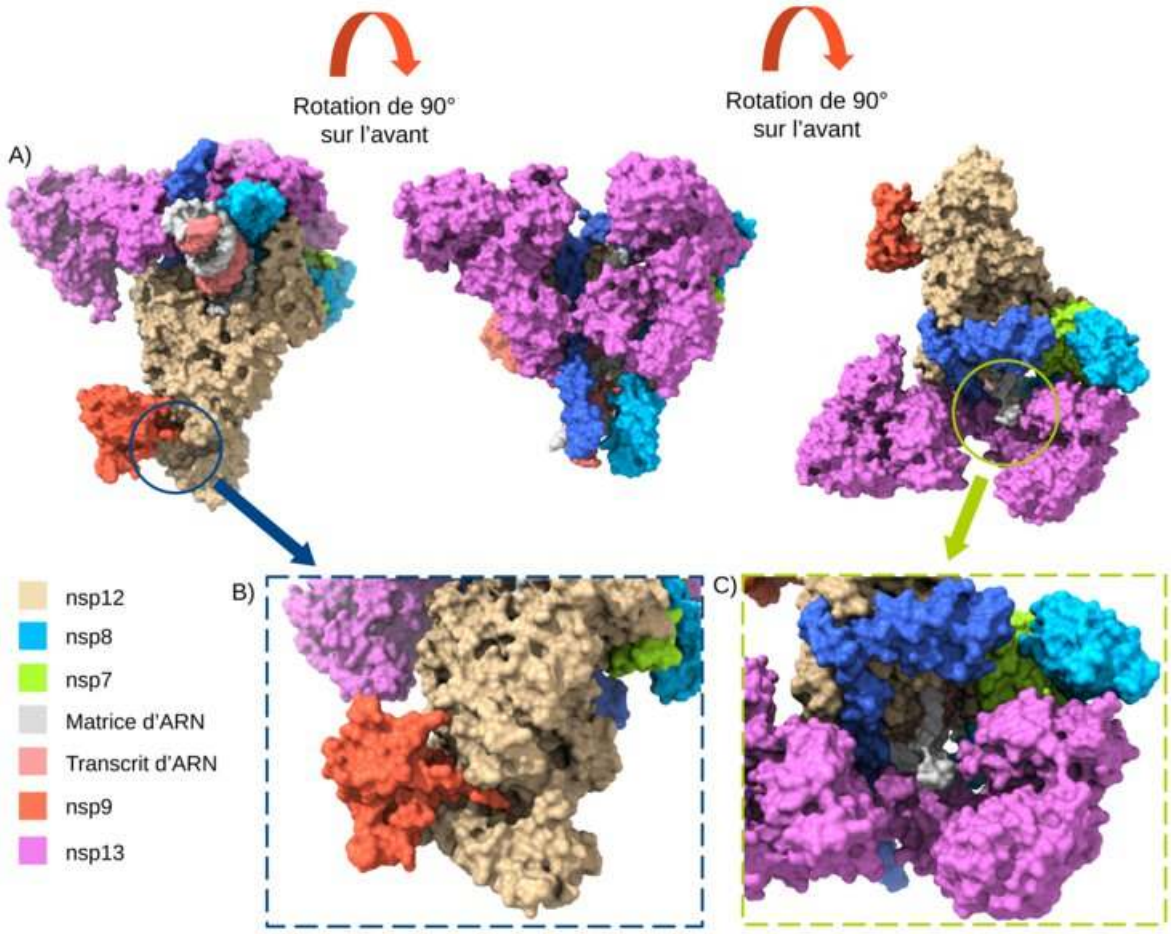


Figure 22 : Structure du Complexe de réplication et transcription (RTC) incluant nsp13 et nsp9

A) Vue générale du complexe RTC (PDB : 7EGQ) avec ses protéines partenaires en vue de face : nsp12 (en beige), les deux nsp8 (en bleu), nsp7 (en vert), nsp9 (en rouge), nsp13 (en violet), la matrice d'ARN (en gris) et le transcrit d'ARN (en rose). La structure est présentée avec deux rotations successives vers l'avant de 90°, exposant différentes perspectives du complexe et l'emplacement des sous-unités nsp13 et nsp9. En B) un zoom sur l'interaction entre nsp9 (en rouge) et le domaine NiRAN de nsp12 (en beige). En C) un autre zoom sur la région où nsp13 (en violet) interagit avec nsp12 et l'ARN, mettant d'entre voir ici la conformation de nsp13 et de son l'activité hélicase pour le déroulement de l'ARN viral pendant la réplication et la transcription.

La protéine nsp13 est l'hélicase du complexe, avec son activité d'ATPase, elle utilise l'énergie dérivée de l'hydrolyse de l'ATP pour créer un mouvement directionnel le long de l'ARN. Elle interagit directement avec nsp12 et facilite l'accès de cette dernière au brin matrice en favorisant le dépliement des structures secondaires en double brin de l'ARN précédemment synthétisé (Voir Figure 22 C ; Yazdi *et al*. 2022). Nsp13 a ainsi un rôle direct dans la régulation de la vitesse de la synthèse de l'ARN. Une structure en Cryo-EM (comme PDB : 7CYQ) permet d'observer l'association de

deux nsp13 sur la partie supérieure de la RdRp, entre « le pouce » et « les doigts » de cette dernière, après les deux nsp8 (Voir Figure 22).

En cas de mésappariements et d'arrêts de la transcription par la RdRp, la matrice va être tirée en arrière via la nsp13, phénomène appelé *backtracking*, et ainsi permettre l'intervention de l'ExoN pour l'excision.

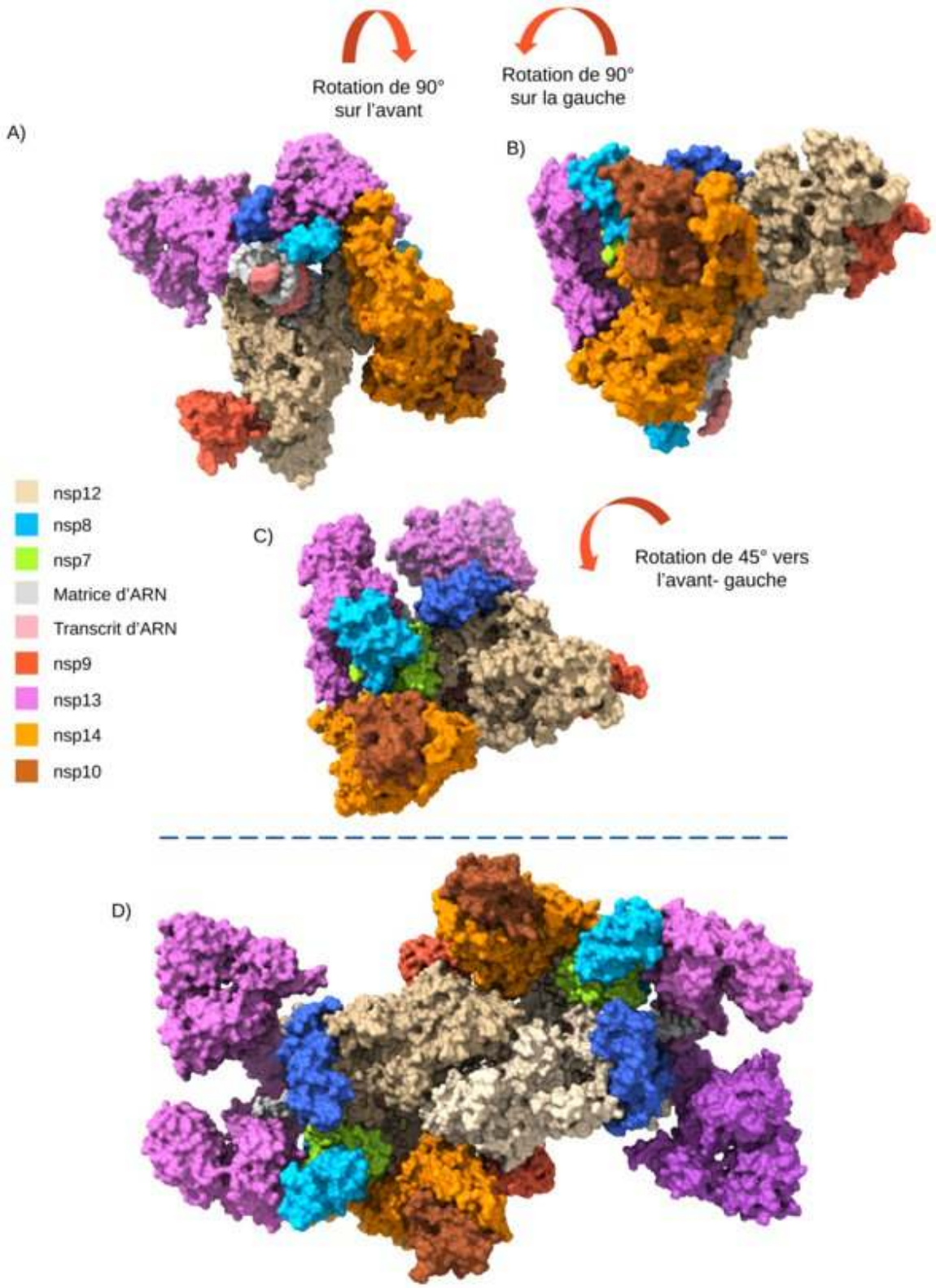


Figure 23 : Structure hypothètique du Complexe de réplication et transcription (RTC) complet, formant un mono ou di-mère

Présentation d'hypothèse d'assemblage d'un monomère de RTC à partir de la structure d'un dimère de RTC (PDB : 7EGQ). A) Vue générale du RTC incluant nsp12 (en beige), les deux nsp8 (en bleu), nsp7

(en vert), nsp9 (en rouge), nsp13 (en violet), nsp14 (en orange), nsp10 (en maron), la matrice d'ARN (en gris), et le transcrit d'ARN (en rose). La nsp14, en association avec nsp10, est impliquée dans l'excision des mésappariements via son activité exonucléase, pourrait se positionner près du site actif de la RdRp dans les figures A) B) C). D) Représentation un dimère possible du RTC (PDB : 7EGQ), une configuration potentiel pour stabiliser l'interaction entre nsp14/nsp10 et les autres composants du complexe. Cette formation permettrait également au domaine méthyltransférase de nsp14, d'intervenir plus facilement dans la maturation de la coiffe 5' de l'ARN viral, étant à proximité de nsp9 et du domaine NiRAN de nsp12, bien qu'aucunes données détaillant d'un telle complexe (avec aussi nsp16) n'ait encore été rapportée.

La dernière paire d'enzymes que l'on peut ajouter au RTC est le couple nsp14/nsp10. Impliqué dans l'excision des mésappariements et les N7-méthylations, selon la structure cryo-EM d'un dimère du RTC (PDB : 7EGQ, présenté en Figure 23), nsp14, stabilisé par la nsp10, se situerait au niveau du « pouce » de la RdRp et de nsp8, ou proche du domaine NiRAN selon comment est scindée la structure.

Cette configuration de nsp14/nsp10 serait possible grâce à la formation d'un dimère du RTC, mais il est difficile de savoir si ce type de structure est réellement possible *in vivo*. Cette structure soulève des questions sur l'emplacement possible de nsp14/nsp10 « en attente » lors de mésappariements (voir Figure 23 A, B, C ; Yan, Yang, *et al*. 2021), car en y regardant de plus près, on aperçoit que cette configuration bloquerait l'entrée de l'ARN dans le site ExoN, comme cela a déjà été observé dans une autre structure (PDB : 7N0B), ce qui a peu de sens au niveau biologique (Imprachim, Yosaatmadja, et Newman 2023). Cette structure en dimère du RTC reste cependant intéressante pour la partie méthylation de l'ARN dans le cadre de la formation de la coiffe. En effet, la proximité du domaine MTase de nsp14 avec nsp9 est plus pertinente dans cette structure pour un processus global comprenant également l'activité de formation de la coiffe en plus de celle de la réplication (Imprachim, Yosaatmadja, et Newman 2023).

Concernant le domaine méthyltransférase de nsp14 qui participe à la maturation de la coiffe 5' de l'ARN viral avec nsp9, le NiRAN et nsp16, aucune donnée structurale de la formation d'un complexe entre ces quatre protéines et/ou en coordination avec d'autres protéines n'a été rapportée dans la littérature. La structure (voir Figure 23 D) pourrait apporter certains éléments de réponse, notamment dans sa forme de dimère (PDB : 7EGQ ; Yan, Yang, *et al*. 2021).

Mais, comme précédemment, bien que le domaine N7-méthyltransférase soit proche de nsp9 et du domaine NiRAN comme « en attente », aucune information ou donnée structurale sur la position de ces protéines lors de leur phase active n'est disponible.

Pour conclure sur cette partie, l'étude des mécanismes de réplication conservés des RdRp *via* les techniques de biologie structurale comme la cristallographie aux rayons X et la cryo-EM est cruciale pour la compréhension globale de ces enzymes, mais également pour le développement de nouveaux inhibiteurs spécifiques.

L'absence d'homologues chez les mammifères en fait des cibles de choix pour le développement de nouveaux antiviraux. Cependant, le taux de mutations généralement élevé attribué aux virus à ARN leur offre une grande plasticité génétique, facilitant l'émergence de variants résistants aux antiviraux (Bordería, Rozen-Gagnon, et Vignuzzi 2015).

Cette émergence de résistances aux antiviraux pourrait être limitée par l'utilisation à grande échelle d'une approche combinatoire de molécules ciblant plusieurs étapes du cycle viral. La présence d'une ExoN chez les Coronavirus (Snijder *et al*. 2003) et les Arenavirus (Hastie *et al*. 2011) complique malgré tout la production de NAs efficaces. Cette dernière doit donc également être ciblée dans notre recherche d'antiviraux.

1.6. Les exoribonucléases virales

Les exoribonucléases virales sont peu fréquentes dans le monde des virus à ARN. Seules deux familles/ordres (les Arenavirus (Hastie *et al*. 2011) et les Nidovirus (Snijder *et al*. 2003)) ont été identifiées comme en possédant.

1.6.1. La nsp14 des Coronavirus

La nsp14 est une enzyme bifonctionnelle dont les deux domaines (ExoN et MTase), sont localisés a ses extrémités N-terminale et C-terminale respectivement. L'ensemble de nsp14 peut être distingué en quatre régions (présenté en Figure 24) :

1) La région N-terminale flexible qui est en contact avec son cofacteur nsp10. Nécessaire à l'activité Exonucléase (ExoN), nsp10 va permettre de structurer la région N-ter de nsp14, permettant ainsi à la poche catalytique de nsp14 d'adopter la conformation active, ainsi que la stabilisation de l'ARNds et de son nucléotide à exciser (en orange sur la Figure 24).

2) En suivant, le domaine ExoN a proprement parler. Cette ExoN appartient à la superfamille des DEDD (sous-famille DED/EDh ; Zuo 2001) dont quatre acides aminés sont conservés (Asp-Glu-Asp-Asp ; Bouvet, Imbert, *et al*. 2012).

3) Les deux domaines sont ensuite séparés par une région charnière, formée d'une boucle et de trois brins, apportant une grande flexibilité de part et d'autre à cette dernière.

4) La dernière partie est le domaine N7-méthyltransférase (N7-MTase) sur lequel nous reviendrons dans la partie « 1.7.1 La N7-méthyltransferase nsp14 des Coronavirus»

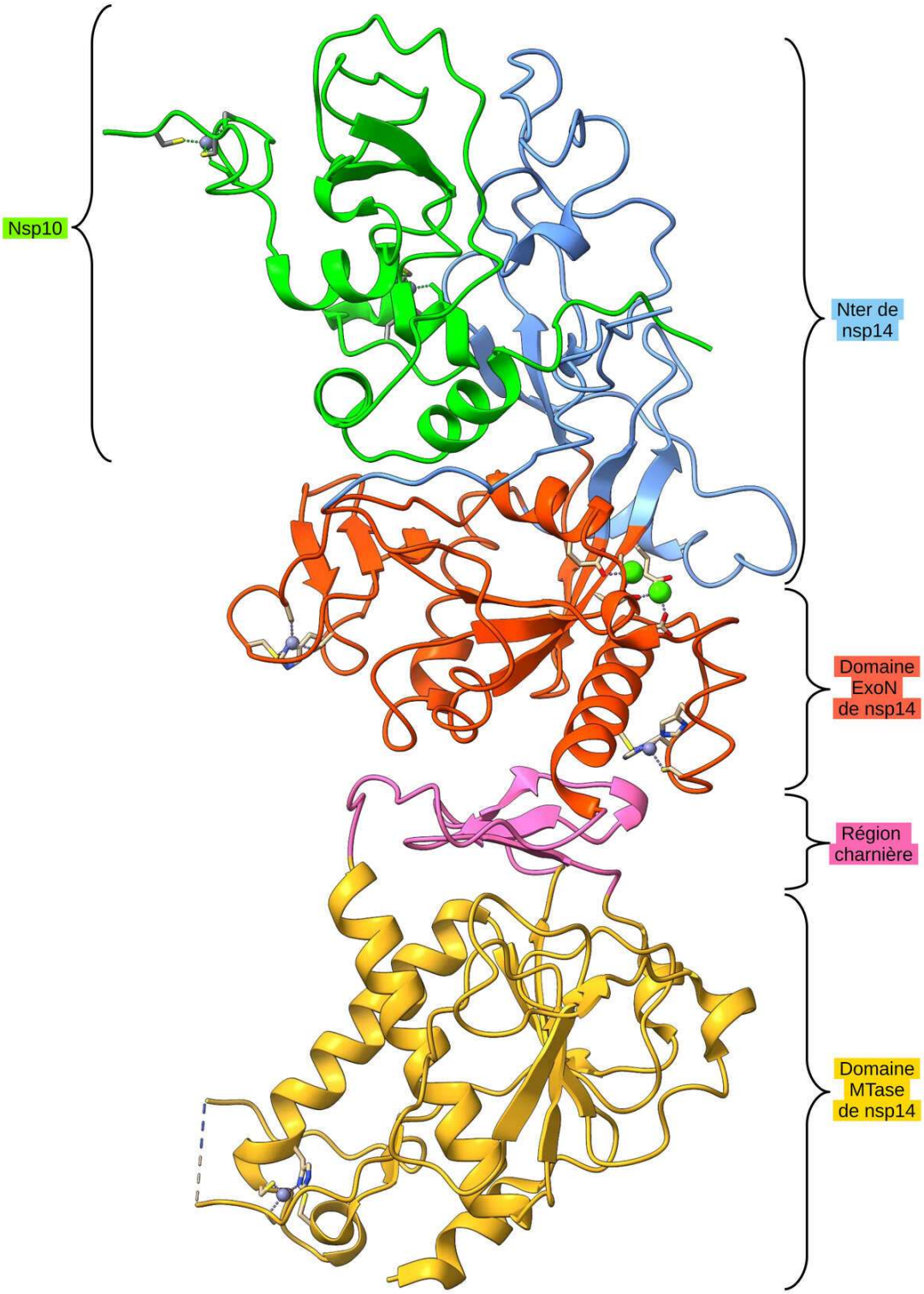


Figure 24 : Structure Cryo-EM du complexe nsp14/nsp10 à 3.90 Å

Structure du complexe nsp14/nsp10 (PDB : 7N0B) obtenue en Cryo-EM à 3.90 Å. Les différents domaine et régions ont été indiqué : nsp10 en vert, l'interface de nsp14 avec nsp10 en bleu, le domaine ExoN de nsp14 en orange, la région charnière en rose, le domaine MTase en jaune. Les 2 ions Mg^{2+} du domaine ExoN sont représentés en vert, et les ion Zn^{2+} des 5 doigts de zincs en violet (Liu *et al*. 2021).

Fait notable, la présence de doigts de zinc qui participent à structurer les protéines nsp14 (deux dans le domaine ExoN et un dans le domaine MTase), et deux autres dans nsp10. Leurs présences sont visible sur la structure nsp14/nsp10 SARS-CoV-1 (PDB : 5NFY) et sur une structure partielle de domaine Nter de nsp14 avec nsp10 de SARS-CoV-2 (PDB : 9FZ4). Une fois la poche catalytique formée avec le placement de ses quatre acides aminés conservés (Asp-Glu-Asp-Asp), deux cations métalliques (Mg^{2+} et/ou Mn^{2+}), et d'une molécule d'H_2O, permettront l'activité ExoN avec l'hydrolyse d'un ribonucléotide ou de NAs (Steitz et Steitz 1993). En effet, l'enzyme est spécifique des ribonucléotides, des matrices double brin et fonctionne dans le sens 3'->5' (Saramago *et al*. 2021).

In vivo, l'ExoN intervient après mésappariement sur le néo-brin d'ARN (3') lors de la transcription via la RdRp de nsp12. *In vitro*, l'ExoN permet la dégradation des nucléotides situés en 3' sur de l'ARNds.

Son inactivation *in vivo* augmente la quantité de mutations dans le génome viral et finit par compromettre le message génétique de ce dernier (Ferron *et al*. 2018).

1.7. Les méthyltransférases virales

Ici, nous nous intéresserons aux méthyltransférases virales spécifiques des virus à ARN, et plus particulièrement à celles des Coronavirus.

1.7.1. La N7-méthyltransferase nsp14 des Coronavirus

Les coiffes N7-méthylguanosine (^{7Me}G) sont une caractéristique structurale des ARNm eucaryotes, mais les virus d'organismes eucaryotes utilisent également cette spécificité. Cela leur permet d'être moins reconnus par l'immunité inné et les exonucléases cellulaires, mais également d'être plus facilement pris en charge par la machinerie de traduction de l'hôte. Ainsi, les Coronavirus ne font pas exception et ont une activité N7-guanine méthyltransférase (N7-MTase) qui est portée par la partie C-terminale de nsp14 (Voir Figure 24 ; Chen *et al*. 2009).

L'analyse structurale de cette dernière a révélé un repliement non-Rossmann (Ferron, Subissi *et al*. 2018), distinguant cette enzyme des méthyltransférases cellulaires et virales communément connues (Xie et Bourne 2008, Gana, Rao *et al*. 2013). Comme précédemment présenté (Voir Figure 24) ce domaine de nsp14 possède également un doigt de zinc qui l'aide à se structurer. Pour l'activité N7-MTase, l'enzyme utilise le S-adénosyl méthionine (SAM) en tant que donneur de méthyle, et transfert ce dernier en position N7 du GMP de la futur coiffe (Voir Figure 17).

Il existe plusieurs structures pour ce domaine, notamment pour SARS-CoV-1 avec une structure complète nsp14/10 (PDB : 5NFY) et chez SARS-CoV-2, il en existe des partielles comme (PDB : 8FRJ).

En raison de sa conservation chez les Coronavirus et de ses caractéristiques structurale uniques, la N7-MTase du CoV constitue une cible intéressante pour le développement de médicaments antiviraux (Sun, Wang *et al*. 2014, Aouadi, Eydoux *et al*. 2017, Ahmed-Belkacem, Sutto-Ortiz *et al*. 2020), afin de lutter contre le SARS-CoV-2 ou les futures menaces émergentes liées au CoV.

Ainsi, les doubles activités de nsp14 jouent un rôle primordial dans le cycle de vie des Coronavirus. D'une part, l'activité 3′->5′ ExoN (avec nsp10) est impliquée directement dans la synthèse des ARN viraux en permettant de maintenir la fidélité du génome à long terme. Et d'autres part, l'activité N7-MTase est impliquée dans la modification de la coiffe de l'ARN pour faciliter la traduction et échapper aux défenses de l'hôte. Cela fait de cette protéine une cible de choix pour la conception de médicaments (Hernández *et al*. 2022 ; Hausdorff *et al*. 2023).

1.7.2. La 2’O-méthyltransferase nsp16 des Coronavirus

La 2′-O-méthyltransférase (2-0-MTase) nsp16 des Coronavirus, de par son activité dans le processus de formation de la coiffe, joue un rôle crucial dans la maturation des ARN viraux (ARNv) et dans l'évasion du système immunitaire de l'hôte comme nous l’avons vu précédemment (Paramasivam 2020). Les 2′-O-méthylations (2′-O-Me) contribuent également à la stabilité des ARNv et assurent une traduction efficace en participant à la formation de la coiffe en 5′ des ARN messagers viraux, essentielle pour le recrutement du ribosome.

Cette activité enzymatique est permise par l’hétérodimère formé par nsp16 et son co-facteur nsp10 qui permet à nsp16 d’adopter une conformation active et en augmentant son affinité pour l'ARN substrat. Contrairement à nsp14, nsp16 ne possède pas de doigts de zinc.

La S-adénosyl méthionine (SAM) sert de donneur de groupe méthyle, permettant l’ajout de 2′-O-Me sur le premier ou le second ribose, conduisant à la formation du Cap-1 $^{7Me}GpppA_{2'\text{-}O\text{-}Me}$, qui correspond à la coiffe finale des ARNv, indispensable pour la protection de l'ARN viral contre les exonucléases, mais aussi pour une traduction efficace (Decroly *et al*. 2011).

Des études structurales ont permis de déterminer plusieurs structures cristallographiques de ce complexe enzymatique, offrant des informations détaillées sur son mécanisme d'action. Par exemple, pour la structure du complexe nsp16/nsp10 de SARS-CoV-1 (PDB : 2XYQ), pour SARS-CoV-2 (PDB : 7BQ7) et chez le MERS-CoV (PDB : 5YN5). Ces structures sont très proche les unes des autres et sont essentielles pour notre compréhension globale du mécanisme catalytique et pour le développement d'inhibiteurs spécifiques.

Du fait de son caractère essentiel dans le cycle de vie viral et son rôle dans l'évasion du système l'immunité inné, nsp16 est une cible pour la production d’antiviraux. Ainsi, des efforts de recherche sont également en cours pour identifier des molécules capables d'inhiber l'activité sont activité (Sulimov *et al*. 2022; Hassan *et al*. 2023).

1.8. Les antiviraux

Les antiviraux sont des molécules qui ciblent une ou plusieurs étapes du cycle de réplication d'un ou plusieurs types de virus. L'effet recherché est de bloquer complètement la réplication du virus, ou au moins de la ralentir fortement, et ainsi de ralentir l'infection virale chez l'hôte.

Ces derniers sont donc utilisés pour soigner une maladie pour laquelle il n'existe pas encore de vaccin sûr et efficace, ou pour aider des sujets immunodéprimés et/ou à risque, ou encore lorsqu'un patient présente une réponse immunitaire insuffisante à un vaccin, par exemple. Les antiviraux peuvent également servir à prévenir une infection potentielle, comme la PrEP (Prophylaxie Pré-Exposition) contre le VIH, composée de ténofovir disoproxil fumarate (TDF) et d'emtricitabine (FTC). On peut également citer la ribavirine et le sofosbuvir dans le cas de l'hépatite C.

Les antiviraux sont donc l'un des trois moyens de lutte contre les virus, avec la prévention et la vaccination.

Il existe plusieurs types d'antiviraux. Dans le cadre de la lutte contre le SARS-CoV-2, on peut par exemple citer les anticorps monoclonaux, qui permettent d'éviter l'entrée des virus comme l'Evusheld (tixagévimab/cilgavimab), qui peut être utilisé autant comme PrEP (Rubin 2022) que comme traitement, souvent chez des patients à risque (Glhoom 2024).

On peut également citer les traitements antiprotéase comme le Paxlovid, une combinaison de nirmatrelvir et de ritonavir, inhibant l'activité protéase de nsp5 (Amani et Amani, s. d.).

Enfin, on peut citer la famille des nucléotides analogues (NAs), qui constituent une part importante des antiviraux déjà découverts. Cette famille de molécules sera présentée dans la partie suivante, ainsi que d'autres types d'antiviraux.

1.8.1. Le RTC, une cible thérapeutique

Le complexe de réplication et de transcription joue un rôle central lors de la phase de réplication du virus afin d'assurer sa reproduction. Ainsi, vu du côté de l'expérimentateur et de la lutte antivirale, chaque enzyme qui le compose, et chacune des interactions qu'elles ont entre elles, est une occasion de plus pour trouver un grain de sable pour bloquer le mécanisme de cette machinerie. Lorsqu'on parle d'inhibiteurs et de réplication du génome viral, on pense en premier lieu, et pour des raisons historiques, aux nucléotides analogues (NA). Sous une forme 5'-triphosphate, ces derniers sont incorporés par l'activité RdRp de nsp12 dans le brin d'ARN naissant. Cependant, les Coronavirus ont une résistance naturelle aux NAs grâce à l'activité de relecture et de correction des erreurs d'incorporation fournie par l'activité ExoN de nsp14 et sa co-enzyme nsp10 (Shannon *et al*. 2024).

La recherche de NAs permettant de bloquer efficacement le RTC est primordiale dans la lutte contre les Coronavirus actuels et futurs. Ainsi, l'obtention de NAs spécifiques de la nsp12 qui ne pourraient pas être excisés par la nsp14, ou alors l'utilisation d'au moins deux NA qui bloquent ces deux enzymes, serait une étape importante dans cette lutte.

1.8.2. Les nucléosides, nucléotides et leurs analogues

Les nucléotides naturels sont les « briques » du génome, où l'on peut repérer trois sous-parties structurales.

- La première est un pentose, où, en partant du seul atome d'oxygène du cycle, les 5 atomes de carbone sont numérotés de 1' à 5', dans le sens des aiguilles d'une montre (voir Figure 4). Ce pentose est un ribose dans le cas de l'ARN et un 2'-désoxyribose dans le cas de l'ADN.
- Ce pentose est relié en 1' à une base azotée, par une liaison C-N, qui peut être soit de type purine, soit de type pyrimidine. Classiquement, les bases puriques sont les résidus adénine (A) et guanine (G), et les bases pyrimidiques sont les résidus thymine (T) et cytosine (C) pour l'ADN, ou uracile (U) pour l'ARN.Le pentose et sa base azotée sont appelés nucléoside.
- En 5' du pentose peut alors être relié à 1, 2 ou 3 groupements phosphate peuvent être ajoutés, on parlera alors de nucléoside 5'-mono, di ou tri-phosphate, ou de nuclétide (voir Figure 4).

Les nucléotides ainsi formés peuvent se lier les uns aux autres, toujours dans le sens 5' vers 3' du pentose, avec un seul phosphate entre chaque pentose. Ainsi, le nucléotide tri-phosphate sera ajouté via une réaction de condensation entre le premier et le second phosphate en 5', avec le 3'-OH du nucléotide précédent. Cette nouvelle

liaison phosphodiester ainsi formée entraînera la libération d'une molécule d'eau et des deux phosphates liés (ou pyrophosphate (PPi)).

La diversité chimique et structurale les nucléotides naturels permet ainsi un très grand nombre de modification potentielles pour créer des NAs.

1.8.3. Les nucléotides analogues (NAs)

Un nucléotide analogue est une molécule dont la structure finale imite celle d'un nucléotide naturel, mais qui possède également des groupement chimiques originaux.

Il sont conçus pour interférer avec les processus biologiques impliquant les nucléotides naturels et inhiber la réplication virale. Lorsque ces analogues sont confondus avec des nucléotides naturels et incorporés dans le génome viral, ils peuvent soit empêcher la synthèse en bloquant la réplication du génome, ou bien être incorporé plusieurs fois, et ainsi altérer le génome ce qui le rendra non-fonctionnel par la suite.

1.8.3.1. Les NAs au cours du temps

Les premiers traitements efficaces contre le VIH incluaient des inhibiteurs nucléosidiques de la transcriptase inverse (NRTI), comme l'azidothymidine (AZT), approuvée en 1987 (« State-of-the-Art Conference on Azidothymidine Therapy for Early HIV Infection. » 1990). Dans le cas du VIH, il a rapidement été observé que la combinaison de plusieurs antiviraux (dont des NAs) permettait d'obtenir une meilleure efficacité et de diminuer les chances d'apparition de mutants de résistance. Ainsi, pour les traitements contre le VIH, comme les trithérapies associant trois de ces antiviraux, il y a classiquement deux NRTI, et soit un inhibiteur non nucléosidique de la transcriptase inverse, soit un inhibiteur de la protéase, soit un inhibiteur de l'intégrase (G. Li, Wang, et De Clercq 2022).

Par la suite, d'autres NAs ont été découverts pour répondre aux besoins de la médecine préventive. Par exemple, l'aciclovir est utilisé pour traiter les infections dues au virus de l'herpès, qui a un génome en ADN double brin. Ou encore le sofosbuvir et le remdesivir (Voir Figure 25) qui agissent contre le virus de l'hépatite C, un virus à un génome d'ARN simple brin de polarité positive. Puis, avec l'apparition des premiers Coronavirus humains dangereux en 2003, les premiers NAs contre ces derniers ont également commencé à être développés.

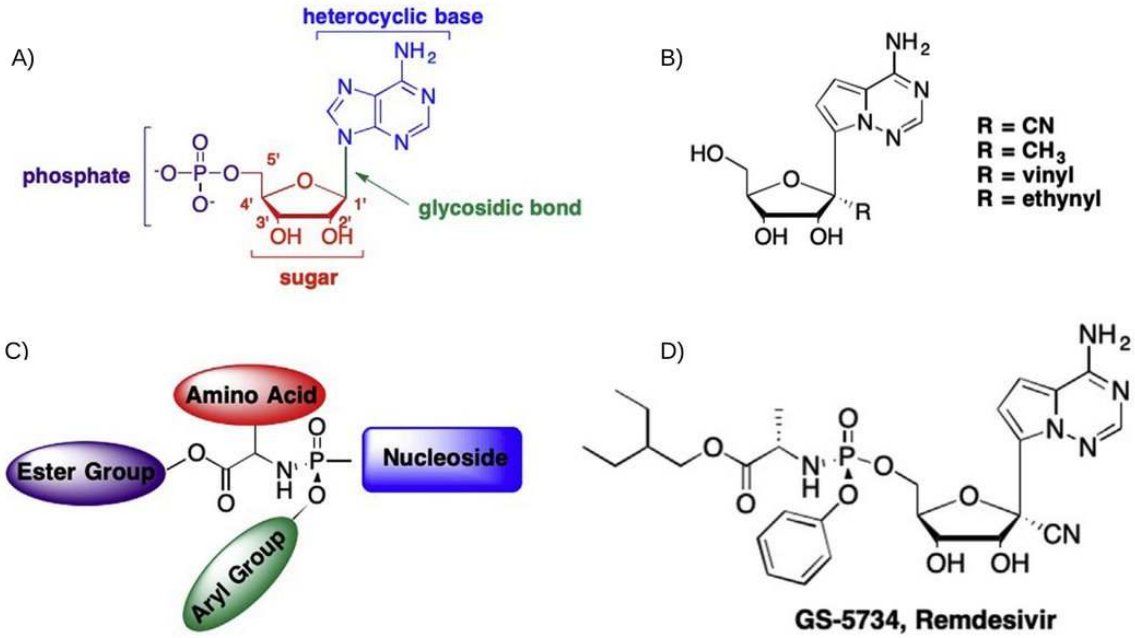


Figure 25 : Schéma des modifications des nucléosides analogues (NA) de l'exemple du Remdesivir

Exemples de plateforme de construction de nucléotide analogue de chez Gilead. A) Sites de modifications potentielles des analogues de nucléos(t)ide. B) Indications des modifications de la première génération de Gilead sur le NA 1′-substituted 4-aza-7,9-dideazaadeonosine C-nucleosides. C) Structure générale d'un McGuigan ProTide D) Structure du McGuigan ProTide GS-5734, le Rendesivir. Figurs extraite de (Yates et Seley-Radtke 2019).

1.8.3.2. NAs contre les Coronavirus

Avec les épidémies du SRAS-CoV puis du MERS-CoV, il s'est avéré que l'absence de traitements vaccinaux ou antiviraux pouvant prévenir ou traiter ces infections virales chez l'humain pouvait devenir problématique en cas de transformation d'une épidémie en pandémie, ou de franchissement de la barrière zoonotique par de nouveaux Coronavirus.

La présence de nsp14 avec son activité de relecture et d'excision des mésappariements rend bien plus complexe le développement de NAs anti-Coronavirus efficaces. Ainsi, malgré une activité de recherche accrue après l'apparition de ces deux virus, aucun NA pleinement satisfaisant n'a pu être découvert.

L'arrivée du SARS-CoV-2 nous a une nouvelle fois prouvé l'importance du développement de ce type de molécules. La pandémie de COVID-19 a donc intensifié à nouveau la recherche et le développement de traitements antiviraux et de vaccins efficaces.

Ainsi, l'une des premières molécules antivirales à avoir été mise sur le devant de la scène était le Remdesivir, développé par Gilead Sciences en 2009 initialement contre l'hépatite C. Le Remdesivir est une prodrogue (GS-441524) qui doit être administrée par voie intraveineuse afin d'être biotransformée en nucléotide analogue dans la cellule, le GS-441524 (5'-triphosphate), pour altérer les génomes viraux lors de son incorporation par la polymérase virale (Bakheit *et al*. 2023).

L'efficacité de cette molécule seule contre l'hépatite C était plutôt modérée, mais des tests *in vitro* ont montré une certaine efficacité du Remdesivir contre quelques virus, notamment des Filovirus (comme Ebola) et des Coronavirus. Le Remdesivir a ainsi été utilisé lors d'un essai clinique pendant une épidémie du virus Ebola en Afrique de l'Ouest de 2013 à 2016. Les premiers résultats de cette étude étaient encourageants, ce qui a permis la poursuite des essais cliniques contre Ebola. Malheureusement, l'efficacité s'est avérée insuffisante pour traiter des patients virémiques, peut-être dû au fait que les traitement interviennent souvent dans des phases trop tardives de l'infection (Cao, Deng, et Dai 2020).

L'émergence du SARS-CoV-2 a permis les premiers essais cliniques randomisés du Remdesivir, d'abord sur plusieurs petits groupes de patients, avec des symptômes légers à sévères, contrôlés en hôpital dans 17 pays. En octobre 2020, les États-Unis ont approuvé le Remdesivir pour un usage médical. En novembre 2020, l'Organisation mondiale de la santé (OMS) a recommandé quant à elle une utilisation conditionnelle du Remdesivir, faute de preuves suffisantes pour appuyer son utilisation à plus grande échelle à cette époque. Puis, en avril 2022, l'OMS a mis à jour ses recommandations après la publication de résultats précliniques positifs au Canada, notamment pour les cas les plus graves (Ali *et al*. 2022).

Aujourd'hui, 4 ans après le début de l'épidémie, plusieurs études démontrent que le Remdesivir semble améliorer la survie globale des patients infectés (Godwin *et al*. 2024). Récemment approuvé par le FDA (Food and Drugs Administration) aux États-Unis pour une utilisation large, et par la HAS (Haute Autorité de Santé) pour une utilisation contre certaines souches du SARS-CoV-2.

On peut citer un autre cas de NA, le Molnupiravir (voir Figure 26), connu depuis une quarantaine d'années comme antiviral contre différents virus. Lors de la pandémie, l'efficacité de cette molécule a également été étudiée lors de phases cliniques, malgré des effets mutagènes soupçonnés dès novembre 2019 dans la cellule de l'hôte. Cependant, les faibles effets bénéfiques du Molnupiravir, surtout s'il n'est pas pris précocement, cumulés avec ses risques cancérigènes, ont fait que cette molécule a été mise de côté en France en particulier (https://www.has-sante.fr/jcms/p_3304143/fr/lagevrio-molnupiravir-covid-19). De plus, des études récentes démontre que le Molnupiravir, de par son mécanisme d'action, serait à l'origine de création de variant chez des patients aillant eu une infection longue au SARS-CoV-2 (Sanderson *et al*. 2023).

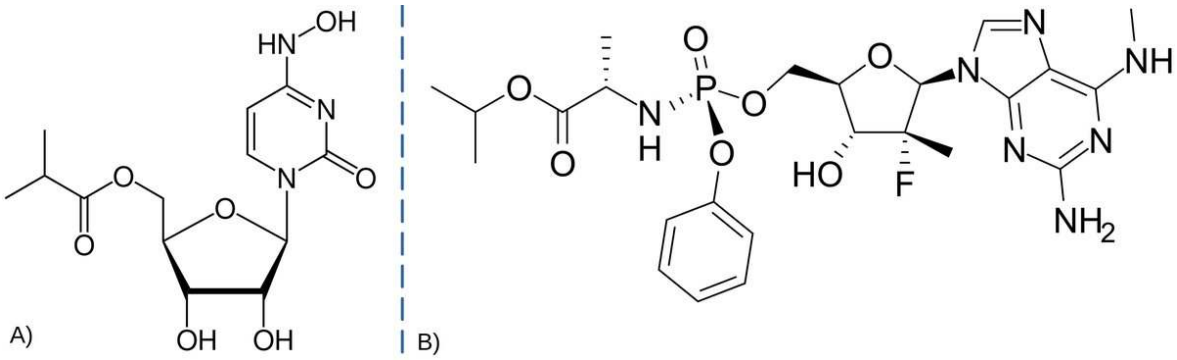


Figure 26 : Schéma des molécules d'exemples de NAs (Prodrugs)

Schéma des prodrugs Molnupiravir (A) et du Bemnifosbuvir (B).

Un cas de NA intéressant est celui du Bemnifosbuvir (voir Figure 26), développé par ATEA Pharmaceuticals. Initialement développé contre l'hépatite C, l'efficacité et le mécanisme d'action de ce NA contre le SARS-CoV-2 ont également été étudiés (Shannon *et al*. 2022). Le Bemnifosbuvir, dont la molécule active 5'-triphosphate est nommée AT-9010, a l'avantage de pouvoir être administré par voie orale. Comparé aux NAs précédemment étudiés, qui sont principalement des mutagènes incorporés par la RdRp de nsp12, l'AT-9010 inhibe la RdRp en bloquant l'activité de polymérisation par terminaison de chaîne, malgré une résistance partielle à l'excision par l'ExoN nsp14/10. De plus, cet NA vient également se loger en profondeur dans la poche catalytique du NiRAN, bloquant ainsi l'activité de ce dernier (Shannon *et al*. 2022). L'inactivation du NiRAN étant létale pour le virus (Lehmann *et al*. 2015) et la synthèse de chaines complètes d'ARN étant grandement réduite, cela rend ce

candidat médicament attrayant pour la lutte contre le SARS-CoV-2. De plus, le fait d'avoir deux cibles permet de limiter le risque de mutation de résistance. Même si la mortalité due au de SARS-CoV-2 est moins importante aujourd'hui, plusieurs essais cliniques pour cette molécule ont eu lieu, avec des résultats prometteurs en phase 2 et des phases cliniques 3 dont les résultats devraient être connus fin 2024.

Pour conclure, la nsp12, de par ses activités dans le RTC, fait d'elle une cible extrêmement intéressante pour le développement d'antiviraux. Son niveau élevé de conservation structurale peut laisser penser que l'apparition de mutations de résistance contre certains NAs doit être limitée, mais reste possible. Cela a déjà été prouvé plusieurs fois pour le Remdesivir, par exemple, où après plusieurs passages de cellules infectées en présence de ce NA, plusieurs laboratoires ont observé l'apparition de différentes mutations comme V166A-N198S-S759A, entre autres (Stevens *et al*. 2022). Pour la dernière mutation, il a été démontré que S759A améliorait la discrimination de la RdRp contre l'incorporation du Remdesivir biotransformé (GS-441524 triphosphate (RTP ; Stevens *et al*. 2022). Les mutations F480L et V557L, quant à elles, modifient la structure du site catalytique de nsp12, ce qui affecte ainsi le contrôle de la fidélité de la RdRp avant l'incorporation de nouveaux nucléotides (Shannon *et al*. 2020).

Les RdRps des Coronavirus sont ainsi capables d'accueillir une grande variété de NAs avec diverses modifications chimiques, que ce soit sur le ribose, la base, ou les deux (Pruijssers et Denison 2019). Cependant, le domaine ExoN est également capable de reconnaître de nombreuses modifications de ces derniers. Cela diminue l'efficacité globale du NA potentiel et augmente les risques de mutations de la nsp12 et donc de résistance de ces virus sur le long terme. Cette caractéristique des Coronavirus ajoute un niveau de complexité supplémentaire lors du développement de NAs. De plus, l'absence de données structurales de nsp14 à résolution cristallographique satisfaisante (<2,5 Å) rend également plus complexe la compréhension globale du mécanisme d'excision des NAs et donc la recherche de stratégies de son contournement avec de nouveaux NAs (Ferron *et al*. 2018).

1.8.4. Les chélateurs d'ions

Dans le cadre de la production de nouveaux antiviraux, l'une des stratégies est la recherche d'inhibiteurs spécifiques des exonucléases (ExoN) virales, qui sont présentes chez les Arenavirus et les Coronavirus. Du fait que l'activité ExoN n'a pas besoin de métabolites particuliers, une autre stratégie devait être mise en place. L'activité ExoN nécessite deux ions métalliques pour stabiliser la molécule d'eau ainsi que le nucléotide à exciser, et ce, avec l'aide des acides aminés du site actif pour catalyser la réaction d'hydrolyse. C'est ainsi que fut étudiée l'utilisation de chélateurs d'ions capables de capturer spécifiquement les deux ions nécessaires pour cette activité, et de rester à l'intérieur de la poche catalytique pour bloquer l'enzyme de cette dernière (Hernández *et al*. 2022). Une fois des inhibiteurs potentiels identifiés, il sera

toujours intéressant de travailler sur différentes parties de la molécule, notamment la partie chélatrice d'ions, mais aussi sur différents groupements autour de cette partie chélatrice d'ions afin qu'elle puisse se fixer spécifiquement à une poche hydrophobe proche du site catalytique. Cela permet d'augmenter l'efficacité et la spécificité de ce type de molécule.

1.8.5. Les analogues du SAM

La molécule de SAM est un métabolite qui sert de substrat pour le transfert de méthyl par les méthyltransférases , et dont le produit de ce métabolite sera le SAH. Le SAM produit par la cellule hôte, et est autant utilisé par cette dernière que par les méthyltransférases virales. Dans le cas des Coronavirus, le SAM est utilisé par nsp14 et nsp16 toutes deux impliqué respectivement dans la N7-méthylation et 2'-O-méthylation lors de la formation de la coiffe.

L'utilisation de SAM de analogue est donc une stratégie permanentant pour bloquer ce processus vital au virus. Différentes de ces molécules sont actuellement étudié dans le cas des Coronavirus, dont certaines sont spécifique de la nsp14 (Hausdorff *et al*. 2023) et d'autres de la nsp16 (Sulimov *et al*. 2022).

1.8.6. Différence entre le RTC des SARS et du MERS

Lors des différentes études menées sur l'AT-9010 et sa prodrogue l'AT-511, des différences importantes d'efficacité du composé ont été observées, selon qu'ils étaient utilisés contre le SARS-CoV et le SARS-CoV-2, ou le MERS. Dans une étude récente (Good *et al*. 2021), des essais sur des cellules Huh-7 infectées par le SARS-CoV ou le MERS, et des cellules HAE infectées par le SARS-CoV-2, puis traitées avec différentes doses d'AT-511, ont été réalisés.

Selon cette étude, la réduction du rendement viral (virus yield reduction (VYR)), qui détermine la concentration efficace requise du composé pour atteindre 90 % d'inhibition (EC90) de l'effet cytopathique induit par le virus (CPE), a montré que la concentration efficace d'AT-511 était de 0,34 µM contre le SARS-CoV, alors qu'elle était de 37 ± 28 µM contre le MERS-CoV. Concernant le SARS-CoV-2, l'EC90 de l'AT-511 était de 0,47 ± 0,12 µM.

D'après cette étude, réalisée sur des modèles cellulaires, l'AT-511 est environ 100 fois plus efficace contre le SARS-CoV-2 que contre le MERS-CoV.

Cette différence d'efficacité des antiviraux, notamment les NAs, entre ces trois Coronavirus a déjà été observée à plusieurs reprises. Les nsp12 de SARS-CoV et de

SARS-CoV-2 partagent 96 % d'identité de séquence, alors que le MERS partage respectivement 72 % et 71 % d'identité de séquence avec ces virus.

Cette différence d'efficacité pourrait s'expliquer par des variations dans la structure et le fonctionnement du complexe de réplication-transcription (RTC) du MERS, comparé aux deux autres Coronavirus. Ces virus font tous partie des bêta-Coronavirus, mais appartiennent à deux sous-groupes différents (groupe B pour les deux SARS, et groupe C pour le MERS ; voir Figure 11). Ces observations soulignent l'importance d'étudier spécifiquement chaque Coronavirus, même au sein de la même famille, car certaines spécificités peuvent considérablement faire varier leur sensibilité aux antiviraux.

C'est pour ces raisons que nous avons voulu nous pencher sur l'étude de la nsp12 du MERS, qui sera abordée dans le chapitre « Chapitre 3 : Étude structurale et fonctionnelle des nsp12 des Coronavirus ».

Ainsi, la découverte d'une molécule active et de son précurseur efficace contre un virus peut également permettre de comprendre certaines spécificités d'autres virus de la même famille, ce qui peut permettre à terme d'adapter des molécules antivirales pour élargir leur spectre d'efficacité.

De même que la stratégie de lutte contre le VIH est devenue réellement efficace avec l'utilisation de plusieurs antiviraux afin de limiter l'apparition de mutations de résistance, il est essentiel aujourd'hui d'étudier les familles de virus émergents et de développer différentes catégories d'antiviraux afin de limiter l'apparition de résistances chez les virus actuels et de mieux nous préparer contre les futures épidémies.

Objectif de la thèse

Dans le contexte de l'épidémie de SARS-CoV-2 de 2020, la recherche d'antiviraux contre ce virus est devenue une priorité. Des repositionnements de molécules, comme le remdesivir, ont eu lieu, en plus de la recherche de nouvelles molécules.

Durant ma thèse, je me suis intéressé à la comparaison structurale et fonctionnelle des systèmes de réplication et de transcription (RTC) des Coronavirus. Pour cela, je me suis d'abord intéressé à la production des trois protéines minimales du RTC, à savoir nsp12, nsp8 et nsp7. Pour cela, j'ai cherché à produire en système de production bactérien les trois protéines du RTC du SARS-CoV-1 et du SARS-CoV-2.

Nsp7 et nsp8 sont les co-enzymes de la nsp12, qui porte l'activité principale de ce complexe minimal. La nsp12 possède l'activité ARN polymérase ARN-dépendante (RdRp) et l'activité nucléotidyltransférase associée au Nidovirus RdRp (NiRAN).

Lors de cette étude comparative, il a été observé qu'un nucléotide analogue, l'AT-9010, pouvait interrompre la synthèse d'ARN et altérer le NiRAN de la nsp12. L'effet de l'AT-9010, dont la prodrogue est l'AT-527, a initialement été trouvé pour lutter contre le virus de l'hépatite C (VHC). Ce composé pourrait donc être repositionné également pour lutter contre le SARS-CoV-2. L'AT-9010 est un terminateur de chaîne du domaine RdRp de la nsp12 et peut également se loger dans la cavité du domaine NiRAN de cette dernière.

Des expériences complémentaires sur cellules ont été effectuées par un laboratoire partenaire. L'efficacité de l'AT-527 a été testée sur des cellules infectées par différents Coronavirus humains, dont le SARS-CoV-1, le SARS-CoV-2 et le MERS-CoV. Les différences de VYR observés interroge. Je me suis donc intéressé au RTC de MERS, et j'ai voulu produire ce complexe enzymatique *in vitro* pour pouvoir plus facilement étudier les déterminants structuraux et fonctionnels pouvant expliquer ces différences observées. J'ai donc réalisé une production similaire à celle des SARS-1 et SARS-2. Après avoir pu produire nsp7 et nsp8, des difficultés plus importantes pour la production de la nsp12 de MERS ont été rencontrées. Cela nous a poussés à expérimenter la production dans d'autres systèmes d'expression.

Il a également été démontré que l'AT-9010 pouvait être partiellement excisé, ce qui est une des particularités des Coronavirus. Cette capacité est apportée par l'activité de relecture de l'exonucléase (ExoN) de la nsp14 (et sa co-enzyme nsp10) qui lui permet l'excision des mésappariements qui pourraient avoir lieu avec ce type de molécules, ce qui peut diminuer l'effet antiviral d'un médicament.

Je me suis donc intéressé dans un second temps à l'étude structurale et fonctionnelle de la nsp14, afin d'identifier les déterminants qui pourraient permettre de trouver un inhibiteur pour la bloquer également. La nsp14 étant une enzyme bifonctionnelle, elle comporte également une activité N7-méthyltransférase (N7-MTase). Ce second domaine situé en C-ter apporte une grande flexibilité à cette enzyme, ce qui peut compliquer son étude par diffraction aux rayons X.

Pour pallier cela, trois familles de VHH ont été produites afin d'aider à la stabilisation et à la cristallisation de la nsp14. Ainsi, les interactions de ces familles de VHH ont été caractérisées via différents essais enzymatiques et analyses bio-physiques, dont du SAXS. Des conditions de cristallogenèse ont également été déterminées, ce qui a permis la résolution de la structure en cristallographie aux rayons X. Cette étape nous a également permis de commencer à cribler différents inhibiteurs de la nsp14.

Chapitre 2 : Matériel & Méthode

2. Matériel & Méthode

2.1. Généralités

Toutes les expériences ont été réalisées avec des réactifs de pureté adaptée aux besoins de chaque expérience. Chaque tampon a été préparé à partir de poudre, à l'exception des tampons utilisés pour la cristallographie, la BLI et les essais enzymatiques, qui ont été préparés à partir de solutions stock. La préparation des milieux de culture, ainsi que des tampons utilisés pour la purification des protéines, la BLI, les essais enzymatiques, la chromatographie par filtration sur gel (GF), la cristallographie, la diffusion multiangle de la lumière (MALS) et la diffusion des rayons X aux petits angles (SAXS), a été réalisée dans de l'eau Milli-Q filtrée à 0,22 µm. Les tampons utilisés pour la BLI, les essais enzymatiques, la GF, la cristallographie, la MALS et la SAXS ont ensuite été refiltrés à 0,22 µm après leur préparation, et dégazés pour les quatre derniers.

2.2. Clonage des plasmides

2.2.1. Constructions plasmidiques

Les plasmides codant pour nsp14, nsp10, nsp12, nsp7 et nsp8, provenant respectivement des isolats Frankfurt-1 pour le SARS-CoV (NCBI ref. seq. : AY291315.1), de Wuhan-Hu-1 pour SARS-CoV-2 (NCBI ref. seq. : YP_009724389.1), et du *Betacoronavirus* England 1pour MERS (NCBI ref. seq. : YP_007188577) pour nsp12, nsp7 et nsp8, ont été optimisés pour l'expression chez *E. coli*, synthétisés, puis clonés dans le plasmide pQE30 (promoteur T5), avec l'ajout en N-terminal d'une étiquette 6 histidines en fusion avec le site de clivage TEV (ENLYFQG), où la dernière glycine a été remplacée par le premier acide aminé de la séquence codante du gène d'intérêt.

Une autre construction plasmidique pour nsp12 de ces trois virus a été réalisée, en clonant leurs gènes respectifs dans le plasmide pJ404 (promoteur T5), cette fois avec la fusion du site de clivage TEV (ENLYFQG), suivie de l'étiquette 6 histidines, ajoutée en C-terminal.

Pour les plasmides des différentes familles de VHH, dans un premier temps les fragments d'ADNc codant pour le VHH des camélidés on été amplifiés (décrit en « 2.7. Isolation ADNc des VHHs », puis clonés dans le plasmide pHEN4 pour créer une

banque. Cette banque de plasmides a ensuite été utilisée pour les différentes étapes du phage display.

Les plasmide d'expression finaux de ces vhh ont été synthétisés précédée de la séquence de sécrétion peLB "AAGTATTTACCGACTGCTGCTGCGGCTGCTGCTGGGCGCAA CCT" en 5' et d'une séquence codant pour un tag x6 histidine en 3'. La séquence de cette protéine de fusion a ensuite été clonés entre les sites de restriction BamHI et HindIII du plasmide pET24 (promoteur T7) chez Twist Bioscience (USA).

Concernant la N7-guanine RNA cap methyltransferase du virus White Bream (NCBI ref. seq. : YP_803214.1), la séquence de Met1333 à Ile1668 a été synthétisée par GenScript, puis clonée dans les plasmides pQE30 (promoteur T5).

2.2.2. Amorces et amplifications par PCR

Les amorces utilisées pour le séquençage, l'amplification de plasmides, la mutagenèse dirigée ou comportant des régions flottantes ont toutes été commandées chez Eurofins Genomics. Elles ont été conçues pour avoir une température d'hybridation (Tm) comprise entre 60 et 62 °C.

Conditions de PCR	
	Volumes
Tp 5x Phusion GC	5µl
H2O qsp 50 µl	Qsp 50µl
dNTP (200µM final)	1µl
DMSO	5µl
Phusion 1/10 dans Tp GC+ (1Unité)	0.5µl
Matrice (de 0,05 à 10ng)	1µl
Amorce 5' 10µM (Stock) 0,5 µM final	2.5µl
Amorce 3' 10µM (Stock) 0,5 µM final	2.5µl

Programme de PCR			
Étapes	T°C	Temps	
Dénaturation initiale	98°C	3 min	
Dénaturation	98°C	10s	
Hybridation	Tm-2°C	20s	x30-35 Cycles
Extension	72°C	15s/kbs	
Extension finale	72°C	5-10min	
Stockage	4°C	Frigo	

Tableau 3 : Conditions et programme type des PCR

Les amorces conçues ont été utilisées à une concentration finale de 10 µM avec la polymérase Phusion™ (réf. F530L de ThermoFisher) pour amplifier le gène d'intérêt

selon le programme indiqué dans le Tableau 3. La taille des produits d'amplification PCR a été évaluée par électrophorèse sur gel d'agarose à 0,8 % dans du tampon TAE.

2.2.3. Mutagenèse dirigée par PCR

Lors des mutagenèses dirigées (mutations ponctuelles ou délétions), les amorces ont été conçues pour posséder environ 15 nucléotides de région flottante, permettant ainsi, après la PCR, d'obtenir une région chevauchante d'environ 30 nucléotides au total. Une PCR a ensuite été réalisée avec une faible quantité de matrice (voir Tableau 3). Par la suite, le plasmide matrice a été digéré avec l'enzyme DpnI à 37 °C pendant 2 heures, avant d'être transformé dans des cellules *DH5α* ou *XL1-Blue*, comme indiqué dans la section « 2.3. Transformation de plasmides ». Les plasmides ont été extraits par minipréparation à l'aide du kit Nucleospin Plasmide de Macherey-Nagel (réf. 740588.250), puis séquencés chez Eurofins Genomics à l'aide du séquençage de Sanger.

2.3. Transformation de plasmides

Environ 100 ng de plasmide ou de produit de PCR ont été transformés dans des cellules DH5α pour l'amplification du plasmide, ou dans des cellules C2523 (NEB réf. C2523) pour les plasmides sous contrôle du promoteur T5, co-transformées ou non avec le plasmide chaperon pGT-f2 (Takarabio réf. 3340), ou dans des cellules C2566 (NEB réf. C2566) pour les plasmides sous promoteur T7, afin d'exprimer la protéine.

Les cellules et l'ADN plasmidique ont été incubés pendant 30 minutes sur glace. Ensuite, un choc thermique a été appliqué à 42 °C pendant 45 secondes, suivi d'une incubation sur glace pendant 3 à 5 minutes. Les cellules transformées ont ensuite été incubées après l'ajout de 950 µl de milieu Luria Broth (LB, AthenaES) ou de milieu de récupération d'expression (Lucigen) pour les transformations difficiles, pendant 1 heure à 37 °C sous agitation à 180 rpm. Une courte centrifugation a permis de précipiter les cellules sous forme de culot, qui ont ensuite été remises en suspension et étalées sur des plaques de gélose LB contenant les antibiotiques appropriés. Les plaques ont été incubées à 37 °C pendant la nuit.

2.4. Expression et purifications des protéines

2.4.1. nsp12 de SARS-CoV

Après transformation du plasmide pJ404 (étiquette 6 histidines en C-terminal) ou pQE30 (étiquette 6 histidines en N-terminal) codant pour nsp12 dans les cellules C2523 (NEB réf. C2523), co-transformées avec le plasmide chaperon pGT-f2 (Takarabio réf. 3340), les cultures ont été réalisées à 37 °C. Lorsque la DO600 a atteint entre 0,4 et 0,6, les cultures ont été retirées de l'incubateur et refroidies pendant 15 minutes à température ambiante avant d'être induites avec 250 µM d'IPTG, 2 % d'éthanol et 5 µg/L de tétracycline pour l'induction des chaperons. Après induction, la culture a été poursuivie à 23 °C sous agitation (200 rpm). Les cellules ont été récoltées après 18 à 20 heures de culture par centrifugation à 8 000 g à 4 °C pendant 10 minutes. Les culots cellulaires secs ont ensuite été stockés à -80 °C. Le milieu LB (AthenaES #0103) a été utilisé pour ces cultures.

Les culots cellulaires ont été ressuspendus avec 50 ml/L de culture dans le tampon A1 (50 mM Tris pH 8, 300 mM NaCl, 5 mM $MgSO_4$, 10 % glycérol, 0,5 mM TCEP), auquel ont été ajoutés 0,5 mg/ml de lysozyme, 10 µg/ml de DNase, 1 mM de PMSF et 0,2 mM de benzamidine. Le mélange a été laissé sous agitation magnétique pendant 45 minutes à 4 °C, puis une quantité de sel en poudre suffisante pour atteindre une concentration finale de 1 M a été ajoutée. Le mélange a été de nouveau agité pendant 45 minutes à 4 °C, puis centrifugé pendant 30 minutes à 30 000 g. Le surnageant obtenu a été placé sous agitation magnétique pendant 1 heure avec 2 ml de résine Talon Superflow (#28957502) par litre de culture, préalablement équilibrée avec le tampon A1.

Une purification par lot a été réalisée, avec plusieurs étapes de lavage correspondant à 10 volumes de colonne (CV) : d'abord avec le tampon W1 (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol, 5 mM imidazole, 0,5 mM TCEP), suivi du tampon W2 (50 mM Tris pH 8, 1 M NaCl, 10 % glycérol, 5 mM imidazole, 0,5 mM TCEP), puis à nouveau avec le tampon W1 avant d'effectuer l'élution avec le tampon d'élution (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol, 250 mM imidazole, 0,5 mM TCEP).

Le pic d'élution a été concentré à l'aide d'un concentrateur Amicon Ultra 15 (#UFC903024) jusqu'à environ 3,5 mg/ml, puis centrifugé à 16 000 g pendant 5 minutes à 4 °C avant d'être injecté sur une colonne d'exclusion de taille Superdex S200 16/60 (#28989335), préalablement équilibrée avec le tampon GF2 (25 mM Hepes pH 8, 300 mM NaCl, 5 mM $CaCl_2$, 10 % glycérol). Le pic correspondant à la chromatographie d'exclusion de taille a été concentré à environ 9 mg/ml (soit environ

85 µM) à l'aide d'un Amicon Ultra 4 (#UFC803024). La nsp12 ainsi purifiée a été flash congelée dans de l'azote liquide, puis conservée à -80 °C pour une utilisation future.

2.4.2. nsp12 de SARS-CoV-2

Après la transformation du plasmide pJ404 (étiquette 6 histidines en C-terminal) ou pQE30 (étiquette 6 histidines en N-terminal) codant pour nsp12 dans les cellules C2523 (NEB réf. C2523), co-transformées avec le plasmide chaperon pGT-f2 (Takarabio réf. 3340), les cultures ont été réalisées à 37 °C. Lorsque la DO600 a atteint entre 0,4 et 0,6, les cultures ont été retirées de l'incubateur et refroidies pendant 15 minutes à température ambiante avant l'induction du plasmide avec 250 µM d'IPTG, 2 % d'éthanol, et 5 µg/L de tétracycline pour l'induction des chaperons. Après l'induction, la culture a été poursuivie à 23 °C sous agitation (200 rpm). Les cellules ont été récoltées après 18 à 20 heures de culture par centrifugation à 8 000 g à 4 °C pendant 10 minutes. Les culots cellulaires secs ont ensuite été stockés à -80 °C. Le milieu LB (AthenaES #0103) a été utilisé.

Les culots secs ont été resuspendus dans 50 ml/L de culture de tampon A1 (50 mM Tris pH 8, 300 mM NaCl, 5 mM MgSO4, 10 % glycérol, 0,5 mM TCEP) auquel ont été ajoutés 0,5 mg/ml de lysozyme, 10 µg/ml de DNase, 1 mM de PMSF, et 0,2 mM de benzamidine. Le mélange a été laissé sous agitation magnétique pendant 45 minutes à 4 °C, puis du sel en poudre a été ajouté pour atteindre une concentration finale de 1 M. Le mélange a été agité pendant encore 45 minutes à 4 °C, puis centrifugé pendant 30 minutes à 30 000 g. Le surnageant a été récupéré et placé sous agitation magnétique pendant 1 heure avec 2 ml de résine Talon Superflow (#28957502) par litre de culture, préalablement équilibrée avec le tampon A1.

Une purification en batch a été réalisée, avec un lavage de 10 volumes de colonne (CV) avec le tampon W1 (50 mM Tris pH 8, 1 M NaCl, 10 % glycérol, 5 mM imidazole, 0,5 mM TCEP), avant de procéder à l'élution avec le tampon d'élution (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol, 250 mM imidazole, 0,5 mM TCEP).

Le pic d'élution a été concentré à l'aide d'un concentrateur Amicon Ultra 15 (#UFC903024) jusqu'à environ 3,5 mg/ml, puis centrifugé à 16 000 g pendant 5 minutes à 4 °C avant d'être injecté sur une colonne d'exclusion de taille Superdex S200 16/60 (#28989335), préalablement équilibrée avec le tampon GF2 (25 mM Hepes pH 8, 300 mM NaCl, 5 mM CaCl2, 10 % glycérol). Le pic correspondant à la chromatographie d'exclusion de taille a été concentré à environ 9 mg/ml (environ 85 µM) à l'aide d'un Amicon Ultra 4 (#UFC803024). La nsp12 ainsi purifiée a été flash-congelée dans de l'azote liquide, puis conservée à -80 °C pour une utilisation future.

2.4.3. nsp7, nsp8 ou nsp8L7 de SARS-CoV et SARS-CoV-2

Après la transformation séparée des plasmides pQE30 (étiquette 6 histidines en N-terminal) codant pour nsp7, nsp8 ou nsp8L7 dans les cellules C3037 (NEB réf. C3037), les cultures ont été réalisées à 37 °C. Lorsque la DO600 a atteint entre 0,6 et 0,8, une induction avec 100 µM d'IPTG a été effectuée. Après l'induction, la culture a été poursuivie à 17 °C sous agitation (200 rpm). Les cellules ont été récoltées après 18 à 20 heures de culture par centrifugation à 8 000 g à 4 °C pendant 10 minutes. Les culots secs ont ensuite été stockés à -80 °C. Le milieu LB (AthenaES #0103) a été utilisé.

Les culots secs ont été resuspendus dans 50 ml/L de culture de tampon A1 (50 mM Tris pH 8, 300 mM NaCl, 5 mM MgSO4, 10 % glycérol, 0,1 % Triton, 0,5 mM TCEP) avec l'ajout de lysozyme (1 mg/ml), DNase (10 µg/ml), PMSF (1 mM), et benzamidine (0,2 mM). Le mélange a été laissé sous agitation magnétique pendant 45 minutes à 4 °C, puis soniqué pendant 45 secondes (1 s ON ; 1 s OFF à 45 % d'amplitude) x3, avant centrifugation pendant 30 minutes à 30 000 g. Le surnageant a été récupéré et placé sous agitation magnétique pendant 1 heure avec 1,5 ml de résine Talon Superflow (#28957502) par litre de culture, préalablement équilibrée avec le tampon A1.

Une purification en batch a été réalisée avec plusieurs étapes de lavage correspondant à 5 volumes de colonne (CV). D'abord, avec le tampon W1 (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol, 10 mM imidazole, 0,5 mM TCEP), puis avec le tampon W2 (50 mM Tris pH 8, 1 M NaCl, 10 % glycérol, 5 mM imidazole, 0,5 mM TCEP), suivi d'un lavage supplémentaire avec le tampon W1 (50 mM Tris pH 8, 300 mM NaCl, 5 mM imidazole, 0,5 mM TCEP) et enfin avec le tampon W2 (50 mM Tris pH 8, 1 M NaCl, 10 mM imidazole, 0,5 mM TCEP), avant d'effectuer l'élution avec le tampon d'élution (50 mM Tris pH 8, 300 mM NaCl, 300 mM imidazole, 0,5 mM TCEP).

Le pic d'élution a été placé dans un sac de dialyse (Spectrumlabs #132655) avec l'ajout de 1 mg de protéase TEV pour 10 mg de protéine purifiée, puis placé dans un tampon de dialyse (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol, 0,5 mM TCEP) sous faible agitation pendant une nuit. Le lendemain, une contre-purification sur la résine Talon Superflow (#28957502) a été réalisée, puis l'éluat a été concentré avec un concentrateur Amicon Ultra 15 (#UFC903024). Après centrifugation à 16 000 g pendant 5 minutes à 4 °C, la solution a été injectée sur une colonne d'exclusion de taille Superdex S75 16/60 (#28989333), préalablement équilibrée avec le tampon GF (25 mM Hepes pH 8, 300 mM NaCl, 5 mM MgCl2, 0,5 mM TCEP). Le pic de chromatographie d'exclusion de taille a été concentré avec un Amicon Ultra 4 (#UFC800324) jusqu'à une concentration d'environ 800 µM. Les protéines purifiées

ont été flash-congelées dans de l'azote liquide, puis conservées à -80 °C pour une utilisation future.

2.4.4. nsp14 et nsp10 de SARS-CoV-2

Après transformation des plasmides pET28 codant pour nsp14 ou nsp10 dans les cellules C2566 (NEB réf. C2566), les cultures ont été réalisées à 37 °C. Lorsque la DO600 a atteint entre 0,4 et 0,6, une induction a été effectuée avec 100 µM d'IPTG. Après induction, la culture a été poursuivie à 18 °C sous agitation (200 rpm). Les cellules ont été récoltées après 18 à 20 heures de culture par centrifugation à 8 000 g à 4 °C pendant 10 minutes, puis les culots secs ont été stockés à -80 °C. Le milieu LB (AthenaES #0103) a été utilisé.

Les culots secs ont été resuspendus dans 50 ml/L de culture dans le tampon A1 (50 mM Tris pH 8, 300 mM NaCl, 5 mM MgSO4, 10 % glycérol), auquel ont été ajoutés 0,1 mg/ml de lysozyme, 10 µg/ml de DNase, et 1 mM de PMSF. Le surnageant a ensuite été placé sous agitation magnétique pendant 30 minutes à 4 °C, avant d'être soniqué pendant 45 secondes (2 s ON ; 2 s OFF à 40 % d'amplitude) x3. Le mélange a été à nouveau agité pendant 30 minutes à 4 °C, puis du sel en poudre a été ajouté pour atteindre une concentration finale de 1 M. Après 30 minutes supplémentaires d'agitation à 4 °C, la solution a été centrifugée pendant 30 minutes à 30 000 g. Le surnageant a été récupéré et placé sous agitation magnétique pendant 30 minutes avec 1 ml de résine Talon Superflow (#28957502) par litre de culture, préalablement équilibrée avec le tampon A1.

Une purification en batch a ensuite été réalisée, avec un lavage au tampon W (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol, 5 mM imidazole). Le tampon d'élution (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol, 250 mM imidazole) a ensuite été incubé pendant 10 minutes sur la résine avant d'être récupéré. Le pic d'élution a été concentré à l'aide d'un Amicon Ultra 15 (#UFC901024) jusqu'à une concentration de 4 ou 8 mg/ml, selon la protéine (nsp14 ou nsp10), puis centrifugé à 16 000 g pendant 5 minutes à 4 °C avant d'être injecté sur une colonne d'exclusion de taille Superdex S75 16/60 (#28989333), préalablement équilibrée avec le tampon GF2 (25 mM Hepes pH 8, 300 mM NaCl, 10 % glycérol). Selon l'utilisation future du lot de protéines (cristallographie et BLI ou essais biophysiques et enzymatiques), 5 mM de CaCl2 ou 5 mM de MgCl2 ont été ajoutés respectivement. Dans le cas des batch pour le BLI, le glycérol était aussi retiré du tampon de purification.

Le pic de la chromatographie d'exclusion de taille a été concentré à l'aide d'un Amicon Ultra 4 (#UFC801024) jusqu'à environ 4 mg/ml (environ 72 µM) pour nsp14 et jusqu'à environ 8 mg/ml (environ 400 µM) pour nsp10. La nsp14 ainsi purifiée a été en partie utilisée directement pour la cristallographie ou le SEC-MALS, tandis que le reste des échantillons a été flash-congelé dans de l'azote liquide et conservé à -80 °C pour

des essais biophysiques et enzymatiques. Du TCEP à 0,5 mM a été ajouté à tous les tampons au cours de la purification.

Pour la cristallographie, une étape de clivage à la TEV a été ajoutée après la purification en batch sur la résine Talon Superflow (#28957502) et concentration avec l'Amicon Ultra 15 (#UFC901024). Dans ce cas, 1 mg de protéase TEV a été ajouté pour 10 mg de protéine purifiée, puis le mélange a été placé dans un sac de dialyse (Spectrumlabs #132700) dans le tampon de dialyse (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol) et incubé sous faible agitation pendant une nuit. Le lendemain, une contre-purification sur la résine Talon Superflow (#28957502) a été réalisée, suivie de l'étape de gel filtration et de concentration de l'échantillon, comme décrit précédemment. Du TCEP à 0,5 mM a également été ajouté à tous les tampons tout au long de la purification.

2.4.5. VHH

Après transformation séparée des plasmides pET24 dans les cellules C2566 (NEB réf. C2566), les cultures ont été réalisées à 37 °C. Lorsque la DO600 a atteint entre 0,4 et 0,6, l'induction a été effectuée avec 1 mM d'IPTG. Après l'induction, la culture a été poursuivie à 28 °C sous agitation (200 rpm). Les cellules ont été récoltées après 18 à 20 heures de culture par centrifugation à 8 000 g à 4 °C pendant 10 minutes, puis les culots secs ont été stockés à -80 °C. Le milieu TB (AthenaES #0104) a été utilisé.

Les culots secs ont été resuspendus dans 50 ml/L de culture dans le tampon A (20 mM Hepes pH 7,5, 150 mM NaCl, 5 mM imidazole) avec l'ajout de lysozyme (0,1 mg/ml), DNase (10 µg/ml) et PMSF (1 mM). Le mélange a ensuite été placé sous agitation magnétique pendant 30 minutes à 4 °C, puis soniqué pendant 45 secondes (2 s ON ; 2 s OFF à 40 % d'amplitude) x3. Le mélange a été de nouveau agité pendant 30 minutes à 4 °C, puis centrifugé pendant 30 minutes à 30 000 g. Le surnageant a été utilisé pour être injecté sur une colonne HisTrap HP 5 ml (Cytiva #17524701) préalablement équilibrée avec le tampon A et monitorée sur un système ÄKTA Pure. Une étape de lavage de 5 volumes de colonne (CV) a été réalisée avec 10 % de tampon B (20 mM Hepes pH 7,5, 150 mM NaCl, 500 mM imidazole), suivie d'un gradient d'élution allant jusqu'à 100 % sur 10 CV.

Le pic d'élution a été concentré à l'aide d'un Amicon Ultra 15 (#UFC901024) jusqu'à environ 6 mg/ml, puis centrifugé à 16 000 g pendant 5 minutes à 4 °C avant d'être injecté sur une colonne d'exclusion de taille Superdex S75 16/60 (#28989333) préalablement équilibrée avec le tampon GF1 (20 mM Hepes pH 7, 150 mM NaCl). Le pic de la chromatographie d'exclusion de taille a ensuite été concentré avec un Amicon Ultra 4 (#UFC801024) jusqu'à environ 6 mg/ml (environ 400 µM), et conservé à 4 °C. Du TCEP à 0,5 mM a été ajouté à tous les tampons au cours de la purification.

2.4.6. nsp12 de MERS

Plusieurs essais de production ont été réalisés avec les deux plasmides (pJ404, étiquette 6 histidines en C-terminal, ou pQE30, étiquette 6 histidines en N-terminal) codant pour la nsp12 de MERS, ainsi qu'avec différentes souches cellulaires, comme décrit dans la section « 2.5. Essais d'expression à haut débit ». Les deux meilleurs condition retenue pour le scale-up est celle utilisant le plasmide pQE30 (étiquette 6 histidines en N-terminal), transformé dans les cellules C2523 (NEB réf. C2523, n'a pas de gène répresseur LacI) et C3037 (NEB réf. C3037, a le gène répresseur LacI). Les cultures pour cette condition ont été réalisées à 24 °C sous agitation continue à 200 rpm dans du milieu LB (AthenaES #0103). Lorsque la DO600 a atteint entre 0,2 et 0,3, une induction avec 0,25 µM d'IPTG a été effectuée. Les cellules ont été récoltées après 18 à 20 heures de culture par centrifugation à 8 000 g à 4 °C pendant 10 minutes. Les culots de cultures de 500 ml ont été resuspendus dans un tampon de lyse (50 mM Tris pH 8, 300 mM NaCl, 5 mM MgSO4, 10 % glycérol, 0,5 mM TCEP), puis transférés dans des tubes Falcon de 50 ml et stockés à -80 °C.

Les culots humides ont été placés à -20 °C la veille de la purification. Le lendemain, les éléments suivants ont été ajoutés à chaque culot humide : 1 mg/ml de lysozyme, 20 µg/ml de DNase, 1 mM de PMSF, et 0,2 mM de benzamidine. Les tubes Falcon ont ensuite été placés dans un bain-marie à 37 °C pendant 10-15 minutes pour décongélation. Les surnageants ont été réunis et soniqués pendant 1 minute (1 s ON ; 1 s OFF à 40 % d'amplitude) x1. Le mélange a ensuite été placé sous agitation magnétique pendant 30 minutes à 4 °C, puis la quantité nécessaire de sel en poudre a été ajoutée pour atteindre une concentration finale de 1 M, suivie de 30 minutes supplémentaires d'agitation à 4 °C. Le lysat a été dilué par l'ajout de deux volumes de tampon de lyse avec 1 M de NaCl avant d'être injecté sur une colonne HisTrap HP 5 ml (Cytiva #17524701), préalablement équilibrée avec le tampon A (50 mM Tris pH 8, 1 M NaCl, 10 % glycérol, 5 mM imidazole, 0,5 mM TCEP) et monitorée sur un système ÄKTA Pure. Une étape de lavage de 5 volumes de colonne (CV) a été effectuée avec 10 % de tampon B (20 mM Hepes pH 7,5, 300 mM NaCl, 500 mM imidazole), suivie d'un gradient d'élution progressant jusqu'à 100 % sur 10 CV.

Le pic d'élution a été concentré avec un concentrateur Amicon Ultra 15 (#UFC903024) jusqu'à environ 3,5 mg/ml, puis centrifugé à 16 000 g pendant 5 minutes à 4 °C avant d'être injecté sur une colonne d'exclusion de taille Superdex S200 16/60 (#28989335), préalablement équilibrée avec le tampon de chromatographie par filtration sur gel (GF) (25 mM Hepes pH 8, 300 mM NaCl, 5 mM MgCl2, 10 % glycérol). Le pic de chromatographie GF a ensuite été concentré avec un Amicon Ultra 4 (#UFC803024), flash-congelé dans de l'azote liquide, puis stocké à -80 °C pour une utilisation future.

2.4.7. nsp8, nsp7, de MERS

Après transformation séparée des plasmides pQE30 (étiquette 6 histidines en N-terminal) codant pour nsp7, nsp8 ou nsp8L7 dans les cellules C3037 (NEB réf. C3037), les cultures ont été réalisées à 37 °C. Lorsque la DO600 a atteint entre 0,6 et 0,8, les cultures ont été retirées de l'incubateur et refroidies pendant 15 minutes à température ambiante avant d'induire l'expression avec 100 µM d'IPTG et 2 % d'éthanol. Après l'induction, la culture a été poursuivie à 17 °C sous agitation (200 rpm). Les cellules ont été récoltées après 18 à 20 heures de culture par centrifugation à 8 000 g à 4 °C pendant 10 minutes, puis les culots secs ont été stockés à -80 °C. Le milieu LB (AthenaES #0103) a été utilisé.

Les étapes de purification suivantes sont identiques à celles décrites pour la purification de la nsp12 de SARS-CoV dans la section « 2.4.1. nsp12 de SARS-CoV » jusqu'à la fin de l'IMAC (chromatographie d'affinité sur métal immobilisé). Après cette étape, le pic d'élution a été placé dans un sac de dialyse (Spectrumlabs #132655), auquel 1 mg de protéase TEV a été ajouté pour 10 mg de protéine purifiée. Le sac a ensuite été placé dans un bécher contenant du tampon de dialyse (50 mM Tris pH 8, 300 mM NaCl, 10 % glycérol, 0,5 mM TCEP) et agité doucement pendant la nuit.

Le lendemain, une contre-purification sur la résine Talon Superflow (#28957502) a été réalisée, puis l'éluat a été concentré à l'aide d'un Amicon Ultra 15 (#UFC903024), centrifugé à 16 000 g pendant 5 minutes à 4 °C, et injecté sur une colonne d'exclusion de taille Superdex S75 16/60 (#28989333) préalablement équilibrée avec le tampon GF (25 mM Hepes pH 8, 300 mM NaCl, 5 mM MgCl2, 0,5 mM TCEP). Le pic de la chromatographie d'exclusion de taille a été concentré à environ 800 µM à l'aide d'un Amicon Ultra 4 (#UFC800324). Les protéines purifiées ont été flash-congelées dans de l'azote liquide, puis conservées à -80 °C pour une utilisation future.

2.5. Essais d'expression à haut débit

Toutes les étapes ont été réalisées dans des plaques 24 puits (DW24) comme dans le protocole standard du laboratoire (Saez *et al*. 2014) et les modifications sont décrite ci dessous.

Les deux plasmides nsp12 de MERS (pJ404 ou pQE30) décrit dans la partie « 2.2.1. Constructions plasmidiques » ont été transformé dans 12 souches de cellules *E.coli*. Puis pour la production, chaque DW24 est divisé en deux, 12 x 2 mL pour ainsi avoir des culture de 24ml final dans le milieu choisi (LB, TB, ou NZ) et supplémenté avec l'antibiotique correspondant au plasmide. 500µl de préculture est alors ajouté. Un choix de 48 duo cellules/ plasmide ont été sélectionné (Voir Annexes C1). La culture est réalisé à 25°C dans un incubateur de secousses de Microtron (INFORS-HT, Suisse ; 600 rpm). L'expression des plasmides est lancé environ 5h après le début

de la culture, soit à une Do600 d'environ 0.4, avec 0.5µM d'IPTG. La suite de la culture est réalisé à 25°C en continue pendant 16h. Les cellules sont ensuite été collectées par centrifugation et chaque puits est resuspendu dans 0,5 mL de tampon de lyse (50 mM Tris, 300 mM NaCl, 10 mM Imidazole, pH 8,0, 0,25 mg/mL Lysozyme) et 15 min agité à 20°C dans un incubateur de secousses de Microtron (INFORS-HT, Suisse ; 800 rpm) avant que les plaques DW24 soient congelées à une température de -20°C.

2.6. Essais de purification à haut débit

Les 24 plaques DW24 correspondant aux 48 duo cellules/plasmide ont été décongelés, pendant 10 min à 37°C au bain-marie , avant d'être placé 15 min sous agitation à 20°C dans un incubateur de secouage de Microtron (INFORS-HT, Suisse ; 800 rpm). Cela permet aux cultures d'etre lysées par le lysozyme et la solution est devenue visqueuse. Après l'ajout de 10 pg/mL d'ADNse et de 20 mM MgSO 4 dans chaque puits et laissé à nouveau 10 min sous agitation à 20°C. Chaque lysat bactérien (12 x 0,5 mL) sont alors regroupé et transférées dans un nouveau DW24 pour avoir 24 lysats distincts de 6 mL par DW24 (soit plus que 2 DW24 au total). Pour assurer une lyse cellulaire complète, le DW24 a été soniqué dans un sonicateur de plaques (processeur ultrasonore XL, Misonix Inc., USA) pendant 5 min (puissance 5, 30 s ON/OFF cycles). La purification a été réalisée par chromatographie d'affinité des métaux immobilisés (IMAC) comme décrit dans l'article (Sequeira *et al*. 2017), où toutes les étapes ont été automatisées sur un robot Tecan Freedom EVO 200 (Suisse) contenant un collecteur sous vide. Brièvement, les 6 mL de lysates de cellules brutes ont été incubés avec 4 x 200 uL de résine Ni-Sépharose 6 Fast Flow (GE Healthcare, 17-5318-02) lié avec du nickel. Puis le mélange surnageant/ résine est transférés (4 puits à la fois) dans des plaques filtrantes à 96 puits (20 µm ; Macherche-Nagel). Les puits ont été lavés deux fois avec 800 pL de tampon A (50 mM Tris, 300 mM NaCl, 10 mM Imidazole, pH 8) suivis de trois lavages avec 800 µL de tampon B(Tris 50 mM, NaCl 300 mM, Imidazole 50 mM, tampon pH 8). Les protéines ont ensuite été éluées de la résine avec 1 mL de tampon d'élution (Tris mM, NaCl 300 mM, Imidazole 250 mM, pH 8). La collecte des élutions est faite à nouveau dans une DW24, mettant en commun les 4 x 1 mL d'élution d'une protéine purifié dans un seul puits du DW24. Cette procédure a été reproduite 2 fois en dans la journée pour purifier 48 conditions proteines/milieu.

2.7. Isolation ADNc des VHHs

Un lama glama a été immunisé avec la protéine purifiée nsp14 (200µl à 1mg/ml). Après trois injections à 21 jours d'intervalle, une population lymphocytaire enrichie a été isolée par centrifugation en gradient de Pancoll. L'ADNc premier brin a été synthétisé à partir de l'ARN total purifié par transcription inverse. Une stratégie de trois

PCR a été utilisée pour amplifier spécifiquement le fragment du gène vhh à partir de l'ADNc.

2.8. Phage display

Les fragments PCR ont été clonés dans le vecteur phagemid pHEN4 pour créer une banque de phages VHH (décrit en « 2.2.1.Constructions plasmidiques »). Le phage display a été utilisé pour enrichir les nanocorps spécifiques à la cible contre nsp14.

2.9. Test Elisa

Après trois cycles d'amplification, des clones individuels ont été sélectionnés et utilisés pour un dernier test de liaison ELISA. 24 clones ont été envoyés pour séquençage (Eurofins Genomics), ce qui a permis d'obtenir 3 familles de nanocorps.

2.10. Assemblage & purification des complexes avec nsp14

Les complexes destinés à la cristallographie ou au SEC-MALS ont été assemblés de la manière suivante. La protéine fraîchement purifiée a été mélangée avec les autres protéines (nsp10 ou VHH) dans un ratio molaire de 1:1,2 à température ambiante pendant 20 minutes. Le complexe a ensuite été centrifugé à 16 000 g pendant 5 minutes à 4 °C avant d'être injecté sur une colonne d'exclusion de taille Superdex S200 10/300 GL (#28990944), préalablement équilibrée avec le tampon GF3 (25 mM Hepes pH 8, 300 mM NaCl, 5 mM CaCl2, 0,5 mM TCEP). Le pic de chromatographie d'exclusion de taille a été concentré avec un Amicon Ultra 4 (#UFC801024) jusqu'à une concentration d'environ 4 mg/ml.

2.11. BLI (Bio-Layer Interferometry)

Le vhh a d'abord été biotinylée à l'aide du kit EZ-Link NHS-PEG4-Biotin (Perbio Science, France). La réaction a été stoppée en éliminant l'excès de biotine à l'aide d'une colonne Zeba Spin Desalting (Perbio Science, France). Les études BLI ont été réalisées dans des plaques noires à 96 puits (Greiner) à 22°C à l'aide d'un OctetRed96 (ForteBio, USA). Les pointes de biocapteurs en streptavidine (#18-5019, ForteBio, USA) ont d'abord été hydratées avec 0,2 ml de tampon cinétique (CB)(25mM Hepes pH8, 300mM NaCl, 5mM CaCl2, 0.05% Tween 20) pendant 20 min, puis chargées de

vhh biotinylée (5 µg/ml dans CB). L'association du vhh avec nsp14 (50nM) a été suivie pendant 400 secondes, et de 1200s pour l'étape de dissociation dans CB.

Pour les expériences de compétitions les vhh était d'abord fixé sur la sonde avant d'ajouté nsp14 ou le complexe nsp14/10 (préassemblé dans un ratio de 1 :1.2 pendant 20 minutes avant l'expérience) pendant 400 secondes, avant de procédé à la phase de dissociation soit avec nsp10 dans du CB pour le premier cas, soit dans du CB pour le second, pendant 1200 secondes. Les données ont été analysé sur le logiciel ForteBio Data Analysis.

2.12. Caractérisation en double référent des VHH1 et VHH2

La caractérisation en double référent pour calculer le Kd apparent des vhh1 & 2 a été fait via le protocole de mise en place précédent et la méthodologie décrit dans la littérature (Orthwein *et al*. 2021) avec une gamme de nsp14 réalisé via des dilution en série, allant de 100nM à 1.56nM. Les données ont été analysé sur le logiciel ForteBio Data Analysis.

2.13. SEC-MALS

L'analyse par SEC-MALS a été réalisée à l'aide d'un système HPLC (Thermo Scientific) équipé d'une colonne Superose6 10/300 GL (CYTIVA #29091596) pour les complexes nsp14/nsp10 et nsp14/nsp10/VHH, ou d'une colonne Superdex S200 10/300 GL (#28990944) pour nsp14, VHH, et nsp14/VHH. Ce système était couplé à un dispositif de « multi angle light scattering » à 8 angles à l'aide d'un photomètre laser DAWN (MALS ; Dawn 8, Wyatt Technology) et à un détecteur d'indice de réfraction différentiel Optilab TrEX (Wyatt Technology). Dans un premier temps, la colonne a été équilibrée avec le tampon (25 mM Hepes pH 8, 300 mM NaCl, 0,5 mM TCEP) pour les deux premiers complexes, ou avec le tampon (25 mM Hepes pH 7, 300 mM NaCl, 5 mM CaCl2, 0,5 mM TCEP) pour les suivants, pendant plusieurs heures, jusqu'à ce que la ligne de base du signal Optilab TrEX soit stable. Ensuite, 80 µL de chaque complexe purifié a été injecté à un débit de 0,6 ml/min, et les données de diffusion de la lumière ainsi que de concentration ont été analysées dans ASTRA pour déterminer la masse moléculaire et la distribution de taille. L'élargissement et l'alignement des bandes inter-détecteurs ont été appliqués à l'aide des algorithmes d'ASTRA 7.3.2 (Wyatt Technology, USA), en utilisant la protéine standard BSA, conformément aux méthodes recommandées par le fabricant.

2.14. SEC-SAXS

2.14.1.1. Préparation des échantillons

Les complexes sont assemblé directement au Synchrotron Soleil sur la ligne Swing. Les protéines conservé sur glace durant le transport, sont assemblé dans le ration 1 de nsp14 pour 1.1 des autres protéines de l'échantillon, puis à RT pendant 20 minutes avant d'être centrifugé (16 000g) 5 minutes. Les échantillons sont ensuite transféré dans des vials et placé dans l'HPLC à 20°C.

2.14.1.2. Diffraction des rayons X aux petits angles

Le système est préalablement équilibré avec le tampon de GF3 (25mM Hepes pH8, 300mM NaCl, 5mM CaCl2, 0.5mM TCEP). L'HPLC injecte alors l'échantillon sur une colonne de gel filtration BioSec 3-300 (Agilent, ref. AG5190-2513), puis l'élution passe dans le capillaire, entre la source de rayons x de 12keV et le détecteur EigerX4M.

2.14.1.3. Extraction et analyse des données

Le traitement des données a été effectué avec Foxtrot, une application graphique basée sur Java et développée à SOLEIL. Les données ont ensuite été analysées avec BioXTAS RAW.

2.15. Cristallographie

2.15.1.1. Conditions de cristallographie criblées

L'étape de criblage des conditions de cristallogenèse pour les complexes de protéines nsp14/nsp10, nsp14/vhh1 et nsp14/nsp10/vhh1 a été réalisée sur des microplaques 96 puits (Swissci 3 Lens Crystallisation Plate). Les criblages de tampons de cristallogenèse testés proviennent de Molecular Dimensions (JCSG Plus, PACT Premier, ProPlex, Structure Screen 1 and 2). Les tampons des différents criblages ont été disposés dans les 96 réservoirs des plaques par un robot TECAN (Freedom EVO (TECAN)), puis les mélanges de protéines à environ 4 mg/ml ont été déposés dans les trois puits de chaque réservoir (ratios 300-200-100 nl de protéines : 100 nl de tampon) grâce à un robot Mosquito (SPT Labtech).

2.15.1.2. Conditions de cristallogenèse finale

De plus grandes plaques de cristallographie (Rigaku CombiClover 500 (CJR) #1009542) ont été utilisées pour obtenir les cristaux destinés à la diffraction. Le complexe est dilué avec le tampon GF3 si nécessaire, à une concentration de 4 mg/ml,

puis 1 ou 2 µL sont déposés dans chaque puits de la plaque avant d'être mélangés avec 1 ou 2 µL de tampon de cristallographie (22,5-26,25 % p/v PEG 3350, 0,1 M Bis-Tris propane pH 6,25-7, 0,2 M fluorure de sodium). La plaque est ensuite placée à 20 °C. Les cristaux commencent à apparaître au bout de 2 jours et sont récupérables après 4 à 6 jours.

2.15.1.3. Cryo-protection

Les cristaux sont cryo-protégés avec 40 % de glycérol final.

2.15.1.4. Diffraction

Les cristaux ont été diffracté au synchrotron soleil sur la ligne proxima 1 avec un faisceau de rayon x de 12.4 KeV (λ = 0.978565Å), et collecté à l'aide d'un détecteur EIGER-X 16M (Dectris Ltd).

2.15.1.5. Résolution de la structure

Les images ont été traitées avec AutoPROC de Global Phasing, puis le calcul des phases pour le remplacement moléculaire a été effectué avec Phaser. Les affinement ont ensuite été réalisés avec Buster de Global Phasing, et le model visualisé et optimisé en utilisant Coot.

2.16. Test d'activité ExoN

Les tests de l'activité ExoN de nsp14 du SARS-CoV-2 ont été réalisés avec un ratio molaire de 1:4 pour nsp10 (Bouvet, Imbert, *et al*. 2012) et de 0.05µM final de concentration de nsp14. Pour le VHH, un ratio de de 1:4 est utilisé. Le volume réactionnel total est de 28 µL, dans un tampon contenant 40 mM de Tris pH 8, 5 mM de DTT, 2 mM de MgCl2, et 0,1 mM de Hairpin RNA (HP_A, 5′6FAM-CGAGAGACUCGCGUAGUUUCUACGCG-3′).

Les contrôles positifs ont consisté à réaliser la réaction avec nsp14 et nsp10 uniquement. Les contrôles négatifs ont consisté à remplacer le MgCl2 par 5 mM d'EDTA dans le tampon. La préparation de l'essai enzymatique a consisté à préparer un prémix avec les enzymes d'un côté et un second prémix avec la matrice de l'autre. Les deux prémix ont ensuite été mélangés pour lancer la réaction, incubée à 37 °C, puis un échantillon de 4 µL a été prélevé au temps souhaité. La réaction a été stoppée en mélangeant l'échantillon avec 1 µL de FBD 5x (formamide, 25 mM EDTA, 0,05 % de bleu de bromophénol).

Les expériences ont été analysées sur gel de polyacrylamide dénaturant (20 % d'acrylamide, 7 M d'urée, tampon TBE), et 5 µL d'échantillons ont été déposés après avoir été chauffés à 70 °C pendant 10 minutes, puis refroidis sur glace pendant 2 minutes avant le chargement sur le gel. Les gels ont été visualisés à l'aide d'un Typhoon FluorImager.

2.17. Test d'activité MTase by filter binding assay (FBA)

Les tests de l'activité N7-MTase de nsp14 du SARS-CoV-2 ont été réalisés avec différents assemblages de protéines afin d'observer l'influence possible de ces protéines sur l'activité. Ces essais ont été effectués dans un mélange réactionnel contenant 40 mM de Tris-HCl pH 8.0, 1 mM de DTT, 1 mM de MgCl2, 2 µM de SAM, et 0.1 µM de 3H-SAM (Perkin Elmer), en présence de 0.7 µM d'ARN synthétique GpppAC4 (Ahmed-Belkacem *et al*. 2022).

Les enzymes ont d'abord été mélangées à une concentration finale de 10 nM pour nsp14 et 40 nM pour le VHH, avant l'ajout du substrat ARN et de la SAM dans un volume final de 210 µL. Le mélange a ensuite été incubé à 30°C. Plusieurs aliquotes (35 µL) du mélange réactionnel ont été prélevées pour analyse et la réaction a été arrêtée par dilution au dixième dans de l'eau glacée.

Les échantillons ont ensuite été transférés sur des filtres diéthylaminoéthyle (DEAE ; Perkin Elmer) à l'aide d'un Filtermat Harvester (Packard Instruments). Les filtres retenant l'ARN ont été lavés deux fois avec du formiate d'ammonium 10 mM pH 8,0, deux fois avec de l'eau, puis une fois avec de l'éthanol. Ils ont été imbibés de liquide scintillant (Perkin Elmer), et le transfert de 3H-méthyl sur les substrats d'ARN a été mesuré à l'aide d'un compteur à scintillation liquide Wallac MicroBeta TriLux (Perkin Elmer).

Chapitre 3 : Étude structurale et fonctionnelle des nsp12 des Coronavirus

3. Étude structurale et fonctionnelle des nsp12 des Coronavirus

La nouvelle pandémie de COVID-19 a souligné l'importance du développement de traitements antiviraux ciblant le SARS-CoV-2 et les Coronavirus. Pour cela, les complexes de réplication et de transcription (RTC) des trois principaux Coronavirus doivent pouvoir être étudiés *in vitro*. Plusieurs systèmes de production de protéines existent à cet effet. Le choix d'une production en système bactérien étant plus pratique que la production en système eucaryote, notre laboratoire a choisi ce système pour débuter la production de ces enzymes. Bien que la production de nsp7 et nsp8 se fasse facilement dans ce type de système, la nsp12, étant une grande enzyme fixant l'ARN, a nécessité plus de travail pour sa production dans ce système. Un protocole pour la production de la nsp12 de SARS-CoV avait commencé à être développé avant mon arrivée, et j'ai travaillé à l'amélioration de ce protocole de production, puis à la mise en place de celui pour la purification de la nsp12 de SARS-CoV-2, ainsi qu'aux différents essais pour celle du MERS.

3.1. Études du RTC de SARS-Cov-2

La production de ces protéines nous a permis d'étudier plusieurs candidats médicaments. Un très bon inhibiteur potentiel, dont la prodrogue est l'AT-527, a été sélectionné afin d'étudier plus en profondeur ses mécanismes d'action au niveau moléculaire, notamment par la production de mutants des enzymes de ce complexe, et via la détermination de la structure du RTC de SARS-CoV-2 en présence de cet inhibiteur.

3.1.1. Présentation de l'article 1 « A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase »

Titre de l'article : A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase.

Auteurs : Shannon A, Fattorini V, Sama B, Selisko B, Feracci M, Falcou C, Gauffre P, El Kazzi P, Delpal A, Decroly E, Alvarez K, Eydoux C, Guillemot JC, Moussa A, Good SS, La Colla P, Lin K, Sommadossi JP, Zhu Y, Yan X, Shi H, Ferron F, Canard B.

Référence : Nat Commun. 2022 Feb 2;13(1):621. doi: 10.1038/s41467-022-28113-1. PMID: 35110538; PMCID: PMC8810794.

Dans cet article (Annexe, Article 1), nous explorons les mécanismes moléculaires de l'inhibition de nsp12 par l'AT-527, un analogue de la guanosine converti en sa forme active, l'AT-9010 (Figure 27 a), chez le Coronavirus de type 2 du syndrome respiratoire aigu sévère (SARS-CoV-2).

Différentes analyses biochimiques de l'effet de l'AT-9010 sur l'activité de l'ARN polymérase virale dépendante de l'ARN (RdRp) semblent montrer un fort effet terminateur de chaîne de l'ARN lors de son incorporation. En effet, nous avons comparé l'incorporation et l'élongation d'AT-9010 par les RTC purifiés et recombinants du SARS-CoV et du SARS-CoV-2. En l'absence de GTP, l'AT-9010 est rapidement incorporé dans l'ARN viral, provoquant la terminaison immédiate de la synthèse de l'ARN (Figure 27 a). L'incorporation d'AT-9010 a également été comparée au Sofosbuvir triphosphate (STP), un analogue structurallement lié à l'uracile (Figure 27 b). Bien que le STP puisse être incorporé par nsp12 en remplacement de l'UTP, il n'est pas compétitif avec l'UTP, même à des concentrations 5 fois plus élevées (Figure 27 b), alors que pour l'AT-9010, même à des concentrations de 500 µM d'ATP, aucun allongement de la matrice au-delà de l'AT-9010 déjà incorporé n'est observé (Figure 28 a).

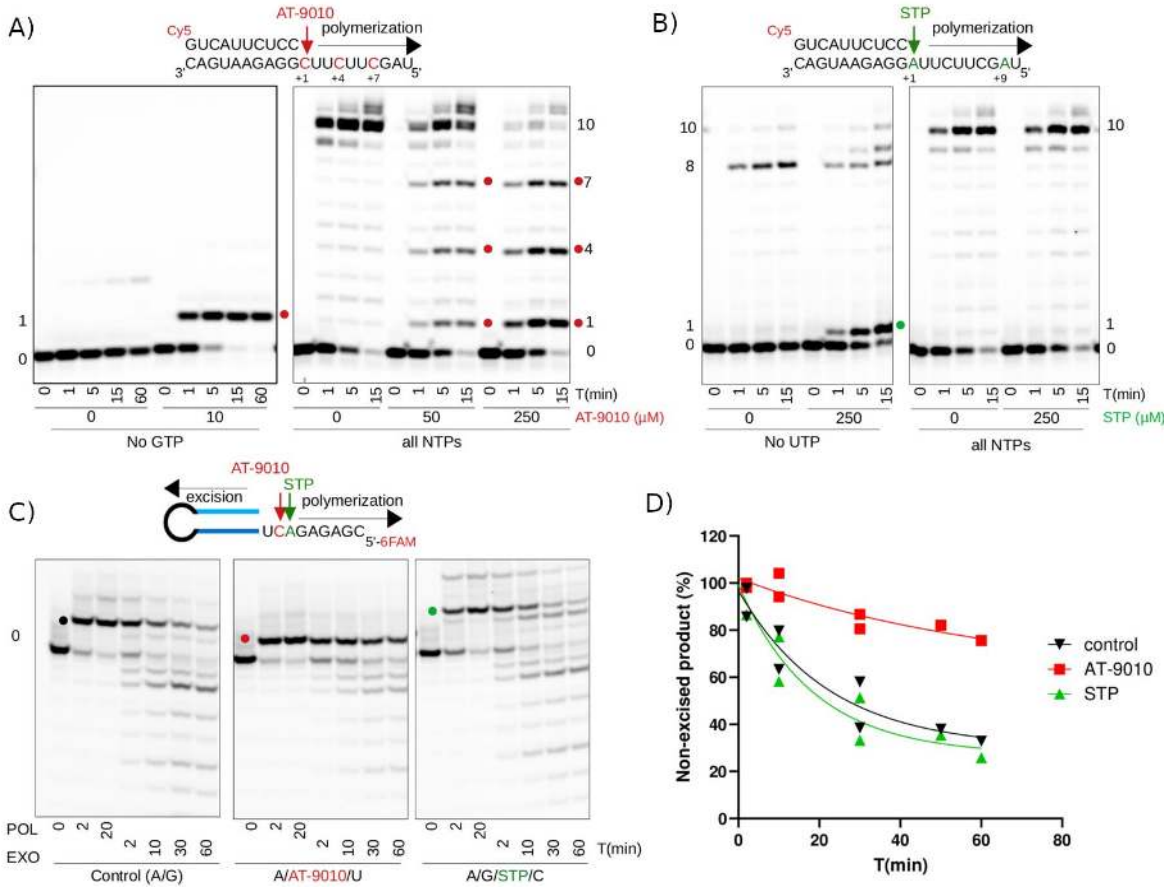


Figure 27 : Incorporation de l'AT-9010 et du Sofosbuvir triphosphate (STP) par le SARS-CoV-2 RTC, et l'excision par le complexe d'exonucléase de SARS-CoV-2 nsp14/10 (ExoN).

a) Chronologie de l'incorporation de l'AT-9010 (points rouges) en tant que substitut au GTP (panneau de gauche, 50 μM de chaque ATP, UTP et CTP), ou en compétition avec le GTP (panneau de droite, 50 μM de chaque NTP) aux concentrations indiquées d'AT-9010 (0 à 250 μM). Le facteur de préférence pour le GTP par rapport à l'incorporation d'AT-9010 est calculé en comparant la quantité d'insertion d'AT-9010 (points rouges) par rapport au produit de pleine longueur à deux concentrations et à trois points temporels. b) Chronologie de l'incorporation du triphosphate de sofosbuvir (STP ; points verts) en tant que substitut à l'UTP (panneau de gauche, 50 μM de chaque ATP, GTP et CTP), ou en compétition avec l'UTP (50 μM de chaque NTP) aux concentrations indiquées (0 ou 250 μM). Pour les deux cas a et b, les réactions ont été réalisées au moins en double à plusieurs concentrations de nucléotides/analogues pour deux substrats ARN différents, avec des résultats cohérents. c) Incorporation du contrôle ATP+GTP (noir, panneau de gauche), ATP + AT-9010 (rouge, panneau central) et ATP + GTP + STP (vert, panneau de droite) par le complexe polymérase (POL), suivie par la chronologie de l'excision avec l'ExoN de nsp14 (EXO). Pour les expériences avec AT-9010 et STP, le nucléotide suivant dans la séquence modèle pour l'incorporation a été également ajouté. d) Quantification du produit restant après l'excision par ExoN montrée en c, un gel représentatif des expériences réalisées en double pour deux ARN distincts. L'ARN marqué à la position 0 sur chaque gel correspond à la taille de l'amorce marquée par fluorescence montrée au-dessus de chaque gel, avant l'élongation.

Le blocage de la polymérase après l'insertion d'un nucléotide analogue (NA) de terminaison de chaîne peut permettre l'excision par l'exonucléase de relecture 3' à 5' nsp14/nsp10 (Ferron *et al*. 2018), atténuant potentiellement l'efficacité de l'AT-9010. Après l'incorporation de l'AT-9010 ou du STP dans l'ARN, ces derniers sont excisés par l'ExoN du SARS-CoV-2 (Figure 28 c). Il est toutefois intéressant de noter une différence d'environ 4,3 fois dans la résistance à l'excision de l'AT-9010 par rapport au STP lorsque les nucléotides analogues sont à l'extrémité 3' de l'ARN (Figure 28 c,d). Compte tenu de l'incorporation hautement compétitive de l'AT-9010 par rapport au GTP, même avec l'activité ExoN de nsp14/10 qui peut potentiellement exciser complètement ces NA, on peut présumer que l'AT-9010 ralentira probablement la réplication virale.

Au niveau structurale, une structure en cryo-EM à une résolution de 2,98 Å du RTC de SARS-CoV-2 (nsp12-nsp7-nsp82) avec ARN, montrant l'AT-9010 lié à trois sites de la nsp12, a été proposée (Figure 28 b). En effet, en plus d'avoir observé une molécule d'AT-9010 incorporée dans l'ARN (Figure 28 b, c, d) et une autre dans le site actif de la RdRP, une troisième molécule a été observée du côté du domaine nucléotidyltransférase associé au Nidovirus RdRp (NiRAN) au niveau N-terminal de la nsp12.

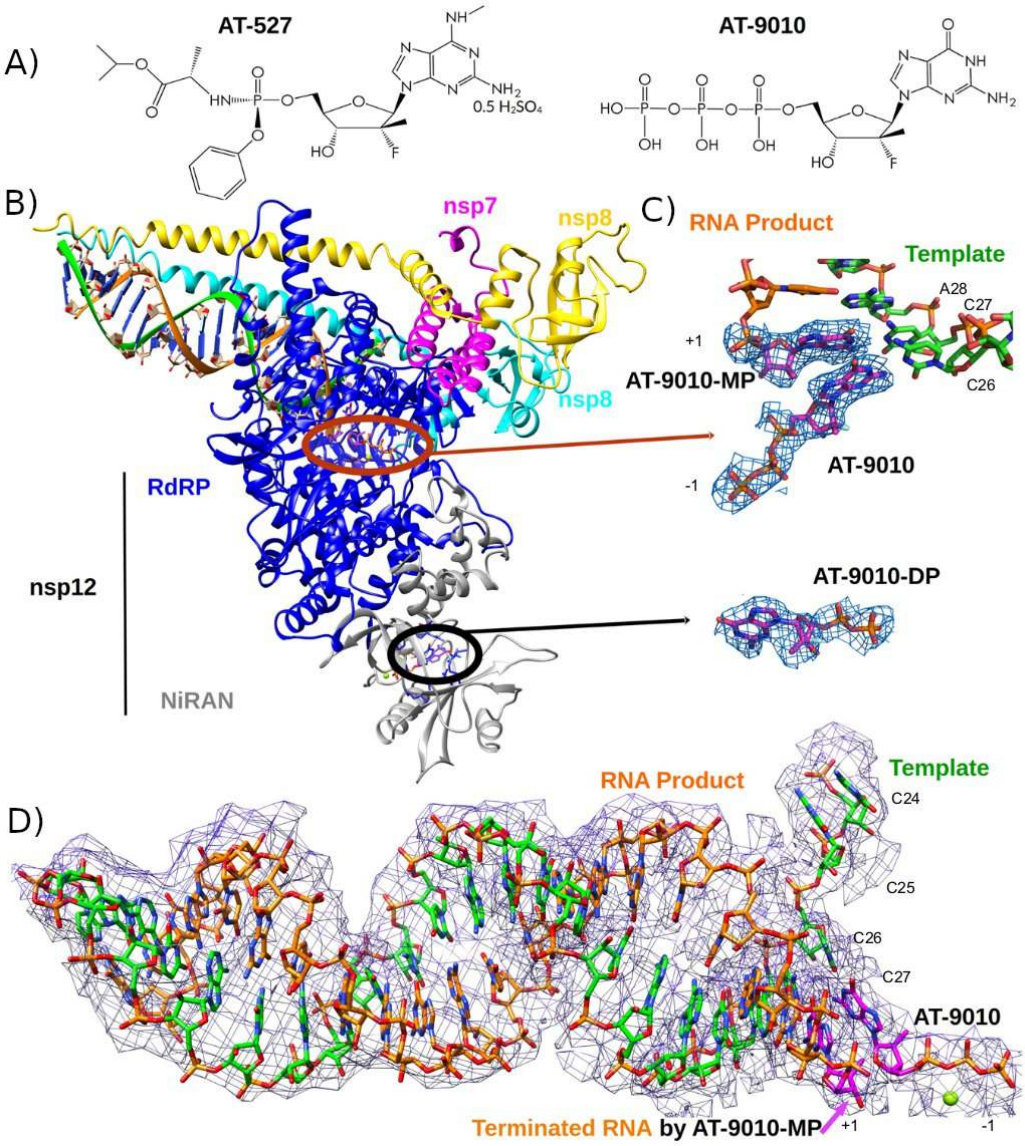


Figure 28 : Structure cryo-EM du RTC avec des molécules liées et des molécules AT-9010 et carte cryo-EM correspondante.

a) Structure de la prodrogue phosphoramidate analogique de la guanine AT-527 (à gauche) et de sa forme triphosphate active AT-9010 (à droite) suite à l'activation par les kinases cellulaires. b) Représentation en ruban et en bâtonnet de la structure cryo-EM du complexe nsp7-(nsp8)2-nsp12 : AT-9010-terminated-RNA : (AT-9010)2. L'ARN, l'AT-9010 et les protéines sont affichés avec les couleurs suivantes : ARN matrice, vert ; produit ARN, orange ; AT-9010, magenta ; nsp7, rose ; nsp81 et nsp82, jaune et cyan respectivement ; et nsp12 en gris et bleu pour les domaines NiRAN et RdRp respectivement. c) Les ovales brun et noir sont agrandis pour les domaines RdRp et NiRAN, respectivement, montrant la carte cryo-EM expérimentale autour de l'AT-9010. Dans le RdRp, un monophosphate d'AT-9010 (AT-9010-MP) est incorporé en position (+1), avec un AT-9010 entrant occupant la position (−1). Dans le domaine NiRAN, l'AT-9010 est lié sous sa forme diphosphate (AT-9010-DP). d) Carte cryo-EM expérimentale pour l'ARN (représentation en bâtonnet), montrant l'AT-9010-MP incorporé (+1) et l'AT-9010 entrant (−1). Les bases de l'ARN matrice impliquées dans

l'interaction avec l'AT-9010 sont numérotées de C24 à C27. Toutes les cartes cryo-EM sont représentées à 3.5σ, représentatives de la carte générale de l'ensemble du complexe à une résolution de 2.98 Å.

L'orientation de cette troisième molécule d'AT-9010 sous sa forme diphosphate (AT-9010-DP) est originale (Figure 29 c) comparée à ce qui a déjà été observé pour l'ADP (J. Chen *et al*. 2020) et le GDP (Yan *et al*. 2021 ; Figure 29 d). Cette configuration inversée de l'AT-9010-DP (Figure 29 e) est permise par la stabilisation de la base guanine de ce dernier grâce à des interactions hydrophobes et à des liaisons hydrogène avec Arg55, Thr120 et Tyr217 (Figure 29 a,b), des résidus conservés dans les séquences nsp12 des CoV, mais pas chez d'autres pseudo-kinases (Gao *et al*. 2020 ; Q. Wang *et al*. 2020). L'AT-9010-DP possède ainsi un mode de liaison unique, dû à la fois à la nature hydrophobe de la cavité du NiRAN et au ribose modifié, ce qui lui permet de se lier de manière stable dans le site actif et de l'inhiber.

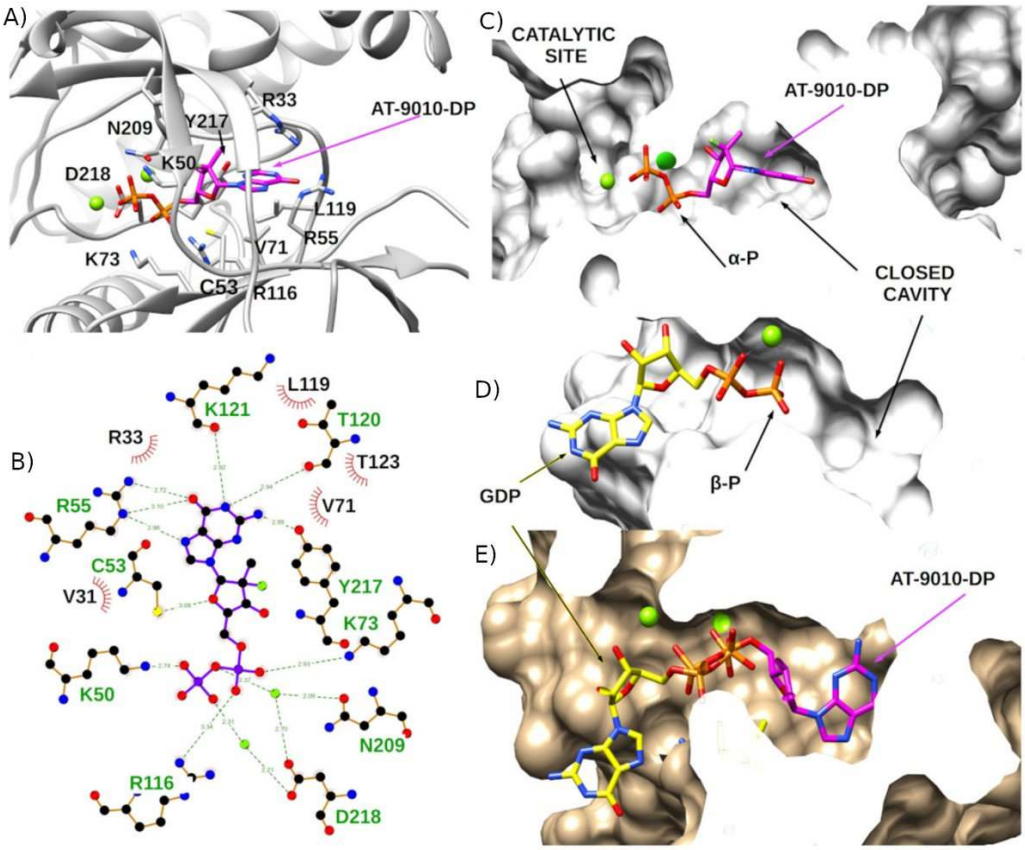


Figure 29 : Base structurale de l'inhibition NiRAN du SARS-CoV-2 par AT-9010.

a) Liaison de l'AT-9010 5'-diphosphate (AT-9010-DP) dans le domaine NiRAN. Les résidus en interaction sont étiquetés et représentés en bâtonnets. b) Représentation en Ligplot 2D des interactions détaillées de la liaison de l'AT-9010-DP dans la cavité NiRAN. c) Représentation en surface du NiRAN coupé orthogonalement en Z montrant l'AT-9010-DP englouti dans la cavité. d) Représentation en

surface du NiRAN coupé orthogonalement en Z montrant la liaison du GDP (PDB 7CYQ) ; l'orientation est tournée d'environ +90° par rapport à c. e) Superposition du GDP et de l'AT-9010-DP montrant le chevauchement du phosphate α de l'AT-9010 avec le phosphate β du GDP et du Mg2+ dans les deux structures, même orientation que d. Le code couleur est le suivant : AT-9010, bâtonnets magenta ; GDP, bâtonnets jaunes ; et Mg2+, vert clair.

Ainsi, dans cet article, nous mettons en évidence le ciblage simultané de l'activité RdRp, ainsi que des activités de NMPylation et de RNAylation du NiRAN, et ce, par le même candidat médicament contre le RTC du SARS-CoV-2 dans une seule étude.

3.1.2. Présentation des mes contributions

Sur ce projet, mon travail a consisté en le développement des protocoles de production et de purification de l'enzyme nsp12, ainsi que la production de divers mutants de l'activité de nsp12, chez SARS-CoV et SARS-CoV-2, permettant ainsi de réaliser différentes expériences d'analyse fonctionnelle, ainsi que la résolution de la structure du RTC avec l'inhibiteur AT-9010.

3.2. Comparaison structurale et fonctionnelle SARS & MERS

Les virus du SARS-CoV et SARS-CoV-2 font partie tous deux du groupe B des β-Coronavirus (Voir Figure 11) et leurs nsp12 partagent 96% d'identité et 98% d'homologie de séquences et ont tout deux pour réservoir naturel principalement la chauve souris et divers hôtes intermédiaire comme le pangolins pour le SARS-CoV. Le virus du MERS lui, fait partie du groupe C des β-coronavirus (Voir Figure 11), et a pour réservoir naturel sont aussi les chauves souris, mais cette fois ci l'hôte intermédiaire est le dromadaire. Il y a ainsi une certaine divergence qui commence à apparaître. Si on compare la séquence de la nsp12 du MERS au deux autres virus, celle ci à respectivement 72% d'identité et 83% d'homologie avec celle de SARS-CoV et 71% d'identité et 82% d'homologie avec celle de SARS-CoV-2.

Lors d'une étude précédente menées sur l'AT-9010 et sa prodrogue l'AT-511, des différences importantes ont également été observées, sur selon que cette NA soit utilisés sur des cellules infecté par le SARS-CoV et le SARS-CoV-2, ou le MERS (Good *et al*. 2021). En effet, des essais sur des cellules Huh-7 infectées par le SARS-CoV ou le MERS, et des cellules HAE infectées par le SARS-CoV-2, puis traitées avec différentes doses d'AT-511, ont été réalisés (Voir Tableau 4).

TABLE 1 *In vitro* activity of AT-511 and other oral antiviral drugs against various human coronaviruses[a]

Virus (genus)	Cell line	Compound	CPE assay		VYR assay	Selectivity index
			EC_{50} (μM [*n*])	CC_{50} (μM)	EC_{90} (μM [*n*])	(CC_{50}/EC_{50})
HCoV-229E (*Alphacoronavirus*)	BHK-21	AT-511	1.8 ± 0.3 [2]	>100	ND	>55
	BHK-21	Sofosbuvir	>100	>100	ND	ND
	Huh-7	AT-511	1.7 ± 0.1 [2]	>86	1.2 ± 0.1 [2]	>50
HCoV-OC43 (*Betacoronavirus*)	Huh-7	AT-511	ND[b]	>86	0.5	>170[c]
	RD	AT-511	2.8	>86	2.2	>30
MERS-CoV (*Betacoronavirus*)	Huh-7	AT-511	26 ± 15 [2]	>86	37 ± 28 [2]	>3.3
SARS-CoV (*Betacoronavirus*)	Huh-7	AT-511	ND[b]	>86	0.34	>250[c]
SARS-CoV-2 (*Betacoronavirus*)	HAE	AT-511	ND[b]	>86[d]	0.47 ± 0.12 [5]	>160[c]
	HAE	Molnupiravir	ND[b]	>19[d]	2.8 ± 1.0 [3]	>4.9[c]

[a]The activity of AT-511 and other antiviral compounds was measured in cells infected with different coronaviruses, using the cytopathic effect (neutral red dye) assay and/or the virus yield reduction (VYR) assay as described in Materials and Methods, to determine the effective concentration required to achieve 50% inhibition (EC_{50}) of the virus-induced cytopathic effect (CPE), the concentration to reduce virus yield by 1 $\log_{10}$ (EC_{90}), and the cytotoxic concentration of the drug to cause death to 50% of viable cells without virus (CC_{50}). Values represent results from single or multiple (mean ± SD [n]) experiments.
[b]Not determined because no cytopathic effect was produced by this virus in this cell line.
[c]CC_{50}/EC_{90}, since EC_{50} could not be determined by measuring CPE in the neutral red assay.
[d]Cytotoxicity assessed by visual inspection of cell monolayers.

Tableau 4 : Activité *in vitro* de l'AT-511 et d'autres médicaments antiviraux oraux contre divers Coronavirus humains. Extrait de (Good *et al*. 2021).

Ces résultats montrent la réduction du rendement viral (virus yield reduction, VYR), qui détermine la concentration efficace requise du composé pour atteindre 90 % d'inhibition (EC90) de l'effet cytopathique induit par le virus (CPE). La concentration efficace d'AT-511 semble être de 0,34 µM contre le SARS-CoV, alors qu'elle est de 37 ± 28 µM contre le MERS-CoV. Concernant le SARS-CoV-2, l'EC90 de l'AT-511 était de 0,47 ± 0,12 µM. Ainsi, d'après cette étude réalisée sur des modèles cellulaires, l'AT-

511 est environ 100 fois plus efficace contre le SARS-CoV-2 que contre le MERS-CoV.

Cette différence d'efficacité des antiviraux, notamment les NAs, entre ces trois Coronavirus a déjà été observée à plusieurs reprises. Ces observations soulignent l'importance d'étudier spécifiquement chaque Coronavirus, même au sein de la même famille, car certaines spécificités peuvent considérablement faire varier leur sensibilité aux antiviraux.

Cela pourrait s'expliquer par des variations dans la structure et le fonctionnement du complexe de réplication-transcription (RTC) du MERS, comparé aux deux autres Coronavirus. Quand on s'attarde sur les représentations ESPript des structures secondaires (voir Figure 30), notamment le domaine RdRp, on peut remarquer quelques différences dans la nature des acides aminés entre les trois Coronavirus.

Il y a peu de différences dans le motif F, contrairement aux motifs A, B, C, D et E, même si les acides aminés directement impliqués dans l'activité catalytique (représentés par les flèches rouges) ne sont pas affectés. Cela pourrait influencer la flexibilité globale de la structure, et potentiellement expliquer le phénotype observé avec l'AT-9010 sur le MERS.

Globalement, les différences structurales observées sont principalement situées dans les régions les moins conservées, comme les boucles périphériques, alors que les régions centrales, notamment autour des motifs catalytiques, restent très conservées entre les trois Coronavirus.

Il est difficile de savoir d'où viennent exactement les différences de phénotype précédemment observées entre les nsp12 des trois virus sans pouvoir étudier séparément chacune des trois enzymes. C'est pour ces raisons que nous avons voulu produire et étudier la nsp12 du MERS et ses partenaires au cours de cette thèse.

Domaine NiRAN



Figure 30 : Comparaison des structures des nsp12 de SARS-CoV, MERS et SARS-CoV-2 sous forme de représentation ESPript

Comparaison des structures des nsp12 de SARS-CoV, MERS et SARS-CoV-2 sous forme de représentation ESPript. Les résidus conservés sont indiqués en rouge, et les éléments secondaires, tels que les hélices α (en forme de spirales) et les brins β (flèches), sont représentés au-dessus des séquences. Le domaine NiRAN se trouve en haut du panel et le domaine RdRp en bas. Sont mis en évidence les motifs catalytiques (A à E) du domaine RdRp et les acides aminés impliqués sont signalés par des flèches rouges. Les flèches orange indiques les acides aminés impliqués dans la réaction catalytique du NiRAN.

3.3. Étude du RTC du virus MERS

3.3.1. Introduction

Après avoir contribué au développement du protocole de purification de la nsp12 du SARS-CoV et à l'adaptation de celui-ci pour celle du SARS-CoV-2, j'ai commencé à travailler sur différents essais de purification de la nsp12 du MERS. En parallèle, j'ai également procédé à la production et à la purification des protéines nsp7 et nsp8 en utilisant les protocoles précédemment établis au sein de l'équipe.

3.3.2. Essais de production et de purification en système bactérien

Lors des premiers essais, la purification de la nsp12 de MERS a été réalisée en utilisant le même protocole optimisé que pour la nsp12 de SARS-CoV. La séquence codante pour la nsp12 de MERS a été clonée dans un plasmide T5 afin d'éviter toute co-contamination possible avec la polymérase T7. Cette séquence a été clonée dans les plasmides pQE30 ou pJ404, en fonction de la position du tag histidine souhaitée (N-terminale ou C-terminale). Lors de la transformation, le plasmide codant pour la protéine ciblée est co-transformé avec le plasmide pGT-f2, qui code pour trois protéines chaperonnes (GroEL, GroES et tig ; Nishihara *et al*. 2000), et 2 % d'éthanol est également ajouté à la culture après induction pour induire cette fois les chaperonnes cellulaires de la bactérie. La culture est ensuite maintenue à 23°C pendant la nuit.

Basé sur l'expérience acquise avec la production de la nsp12 de SARS-CoV en bactérie, plusieurs essais de purification ont été réalisés en variant la densité optique d'induction, les quantités d'IPTG, les températures après induction, les durées avant la récolte des cellules, ou le type de cellules utilisées (C2523 (pRARE/LacI ou pGT-f2) ou C3037 pLacI +/- pGT-f2). Il semble que les conditions utilisant les cellules C3037 (pLacI) co-transformées avec le plasmide pGT-f2 et exprimant la protéine à partir du plasmide pQE30 donnent les meilleurs résultats.

Il avait également été observé pour la purification de la nsp12 de SARS-CoV que la sonication (40 % d'amplitude, 1 s on / 1 s off, pendant 1 minute) pouvait affecter la dégradation de la nsp12 lors de sa purification. Le protocole de base utilise donc uniquement une lyse chimique, sans recourir à une lyse mécanique telle que la sonication. Un essai comparatif avec et sans sonication a été réalisé pour déterminer si le même problème était observé avec cette protéine.

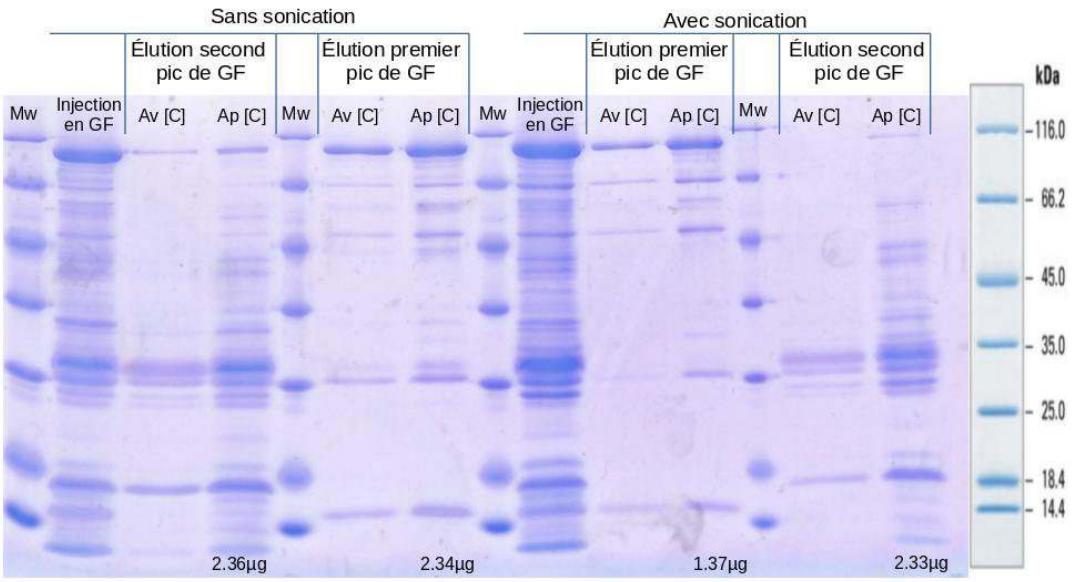


Figure 31 : Gel SDS-Page de purification de nsp12 de MERS avec ou sans étape de sonication.

Gel SDS-PAGE à 12 % d'une purification de la nsp12 de MERS, effectuée avec ou sans étape de sonication (40 % d'amplitude, 1 s on / 1 s off, pendant 1 minute) avant et après injection sur une colonne de gel filtration 16/60 S200. « Mw » désigne le marqueur de poids moléculaire. Le panneau de gauche illustre la purification sans sonication (lyse chimique uniquement), tandis que le panneau de droite montre la purification avec sonication. « Injection en GF » se réfère à l'échantillon concentré après l'étape d'IMAC. « Élution premier pic de GF » désigne le premier pic d'intérêt obtenu après la gel filtration, tandis que « Élution second pic de GF » correspond au second pic, qui représente généralement la forme plus dégradée de la protéine d'intérêt. « Av [C] » indique le pic d'élution avant concentration et « Ap [C] » après concentration. La quantité de protéine déposée sur les pistes « Ap [C] » a été estimée en µg..

Au vu des résultats du gel SDS-PAGE de cet essai (Voir Figure 31), il apparaît que, tout comme la nsp12 de SARS-CoV, celle de MERS est sensible à la lyse mécanique telle que la sonication. Ainsi, à la fin d'une telle purification, la quantité de protéine éluée au premier pic de gel filtration, qui correspond à la forme la moins tronquée de la nsp12, est de 0.194 mg/L de culture sans sonication contre seulement 0.009 mg/L avec sonication. La sonication semble donc avoir un effet extrêmement délétère sur la stabilité de la protéine durant la purification.

Pour tenter de limiter encore la dégradation de la protéine lors de sa purification, l'ajout de divers inhibiteurs de protéase (PMSF, benzamidine) ou d'un cocktail d'anti-protéases (cOmplete™) a également été expérimenté, mais les résultats étaient sensiblement les mêmes.

D'autres types de purification ont alors été entrepris pour essayer de limiter cette autolyse ou clivage par une protéase potentielle d'E. coli. Différents essais de précipitation des contaminants ou de la protéine à l'aide de sulfate d'ammonium ou de kanamycine à différentes concentrations ont été réalisés, mais sans succès ou amélioration notable de la pureté de la protéine cible après purification.

Des essais d'ajout d'étapes de purification sur résine d'échange d'ions (Mono Q et Mono S) ont également été tentés dans différentes conditions de concentration saline. Le but de ces ajouts était d'essayer d'éliminer certains contaminants en les fixant sur la résine (conditions de haute salinité, et protéine d'intérêt non retenue) ou de fixer la nsp12 sur la résine et d'éluer certains contaminants dès le départ. Malheureusement, ces quelques essais n'ont pas permis de limiter la dégradation de la nsp12.

Bien que les conditions générales de production et de purification de la nsp12 de MERS aient été significativement améliorées, la protéine restait trop contaminée ou dégradée pour être utilisée dans des études de caractérisation enzymatique ou structurale.

Pour ces raisons, nous avons choisi de revenir à l'optimisation des conditions de production de la protéine en système bactérien, et pour ce faire, nous avons cette fois-ci utilisé la plateforme de test à haut débit proposée par une autre équipe du laboratoire.

3.3.3. Test à haut débit des contions de production en système bactérien

Pour tenter d'optimiser davantage les conditions de production de la nsp12 de MERS en système bactérien, ainsi que pour limiter sa dégradation lors de la purification, j'ai réalisé des essais de culture dans diverses conditions (types de cellules, plasmides, milieux de culture ; les détails des conditions sont fournis en annexe C1). La purification de la protéine a ensuite été effectuée sur la plateforme de haut débit du laboratoire. Toutes les étapes ont été réalisées dans des plaques de 24 puits (DW24) comme décrit dans le protocole standard du laboratoire (Saez *et al*. 2014), et les modifications apportées sont décrites dans la section « 2.5. Essais d'expression à haut débit ».

Pour cette expérimentation, les deux types de plasmides disponibles (pQE30 et pJ404) ont été transformés chez 12 types de cellules d'expression bactérienne compatibles avec les promoteurs T5. Ensuite, 48 conditions ont été sélectionnées et les cultures ont été maintenues à une température constante de 24°C (avant et après l'induction ; voir Tableau 5).

1	2	3	4	5	6
1 - Tuner - RARElacI	9 - Tuner - RARElacI	17 - Tuner - RARElacI	25 - BL21	33 - C3037	41 - BL21-PGTF2
2 - BL21-PGTF2	10 - BL21-PGTF2	18 - BL21-PGTF2	26 - C2523	34 - ROS2	42 - C2523-RARE
3 - C2523-RARE	11 - C2523-RARE	19 - C2523-RARE	27 - C2530	35 - ROS2-pLysI	43 - C2523-RARElacI
4 - C2523-RARElacI	12 - C2523-RARElacI	20 - C2523-RARElacI	28 - Tuner	36 - ROS2-pLysS	44 - C2523-PGTF2
5 - C2523-PGTF2	13 - C2523-PGTF2	21 - C2523-PGTF2	29 - BL21-PGTF2	37 - BL21	45 - C3037
6 - C3037	14 - C3037	22 - C3037	30 - C2523-RARE	38 - C2523	46 - ROS2
7 - ROS2	15 - ROS2	23 - ROS2	31 - C2523-RARElacI	39 - C2530	47 - ROS2-pLysI
8 - ROS2-pLysI	16 - ROS2-pLysI	24 - ROS2-pLysI	32 - C2523-PGTF2	40 - Tuner	48 - ROS2-pLysS
LB	**TB**	**NZ**	**TB**	**NZ**	
PQE30			**PJ404**		

Tableau 5 : Conditions testées pour la production de la nsp12 de MERS en système bactérien, incluant différentes combinaisons de cellules, plasmides et milieux.

Un système de « score » a été mis en place en fonction du rendement final, de la pureté après la première étape de purification (IMAC-Ni) et de la présence de précipitation visible dans les puits après purification.

Les conditions 3, 5, 6 et 7 (ainsi que la condition 8, qui utilise les mêmes cellules que la condition 7 ; Voir Figure 32 ; les résultats sont visibles en annexe C1 & C2) ont pour point commun d'utiliser le plasmide pQE30 (avec le tag histidine en N-terminale) et d'employer le LB comme milieu de culture. Lorsque le plasmide pJ404, avec le tag de la protéine en C-terminale, est utilisé, il semble y avoir plus de dégradations (Visibles en annexe C1 & C2), notamment avec des bandes d'environ 85 kDa et 25 kDa, comparées aux 109 kDa de la nsp12) et de contaminants. Un milieu riche (TB) ou auto-inducteur (NZ) ne semble pas offrir d'avantage pour cette protéine, avec une augmentation de la quantité de produits de dégradation de la protéine et de contaminants de taille intermédiaire. Fait intéressant, les conditions 3, 5 et 7 utilisant des cellules sans gène répresseur de l’expression lacIq montrent une meilleure expression. À l'inverse, lorsque, comme dans le cas des cellules C2523 de la condition 3, on ajoute un plasmide avec pLacI (condition 4), aucune protéine n'est exprimée (voir Figure 32).

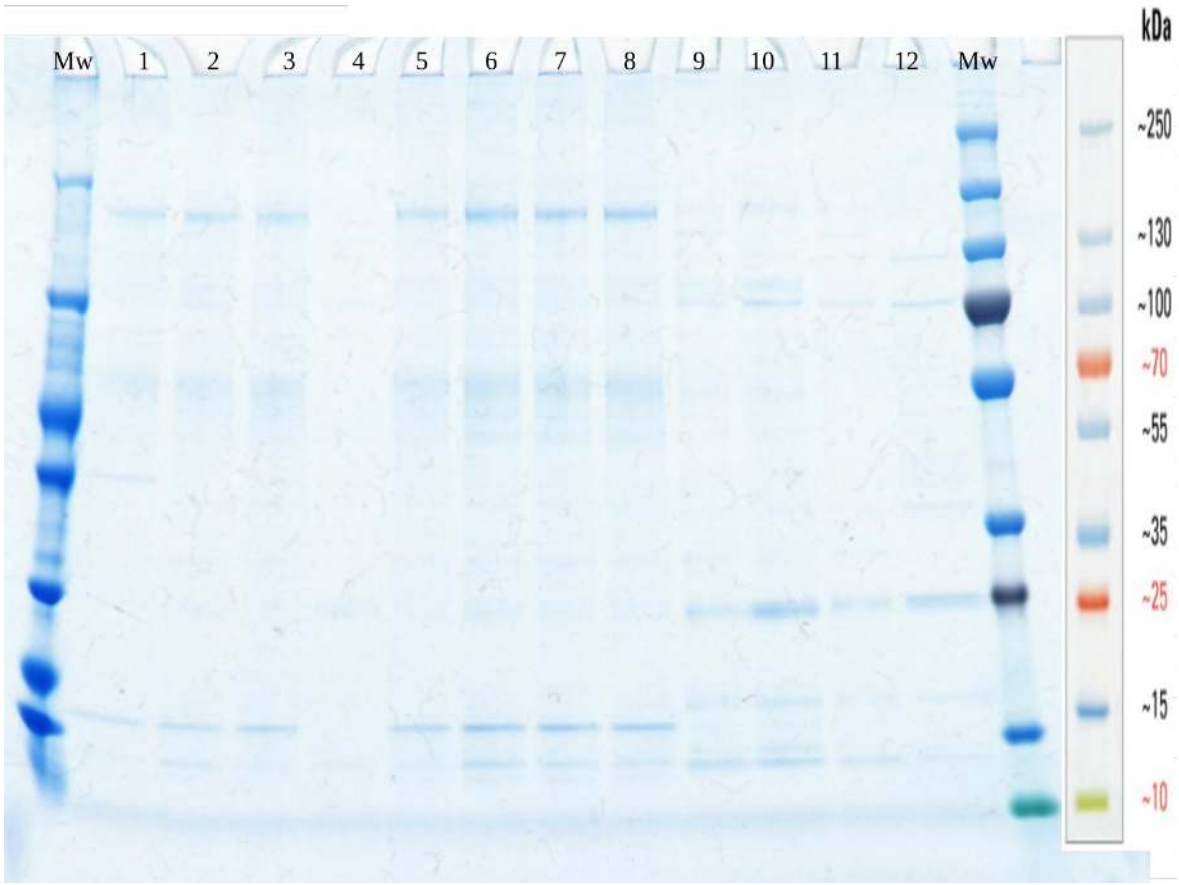


Figure 32 : Gel SDS-PAGE 4-20% des échantillons élution des 12 premières purifications IMAC-Ni effectuées en haut débit.

Se référer aux Tableau 5 pour les correspondances numéro de piste/élution.

Il semble donc qu'une expression lente et continue, sans variation de température importante avant ou après « induction », dans un milieu non riche, favorise la production de nsp12.

La meilleure condition qui avait été précédemment déterminée utilisait également les cellules C3037 avec le plasmide pQE30, comme la condition 6, mais avec une étape de changement de température après induction, passant de 37 °C à 23 °C.

Une autre différence avec les conditions précédemment utilisées est que, dans l'ancien protocole, après avoir récolté les cellules, les culots secs étaient directement congelés à -80 °C. Or, lors de ces essais à haut débit, une fois les cultures centrifugées, les culots sont suspendus dans le tampon de lyse avant d'être congelés (= culots humides). Lors de la mise à l'échelle des conditions de purification de la condition 6, il a été remarqué que l'utilisation de culot humide et l'étape de lyse

chimique de quelques minutes à 37 °C lors de la décongélation et du début de la purification était primordiale pour conserver un profil peu dégradé de la protéine.

À l'inverse, après ces premières étapes de purification différente dans le protocole de lyse chimique, et avec l'ajout d'une étape de sonication à faible puissance, il a été remarqué que la protéine était très sensible à la centrifugation pour clarifier le lysat bactérien (même sans étape de sonication). Il a donc été testé, comme lors de l'essai à haut débit, de faire passer le lysat bactérien dilué trois fois directement sur la colonne IMAC-Ni de 5 ml. Cela nécessite de nettoyer la colonne après chaque utilisation et d'avoir une colonne dédiée, mais cela fonctionne. Il a également été remarqué que la protéine devenait plus sensible à des concentrations d'imidazole au-delà de 250-300 nM au moment de l'élution.

Sur les quatre meilleures conditions trouvées, celle dont la mise à l'échelle a le mieux fonctionné est celle utilisant la condition 3, c'est-à-dire en utilisant les cellules C2523 avec le plasmide pRARE2, qui est pourtant plus proche de la condition 5 qui utilise les mêmes cellules mais avec le plasmide pGT-f2, qui code pour les protéines chaperonnes précédemment citées au lieu des codons rares. La séquence utilisée dans ce plasmide est pourtant une séquence optimisée pour l'usage des codons chez E. coli.

Le gel des différentes fractions d'élution après l'étape d'IMAC-Ni suite à la mise à l'échelle de la condition 3 est à retrouver Figure 33. Les différentes bandes encadrées sur ce gel SDS-PAGE ont été digérées à la trypsine puis analysées par LC-MS/MS sur un spectromètre Orbitrap Q-Exactive. La taille attendue pour cette protéine est de 109 KDa, soit celle représentée en D. Les bandes A sont des contaminants bactériens avec quelques peptides de la nsp12. Les bandes B et C semblent être la nsp12 de MERS accompagnée de quelques contaminants bactériens. Dans les deux cas, il s'agit de la polymérase de E. coli. Pourtant, plusieurs lavages à 1M de NaCl ont été effectués lors de la purification pour éviter les contaminations aux acides nucléiques, mais aussi la co-purification des autres protéines sensibles à ce type de molécules.

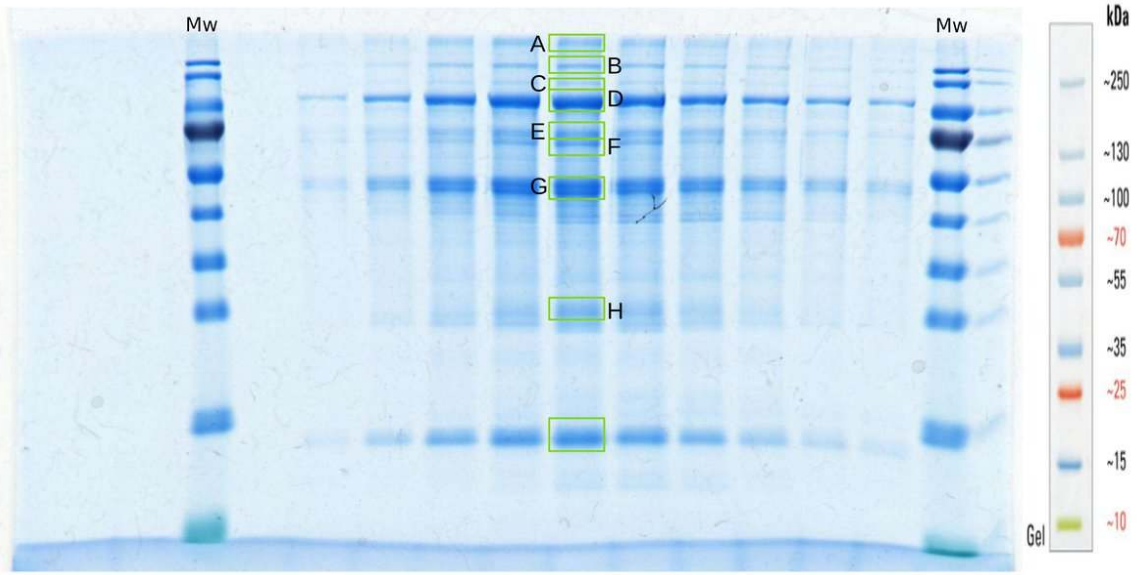


Figure 33 : Gel SDS-Page 12% d'une purification de nsp12 de MERS sur IMAC-Ni de 5ml

Gel SDS-Page 12%, suivant les élutions d'une purification de nsp12 de MERS sur IMAC-Ni de 5ml. Mw correspond aux pistes du marqueur de taille. Les bandes E, F, H, I sont d'autres contaminants. La bande G, quant à elle, est un produit de dégradation de la nsp12 majoritaire dont le poids moléculaire est d'environ 50 kDa. Cela pourrait donc suggérer que la protéine serait auto-lysée ou clivée en deux fragments d'environ 50-55 kDa chacun.

Une analyse de la séquence de nsp12 avec le logiciel PeptideCutter d'Expasy a été effectuée pour trouver d'éventuelles protéases qui cliveraient nsp12 entre une à trois fois. Seule l'hydroxylamine (Bornstein et Balian 1977) ressort de la base de données avec deux sites de coupure potentiels proches du centre de la protéine qui donneraient des fragments d'environ 80 et 30 kDa. Cette protéase ne semble pas être compatible avec notre profil de bande sur gel.

Ainsi, les différentes optimisations de la production ainsi que de la purification de la nsp12 du MERS en système de production bactérien ont été grandement améliorées. Certes, il y a moins de contaminants et de produits de dégradation de la protéine, et cette dernière est moins contaminée par des acides nucléiques endogènes (DO260/280 = 1.06 tout de même), mais la protéine reste instable après l'élution, et a tendance à précipiter rapidement après l'élution. Cela rend délicat l'ajout d'une étape de gel filtration et risque que l'entièreté de l'échantillon précipite sur la colonne.

Cela est peut-être dû à l'augmentation de la pureté de la protéine, qui était potentiellement stabilisée par d'autres protéines bactériennes, comme on a pu le voir avec les bandes de plus haut poids moléculaire où la polymérase de E. coli était co-purifiée (bandes A, B, C visibles sur la Figure 33), ou de ses partenaires naturels que sont nsp7 et nsp8.

Ainsi, peut être contrairement aux nsp12 de SARS-CoV et SARS-CoV-2 que nous arrivons à très bien purifier en système bactérien, et malgré le fait que, au niveau structurale, les trois protéines nsp12 semblent très proches, la nsp12 de MERS pourrait avoir besoin de modifications post-traductionnelles potentielles pour l'aider à maintenir son intégrité.

C'est pour ces raisons que des essais de production de cette protéine dans d'autres systèmes de production, cette fois-ci de type eucaryote (Baculovirus et HEK293), ont commencé à être entrepris.

3.3.4. Conclusions & Perspectives

La recherche des optimisations pour la production et de la purification de la nsp12 du MERS en système bactérien a été un processus complexe, impliquant de nombreuses modifications des conditions de culture, de l'utilisation de différents plasmides et souches, ainsi que l'introduction de différentes techniques de purification.

Bien que des progrès significatifs aient été réalisés, notamment avec une réduction des contaminants bactériens et des produits de dégradation, la protéine reste instable après l'élution et présente une tendance à précipiter rapidement. Cette instabilité rend difficile l'ajout d'une étape de gel filtration a cause du risque de précipitation de l'échantillon sur la colonne.

Plusieurs hypothèses sont possibles pour expliquer cette instabilité : la nsp12 du MERS pourrait être plus sensible à l'absence ses partenaires naturels tels que nsp7 et nsp8, ou moins tolérer la présence d'autres protéines bactériennes co-purifiées comparé aux autres nps12 purifié au laboratoire.

Il est également possible que contrairement aux nsp12 de SARS-CoV et SARS-CoV-2, la nsp12 de MERS ait plus besoin de modifications post-traductionnelles et que ces dernières soient plus importante pour le repliement et la stabilité de la protéine, ce qui est difficilement réalisables dans un système de production bactérien.

Pour ces raisons, des essais de production en systèmes de type eucaryote (Baculovirus et HEK293) ont été initiés. Ces systèmes devraient permettre potentiellement de produire les modifications post-traductionnelles nécessaires à sa bonne production.

En parallèle, cote système de production bactérien, l'utilisation de nouveaux tampons aidant à la stabilisation avec l'ajoute de lysine en plus des détergeants telle que le CHAPS déjà ajouté. Le changement de la base du tampon et essayer un tampon à base de phosphate par exemple, pourrait être explorée pour améliorer la stabilité de la protéine lors des étapes de purification.

Les efforts autour de la recherche de la production de la nsp12 de MERS doivent être poursuivie pour permettre à terme, de mieux études les particularités enzymatiques et structurales de ce RTC. Cela permettra ainsi de mieux comprendre les mécanismes de réplication globaux des Coronavirus, et potentiellement aider à l'identification et le développement de nouvelles molécules et/ou stratégies antivirales.

Chapitre 4 : Étude structurale de méthyltransférases virales

4. Étude structurale de méthyltransférases virales

4.1. Introduction

Les méthyltransférases virales sont des enzymes essentielles dans le cycle de vie des virus à ARN, jouant un rôle clé dans la modification de l'ARN viral pour assurer une réplication efficace et échapper aux défenses immunitaires de l'hôte. L'étude structurale de ces enzymes offre des perspectives prometteuses pour le développement de nouveaux agents antiviraux ciblant spécifiquement ces mécanismes.

Ce chapitre se concentre sur l'analyse structurale de trois méthyltransférases virales distinctes, chacune présentant des caractéristiques a la fois propre et communes, permettant ainsi de faire le parallèle entre différents virus, totalement pour l'étude d'effet antivirale de molécule en cours d'étude.

Dans la première partie, nous examinons une partie des résultats sur la 2'-O-méthyltransférase de la protéine NS5 du virus de la dengue et ces interactions avec l'inhibiteur AT9010 qui on mené à l'article « AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5 », ce qui ouvre la voie à de nouvelles stratégies thérapeutiques contre la dengue.

La deuxième partie est dédiée à l'étude structurale d'une nouvelle N7-méthyltransférase découverte dans l'ORF1a, telle que décrite dans l'article « A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome ». Cette section présentera cette méthyltransférase, l'étude structurale et enzymatique auquel j'ai participé, et une discussion sur son emplacement inhabituel dans le génome viral et les implications fonctionnelles de cette configuration atypique.

Enfin, la troisième partie présentera mes travaux sur la structure de la nsp14 du SARS-CoV-2. Nous détaillerons la structure cristallographique obtenue de nsp14, stabilisée par un nanocorps. Ce résultat sera complété par plusieurs caractérisations structurales et enzymatiques de ce nouveau complexe.

4.2. Étude structurale de la 2'-O-méthyltransférase de NS5 de la Dengue

Le genre des Flavivirus comprend plusieurs pathogènes humains importants, notamment les virus de la fièvre jaune (YFV), Zika (ZIKV), West Nile (WNV), l'encéphalite à tiques (TBEV), l'encéphalite japonaise (JEV) et la dengue (DENV). Ces virus infectent principalement les glandes salivaires de leurs vecteurs arthropodes et se transmettent à l'humain lors des piqûres. Ces virus enveloppés pénètrent dans les cellules hôtes via divers récepteurs cellulaires. Parmi eux, la dengue représente un enjeu majeur avec environ 20 000 décès et plus de 100 millions de cas annuels.

Le génome des Flavivirus pathogènes pour l'humain est constitué d'un ARN simple brin de polarité positive d'environ 11 kb. Cet ARN est coiffé en 5' mais ne possède pas de queue poly-A en 3'. Les protéines non structurales (NS1, NS2a, NS2b, NS3, NS4a, NS4b, et NS5) sont impliquées dans la réplication du génome viral. Parmi celles-ci, la région de la NS5 est la plus conservée. La NS5 possède plusieurs activités enzymatiques, dont l'ARN polymérase ARN-dépendante (RdRp), responsable de la synthèse de l'ARN viral, ainsi que les activités N7-guanine méthyltransférase et 2'-O-méthyltransférase (MTase) nécessaires à la formation de la coiffe de l'ARN viral, essentielle pour l'évasion de la réponse immunitaire de l'hôte.

4.2.1. Présentation de l'article « AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5 »

Titre de l'article : AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5.

Auteurs : Feracci M, Eydoux C, Fattorini V, Lo Bello L, Gauffre P,Selisko B, Sutto-Ortiz P, Shannon A, Xia H, Shi PY, Noel M, Debart F, Vasseur JJ, Good S, Lin K, Moussa A, Sommadossi JP, Chazot A, Alvarez K, Guillemot JC, Decroly E, Ferron F, Canard B.

Référence : Antiviral Research. 2023 Volume 212. doi: 10.1016/j.antiviral.2023.105574

Les vaccins contre YFV, JEV et TBEV existent, mais un vaccin universel contre les quatre sérotypes de la dengue reste un défi. Les vaccins Dengvaxia™ et QDENGA™ sont approuvés, mais leur protection est partielle et diminue avec le temps. La lutte contre les Flavivirus nécessite également des antiviraux efficaces. NS5 constitue une cible privilégiée pour le développement de médicaments antiviraux, tels que les analogues nucléosidiques (NA) comme le sofosbuvir, déjà utilisé avec succès contre le virus de l'hépatite C.

L'analogue nucléosidique (NA) AT-752 est une prodrogue, qui est métabolisée intracellulairement en AT-9010, un inhibiteur de l'analogue de la guanosine 5'. Le nucléotide GTP naturel est un substrat essentiel pour de nombreuses réactions impliquant NS5 au cours du cycle de vie flaviviral.

Dans cet article (Annexe, Article 2), nous avons étudié le mécanisme d'action du composé AT-752 et de son métabolite actif AT-9010 sur l'enzyme NS5 du virus de la dengue (DENV).

Des essais d'activité réalisés sur la MTase de NS5 avec AT-9010 ont révélé une inhibition significative de l'activité 2'-O-MTase, avec une IC50 mesurée à 29,6 µM, contrairement à l'analogue AT-9002, qui possède un groupement hydroxyle à la place du fluor en position 2' et présente une IC50 plus basse à 4,1 µM. Cela suggère que le groupement 2'-OH favorise l'interaction avec la 2'-O-MTase (Figure 34 a).

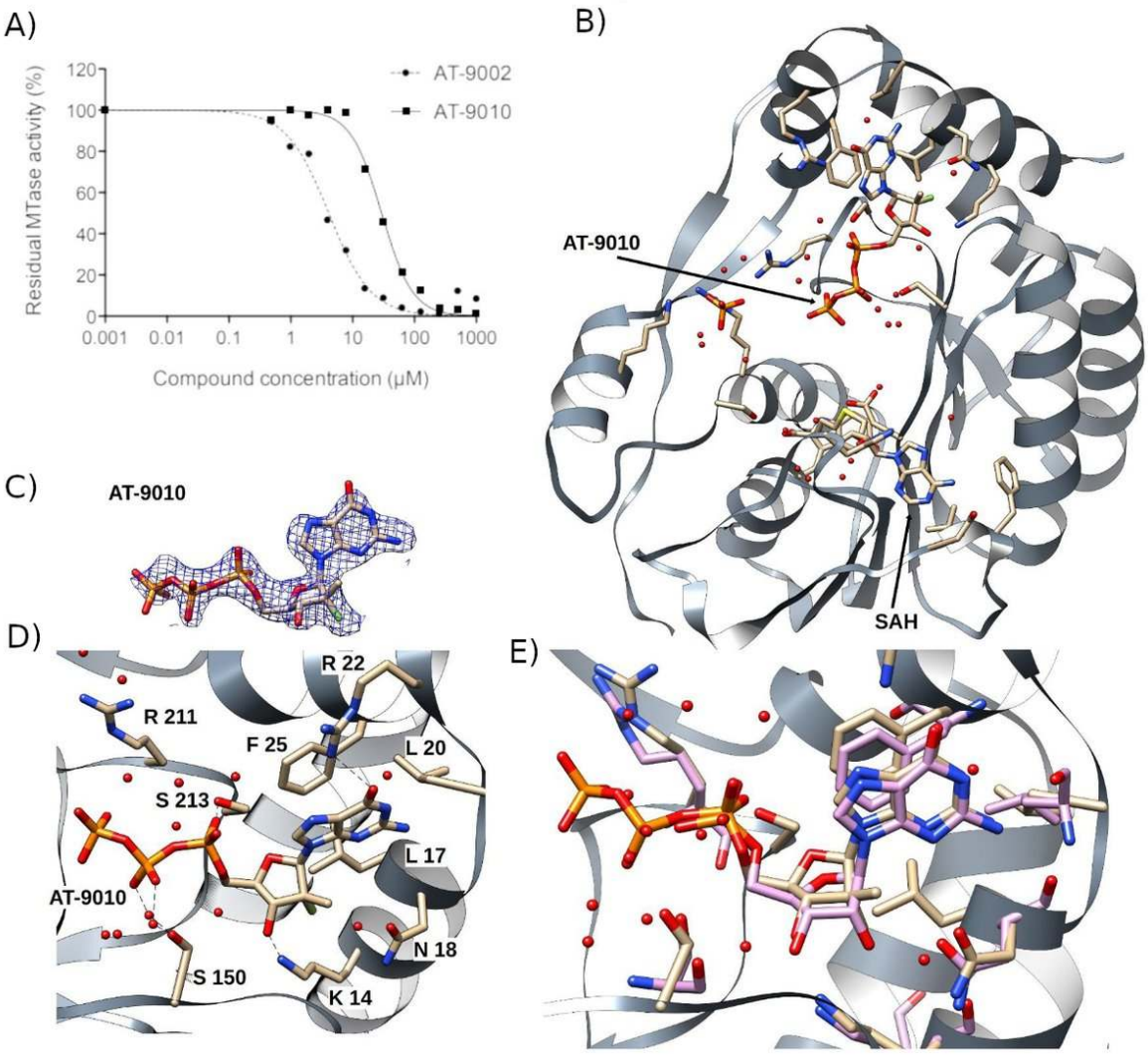


Figure 34 : Base fonctionnelle et structurale de l'inhibition de la MTase du DENV par l'AT-9010.

A) Détermination de la réponse en fonction de la dose en utilisant un test de liaison sur filtre (FBA). Les réactions ont été incubées à 30°C pendant 30 minutes et l'activité enzymatique a été déterminée en

utilisant le FBA comme décrit dans la section Matériel et Méthodes. B) Structure globale de la MTase de DENV3 représentée en rubans bleus complexée avec SAH et AT9010 représentés en bâtonnets. L'AT-9010 est situé au site de liaison du Cap. C) Carte Omit à 1σ de l'AT-9010. D) Vue rapprochée de la liaison de l'AT-9010 dans le site de liaison du GTP. Les résidus en interaction sont étiquetés et montrés en bâtonnets. E) Comparaison structurale de la position d'un GTP (rose) et de l'AT9010 (marron clair) liés dans le site de liaison du Cap. Les structures du site de liaison du Cap appartiennent aux MTases de DENV2 (en rose - PDB 2P1D) et DENV3 (en marron clair) et sont conservées.

La structure cristallographique de la MTase de DENV3 en complexe avec AT-9010, résolue à une résolution de 1,97 Å, montre que l'AT-9010 se lie au site de la coiffe GTP à la place du GTP naturel (Figure 34 e). Les groupes 2'-fluoro et 2'-méthyle ne perturbent pas la position du ribose dans le site actif, et l'interaction avec Lys14 semble jouer un rôle dans l'inhibition de la 2'-O-MTase. Cependant, AT-9010 n'inhibe pas l'activité N7-MTase, ce qui suggère que l'inhibition est spécifique à la méthylation en 2'-O de l'ARN.

En plus de son effet sur l'activité 2'-O-MTase de NS5, l'AT-9010 agit également sur le domaine RdRp de NS5 en tant que terminateur de chaîne non obligatoire. Lors de la synthèse de l'ARN de novo, la polymérase NS5 initie la formation du dinucléotide amorce pppApG, une étape clé de la réplication virale. Les expériences montrent qu'AT-9010 est fortement discriminé par rapport au GTP lors de cette étape d'initiation, avec des IC50 de 315 µM et 107 µM en présence de Mn^{2+} et Mg^{2+}, respectivement (Figure 35 a, b). Bien qu'AT-9010 soit incorporé dans l'ARN sous forme de pppApAT-9010, il ne prolonge pas la chaîne, agissant ainsi comme un terminateur de chaîne. Toutefois, cette inhibition est beaucoup plus faible que lors de l'étape d'élongation de l'ARN, où AT-9010 est discriminé environ 12 fois par rapport au GTP naturel. Cette discrimination diminue lors de l'élongation, ce qui renforce l'idée qu'AT-9010 agit principalement au niveau de cette phase.

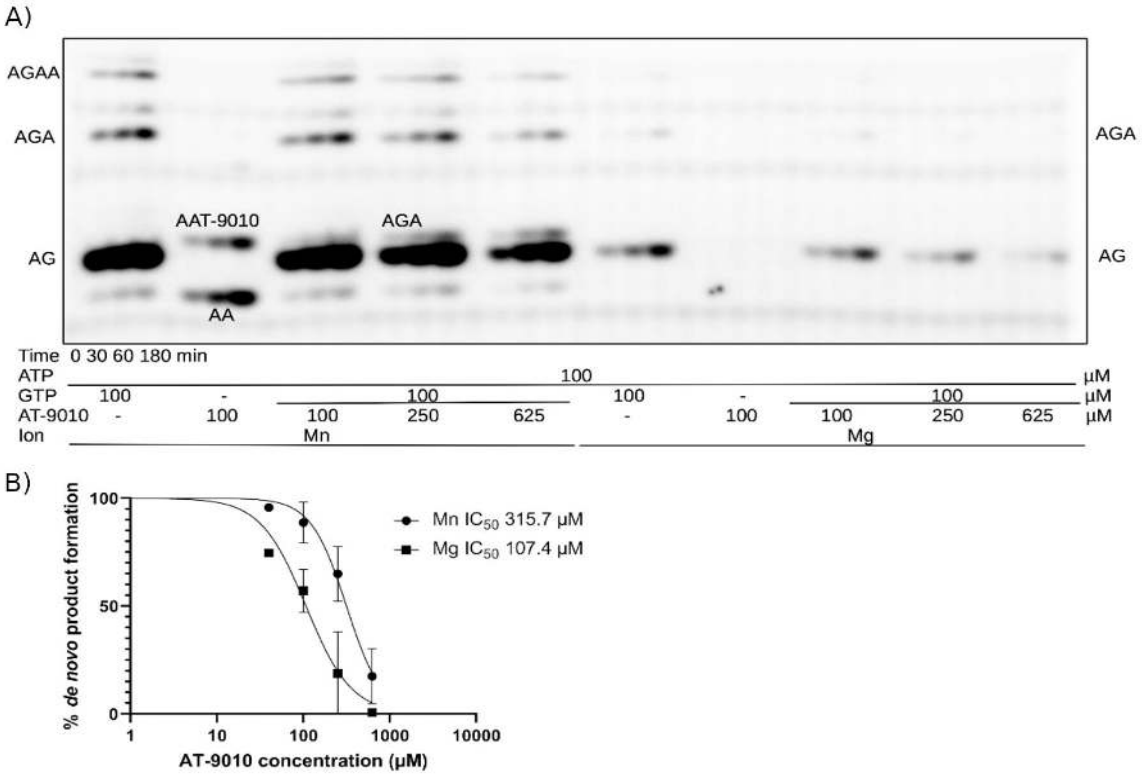


Figure 35 : Incorporation et inhibition de l'AT-9010 lors de la réaction d'initiation de novo par le DENV2 NS5.

A) Analyse par PAGE des réactions d'initiation de novo utilisant un modèle de 10 nucléotides correspondant à l'extrémité 3' du génome (AACAGGUUCU) résultant en la formation de pppApG, pppApGpA et pppApGpApA au fil du temps. Le GTP était soit remplacé par AT9010, soit ajouté en concentrations croissantes aux réactions contenant 100 µM de GTP. B) Inhibition de la formation de produits de novo par AT-9010 basée sur la quantification des bandes de produits de quatre réactions indépendantes (Mn) ou deux (Mg) ; les valeurs correspondantes de IC50 sont données.

En résumé, cet article montre qu'AT-9010 agit sur deux des trois activités de NS5. D'une part, il inhibe efficacement l'activité 2'-O-méthyltransférase (MTase) de NS5, perturbant ainsi la méthylation de l'ARN viral, un processus essentiel pour la maturation de l'ARN et la protection contre la réponse immunitaire de l'hôte. Cela a pu être déterminé grâce à diverses caractérisations enzymatiques et structurales.

D'autre part, AT-9010 agit également comme un inhibiteur compétitif du GTP pendant l'élongation de l'ARN, se comportant comme un terminateur de chaîne non obligatoire lors de la réplication de l'ARN viral. Bien que son effet soit plus limité au niveau de la synthèse de l'amorce pppApG, l'AT-9010 devient beaucoup plus efficace lors de l'élongation de l'ARN. Cette double action sur la méthylation et l'élongation de l'ARN confère à AT-9010 un fort potentiel antiviral. Ces résultats suggèrent qu'AT-9010 pourrait être un candidat prometteur pour le traitement des infections à Flavivirus,

y compris les quatre sérotypes du virus de la dengue. Son mécanisme d'inhibition de NS5 pourrait être exploité pour développer des thérapies antivirales à large spectre et contribuer à mieux contrôler les infections causées par d'autres Flavivirus émergents, tels que ceux responsables de la fièvre jaune ou de l'encéphalite à tiques.

Ce résultat semble à première vu surprenant, mais est mécanistiquement cohérent. En effet la position où se trouve le composé correspond à la position de la coiffe 1 methylé en N7. La N7 méthylation précède la 2'O méthylation qui ne fait pas appel à ce site. En revanche la formation de la coiffe 2 nécessite le correcte positionnement de l'extrémité de la coiffe au site de liaison de la guanosine. Lorsque ce site est occupé par le composé il semble qu'il empêche le positionnement correct de la coiffe pour effectuer la méthylation en 2'O.

4.2.2. Présentation de mes contributions

Pour ce projet, ma contribution a consisté à participer à la collection de données au synchrotron de différents cristaux et à la résolution de la structure cristallographique du domaine méthyltransférase (MTase) de DENV2, en complexe avec l'AT-9010, à une résolution de 1,97 angström. Par l'analyse de la structure, j'ai pu ainsi montré que cette molécule se lie au site de liaison GTP/ARN-cap. La présence de la molécule au site de liaison du GTP suggère un mécanisme d'inhibition pour l'activité 2'-O-MTase, mais pas pour celle de l'activité de N7-méthylation.

4.3. Étude structurale d'une nouvelle N7-méthyltransférase

Parmi les familles de Nidovirus, les Coronavirus sont des génomes de grande taille, supérieurs à 30 kb, et sont les plus étudiés. L'identification et la caractérisation des enzymes impliquées dans le processus de formation de la coiffe ont progressé, mais la compréhension des méthyltransférases dans d'autres familles que les Coronavirus reste limitée. En particulier, l'identification des N7-MTases dans des familles moins étudiées comme les Tobaniviridae a longtemps été mise de côté, faute de séquences suffisantes pour l'étude de leurs génomes.

4.3.1. Présentation de l'article « A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome »

Titre de l'article : A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome.

Auteurs : Shannon A, Sama B, Gauffre P,Guez T, Debart F, Vasseur JJ, Decroly E, Canard B, Ferron F.

Référence : Nucleic Acids Research, Volume 50, Issue 19, 28 October 2022, Pages 11186–11198, doi : 10.1093/nar/gkac876

Les Tobaniviridae sont une famille de virus à ARN positif qui infectent principalement les reptiles et les poissons, comme par exemple le virus White Bream (WBV). Récemment, une nouvelle N7-méthyltransférase a été identifiée chez ce virus. Cette découverte est particulièrement intéressante, car cette enzyme est située dans une région inattendue du génome viral, l'ORF1a, contrairement aux autres MTases habituellement localisées dans l'ORF1b chez les Coronavirus.

Cette nouvelle N7-MTase est non seulement différente par sa position génomique, mais elle présente également une structure unique, avec un repliement de type Rossmann-fold (RF), la première N7-MTase de ce type identifiée dans un virus à ARN positif.

Dans cet article (Annexe, Article 3), nous présentons cette nouvelle N7-MTase, la manière dont nous avons procédé pour la purifier, ainsi que ses caractéristiques biochimiques et structurales.

Tout d'abord, la région putative de cette N7-MTase a été clonée dans un plasmide pQE30, de la méthionine 1333 à l'isoleucine 1668 du génome du virus White Bream (WBV, NCBI Reference YP-803214.1). Par la suite, nous avons procédé à la purification de différentes versions de cette région, avec des régions en 5' et en 3' plus ou moins longues (Figure 36 a).

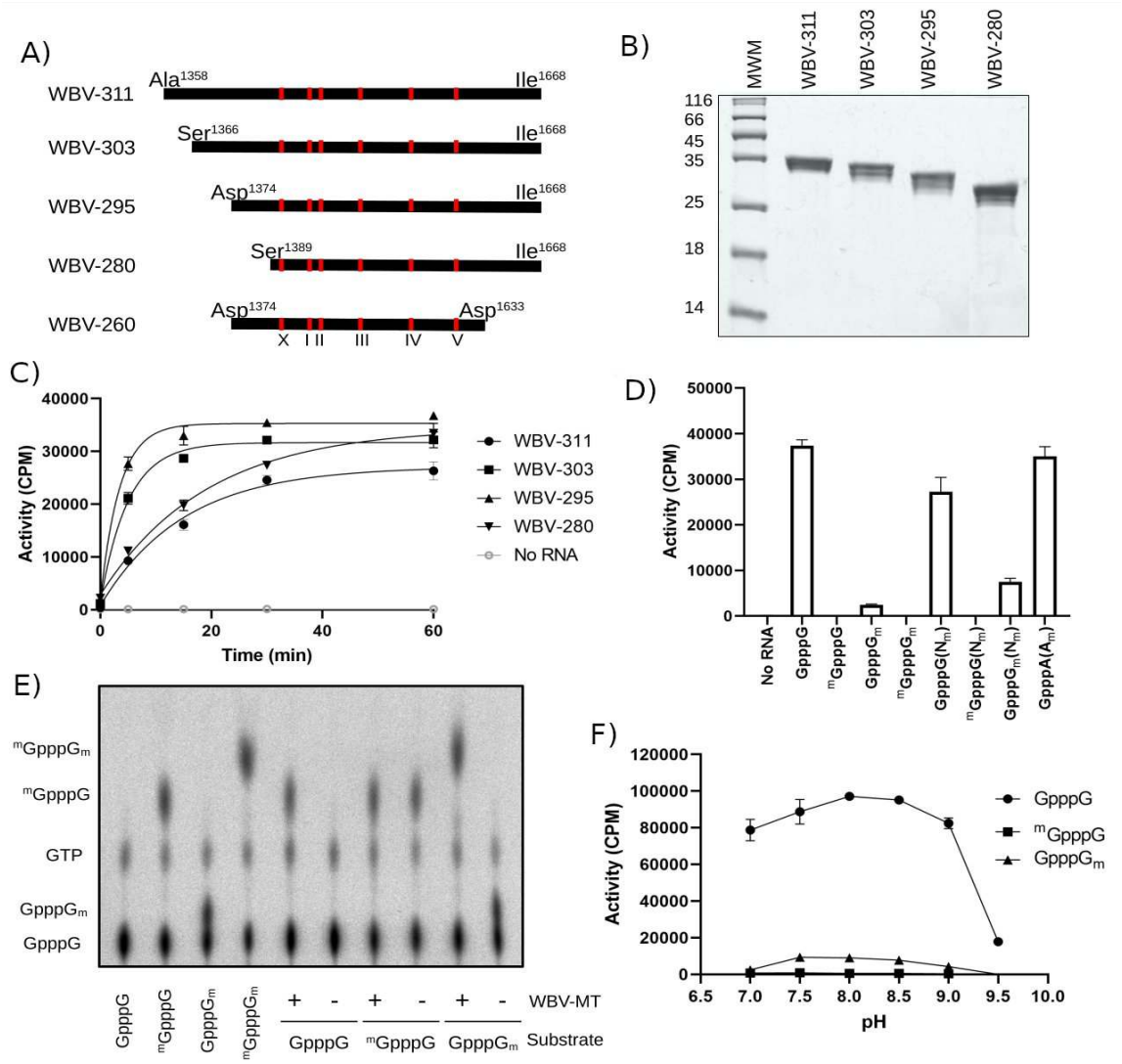


Figure 36 : Panel de figures sur la méthyltransférase cap-dépendante du domaine N7-MT-1a du WBV.

A) Schéma des différents constructs de N7-MT-1a du WBV. Les limites indiquent l'acide aminé de départ et la numérotation par rapport à la polyprotéine 1a, avec la localisation approximative des domaines conservés (I, II, III, IV, V et X) montrée. (B) Analyse SDS-PAGE des différents constructs. (C) Chronologie comparant l'activité de différentes longueurs du domaine présumé N7-MT-1a du WBV. Pour le construct WBV-280, la concentration finale de l'enzyme dans les tests était de 500 nM. Pour

tous les autres constructs, la concentration finale de l'enzyme était de 20 nM. (D) Méthylation de différents substrats ARN13 coiffés préméthylés en position guanine N7 (mGpppN) ou aux groupes 2′O ribose du premier nucléotide (GpppNm), ou les résidus internes de l'ARN (GpppN(Nm)), où N représente tous les résidus. L'activité représente un point temporel unique de 30 minutes mesuré pour le produit WBV-295, et est présentée comme la moyenne ± SD (n = 3). (E) Analyse par chromatographie sur couche mince montrant la migration de différentes structures de coiffe GpppG méthylées. Les quatre premières pistes montrent la migration des contrôles de coiffe, préparés avec la N7-MTase de vaccinia virus commerciale. Le reste du gel montre la modification de différents ARN coiffés (substrat) en présence (+) ou absence (−) de la N7-MT-1a du WBV. (F) Activité de méthylation du produit WBV-295 à différents pH sur différents substrats ARN9 coiffés. Les valeurs tracées représentent un point temporel unique de 30 minutes, présenté comme la moyenne ± SD (n = 3).

Puis, l'activité de la N7-MTase a été testée sur ces différentes versions de la protéine pour déterminer laquelle était fonctionnelle. Parmi ces constructions, la version de l'enzyme Asp1374–Asp1633 s'est avérée la plus active (Figure 36 c). Il a été démontré que cette enzyme méthylait spécifiquement la guanine en position N7 de la coiffe de l'ARN (GpppG), mais ne méthylait pas le ribose en 2'-O, indiquant une spécificité stricte pour la méthylation en N7 (Figure 36 d,e), contrairement à certaines autres MTases virales qui peuvent méthyler plusieurs positions.

Ces essais enzymatiques ont également démontré que l'enzyme ne pouvait pas méthyler les ARNm déjà méthylés en N7, mais fonctionnait sur un ARN déjà méthylé en 2'-O, ce qui pourrait indiquer que la méthylation en N7 précède la méthylation en 2'-O dans la voie de formation de la coiffe des ARNs viraux chez ce virus.

L’influence du pH a également été analysée sur l’activité (Figure 36 f) pour confirmer ce résultat, car cela peut être important pour les différentes méthyltransférases bifonctionnelles trouvées dans d'autres familles virales. Cependant, en ce qui concerne cette MTase de WBV, elle semble incapable de méthyler des résidus internes ou de reconnaître des structures autres que l'extrémité 5' des ARN, ce qui distingue cette N7-MTase des autres chez les Tobaniviridae.

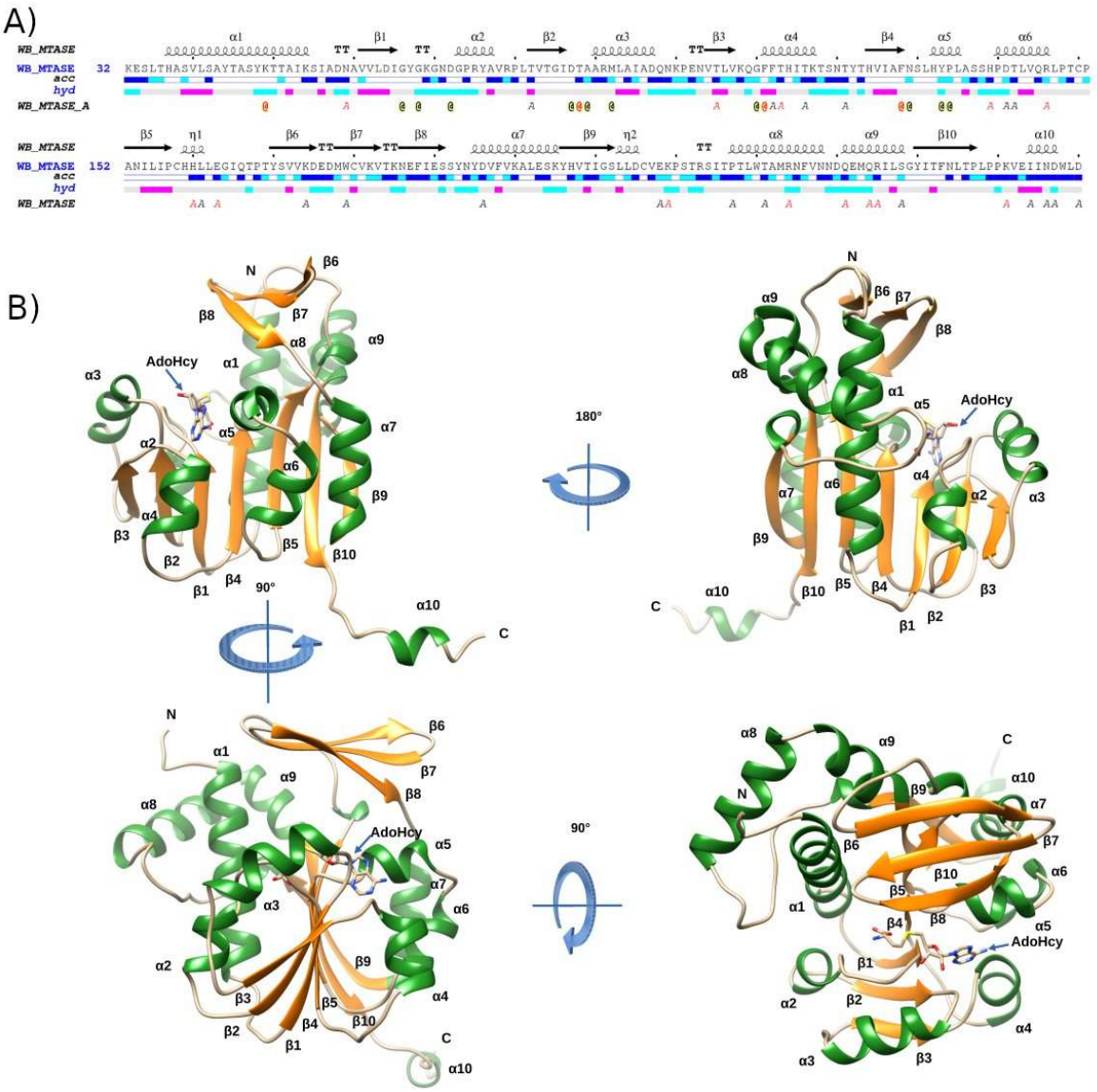


Figure 37 : Séquence et structure de la N7-MT-1a de WBV.

A) Séquence du domaine N7-MT-1a résolu avec les éléments de structure correspondants montrés au-dessus, et accessibilité au solvant (acc ; bleu-exposé, cyan-intermédiaire, blanc-enterré) et hydrophobicité relative des résidus (hyd ; rose-hydrophobe, gris-intermédiaire, cyan-hydrophile). Les résidus interagissant avec le SAH sont marqués d'un @. B) Modèle en ruban du domaine N7-MT-1a coloré par structures secondaires : boucle (tan), hélice α (vert), brin β (or). Les structures secondaires sont étiquetées selon A).

Du côté de la structure de cette nouvelle enzyme, nous avons résolu la structure cristallographique de la N7-MTase du WBV à une résolution de 1,7 Å. Cette structure présente un repliement canonique, avec un motif structural β-α-β alterné, composé d'un feuillet-β fait de 7 brins en parallèle encadré de chaque côté par des hélices-α. Ce motif de type Rossmann-fold (RF), une caractéristique commune des

méthyltransférases dépendantes du S-adénosylméthionine (SAM) (famille de classe I des méthyltransférases). La découverte ne réside donc pas dans son type de repliement, mais dans le fait qu'elle soit spécifique de la N7-méthylation, et qu'il s'agisse de la première à être résolue chez un virus à ARN positif.

Le site actif de l'enzyme présente des caractéristiques conservées parmi les N7-MTases de divers organismes, bien que la N7-MTase du WBV ait peu de similitudes de séquence (<15%) avec celles de virus, parasites ou mammifères, ce qui indiquant une divergence évolutive importante.

La structure présente également une région "3β-flap" (présenté en figure S6 du papier), formée par une courte hélice-α suivie de trois brins β antiparallèles (Figure 37 a,b), un motif retrouvé dans d'autres N7-MTases, comme celles de l'humain ou d'Encephalitozoon cuniculi, un parasite intracellulaire. En revanche, des différences structurales dans la région "α-bundle" ont été observées. Bien que toutes les MTases apparentées possèdent un domaine similaire, la longueur, la position et le nombre d'hélices α varient considérablement entre ces structures.

Pour finir, la variabilité observée dans les régions "3β-flap" et "α-bundle" (présenté en figure S6 du papier) suggère un rôle dynamique dans l'accommodation et la stabilisation de la structure de l'ARN GpppG-cap au cours de la synthèse de la coiffe cap-0, et elles sont des marqueurs structurales des N7-MTases.

Pour conclure, cette étude confirme la découverte d'une N7-méthyltransférase unique et distincte dans l'ORF1a des Tobaniviridae non mammifères, contrairement à celles habituellement retrouvées dans l'ORF1b. Cela permet de mettre en perspective la diversité génétique et évolutive des virus à ARN positif et améliore notre compréhension des stratégies utilisées par ces virus pour s'adapter à différents environnements et hôtes.

4.3.2. Présentation de mes contributions

Pour ce projet, j'ai travaillé sur la création des plasmides codant pour les différentes versions/longueurs de cette N7-MTase. Ensuite, j'ai optimisé et produit ces protéines, d'une part pour réaliser les essais enzymatiques, mais également pour la caractérisation structurale. J'ai également participé à de nombreux essais de cristallogenèse de cette protéine et à l'optimisation des conditions pour améliorer la diffraction de ces cristaux, j'ai aussi participé à l'analyse de la structure.

4.4. Structure de nsp14 du SARS-CoV-2

La nsp14 est une protéine bifonctionnelle qui contient deux domaines enzymatiques distincts. Le premier domaine, situé en N-terminal, assure une activité de relecture par exoribonucléase (ExoN) de type DEDDh, corrigeant les mésappariements introduits par la polymérase virale (RdRp). Cette activité est fortement amplifiée par l'interaction avec le cofacteur nsp10. Le second domaine, en C-terminal, est une N7-méthyltransférase (N7-MTase) dépendante de la S-adénosylméthionine (SAM), jouant un rôle clé dans la formation de la coiffe 5' (cap 0), essentielle pour la stabilité de l'ARN viral et l'évasion de la réponse immunitaire de l'hôte.

Contrairement à la majorité des méthyltransférases qui adoptent un repliement de type Rossmann pour la liaison à la SAM, le domaine N7-MTase de nsp14 présente une architecture unique, distincte de ce repliement canonique. Cette différence structurale offre la possibilité de concevoir des inhibiteurs spécifiquement dirigés contre nsp14, évitant ainsi potentiellement de cibler les méthyltransférases humaines qui adoptent un repliement Rossmann-like. Cette spécificité permet de réduire le risque d'effets secondaires indésirables et de concentrer l'effet antiviral sur la méthyltransférase virale.

Récemment, plusieurs études ont également démontré que certains analogues nucléosidiques (NAs), inhibiteurs potentiels de la RdRp du SARS-CoV-2, étaient partiellement excisés par l'activité ExoN de la nsp14, réduisant ainsi l'effet antiviral de ces composés. Par conséquent, l'activité ExoN, essentielle pour la fidélité de la réplication de ce virus, fait également de la nsp14 une cible de choix pour le développement de nouveaux inhibiteurs. Cela suggère aussi une synergie possible d'un tel inhibiteur avec les NAs déjà à l'étude. Ainsi, la compréhension fine des mécanismes moléculaires sous-jacents aux activités de nsp14 doit être étudiée. Néanmoins, l'instabilité connue de son domaine N-terminal en l'absence de nsp10, ainsi que la présence de la région charnière flexible entre les deux domaines de nsp14, posent de nombreux défis à sa caractérisation structurale par cristallographie. Bien que la structure de nsp14/nsp10 de SARS-CoV-2 a été résolu par cryo-EM, ou domaine par domaine en cristallographie, aucune structure complète de nsp14 n'a été résolu. L'intérêt étant de posséder des conditions expérimentales reproductibles permettant d'étudier les interactions avec des inhibiteurs sur l'ensemble de la proteine.

4.4.1. Historique sur les structure de NSP14 chez SARS-CoV-1 et 2

Une première étude structurale sur le complexe nsp14/nsp10 fut présentée pour le SARS-CoV-1 (PDB : 5C8U, 5NFY ; Ferron *et al*. 2018, Ma *et al*. 2015). Concernant le SARS-CoV-2, les particularités structurales de nsp14 semblent plus complexes à appréhender pour sa cristallisation, malgré une identité de séquence de 95 % avec le SARS-CoV-1. Toutefois, des structures séparées ont été publiées pour le domaine ExoN (PDB : 9FZK) et le domaine N7-MTase (PDB : 8FRJ). Récemment, une structure cristallographique de nsp14 seule, tronquée des 7 premiers acides aminés et sans nsp10, a été obtenue (PDB : 7QGI ; Imprachim, Yosaatmadja, et Newman 2023). Elle révèle une structure plus complète que les structures des deux sous-domaines précédemment citées, bien que certaines boucles flexibles ainsi que la partie dépendante de nsp10 restent non résolues.

4.4.2. Utilisation des nanobodies comme protéine chaperonne de la cristallogenèse

Pour surmonter ces défis de cristallisation de nsp14, nous avons opté pour une approche de stabilisation en utilisant des nanobodies (vhh) de camélidés comme chaperons de cristallisation, capables de reconnaître spécifiquement la nsp14 du SARS-CoV-2. Les nanobodies correspondent au domaine variable (vhh) des anticorps à chaîne lourde (HcAb) des camélidés (Desmyter *et al*., 2015). Les nanobodies sont très stables et peuvent être produits à des niveaux élevés en utilisant des systèmes d'expression bactériens.

4.4.3. Résultats nsp14 SARS-CoV-2

4.4.3.1. Production de nsp14 & nsp10

Avant tout, pour ce projet, j'ai cherché à produire les protéines nsp14 et nsp10 avec un protocole et des conditions de purification optimisés et compatibles avec la pureté de l'échantillon requise pour la cristallographie. Pour cette partie, j'ai donc optimisé mes conditions de purification, notamment avec une étape de lavage à 1M de NaCl pour éliminer les acides nucléiques et contaminants potentiels qui peuvent être co-purifiés avec ces derniers (Détails en M&M « 2.4.4. nsp14 et nsp10 de SARS-CoV-2 »).

4.4.3.2. Production des VHHs

Une fois la production de la nsp14 de SARS-CoV-2 établie, nous avons pu passer à la production des vhh anti-nsp14. Ces vhh ont été produits sur la plateforme de notre laboratoire, et plusieurs familles de nanobodies ont été identifiées (Détails de « 2.7. Isolation ADNc des vhh» à « 2.9. Test Elisa »).

Parmi les trois familles de vhh obtenues, une séquence pour chaque famille a été commandée et clonée dans le plasmide pET24 (promoteur T7) chez Twist Bioscience (USA). Chaque vhh a ensuite été produit et purifié, puis nous avons souhaité caractériser leur affinité en Bio-Layer Interferometry (BLI).

4.4.3.3. Sélection du VHH optimum pour les caractérisations fonctionnelles et structurales

(a) Caractérisation en BLI des vhh contre nsp14

Les premiers essais de caractérisation en BLI ont permis de déterminer que, sur les trois familles de vhh, seules les deux premières, que nous appellerons vhh1 et vhh2, semblent avoir une « bonne affinité » avec nsp14. Le vhh, quant à lui, reste souvent proche du bruit de fond. Il se peut que cette dernière famille de vhh ait été à la limite de sélection lors du test Elisa, réalisé à la fin de la sélection des familles de vhh, et qu'elle ait ainsi été sélectionnée alors qu'elle n'était en réalité pas si affine pour nsp14.

Pour la suite de nos résultats, nous allons ainsi nous concentrer sur le vhh1 et le vhh2.

(i) Kd apparent des vhh1 et vhh2 en double référent

Les Kd apparents des deux vhh ont été déterminés à partir d'une gamme de concentrations en protéines cibles allant de 100 nM à 1,56 nM.

Cette caractérisation a été faite en double référence, comme décrit dans la méthodologie suivante (Orthwein *et al*. 2021). Cela permet de soustraire le bruit potentiel lors des mesures et ainsi d'obtenir une meilleure qualité de nos données.

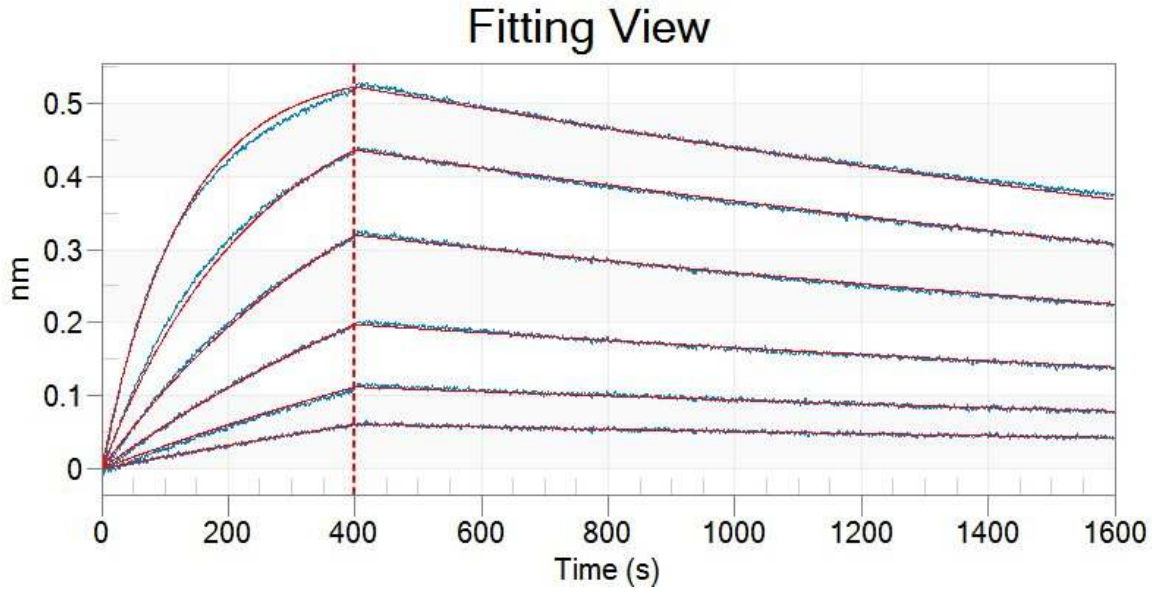


Figure 38 : Graphique de fitting des phases d'associations et de dissociation pour le vhh1

Graphique de fitting des phases d'association (de 0 à 400 s) et de dissociation (de 400 à 1600 s) pour le vhh1 (5 µg/ml) lors des mesures en double référence. Les courbes bleues, de bas en haut, vont de 1,56 nM à 100 nM (par pas de 2 fois la concentration précédente), et en rouge correspondent les courbes de fitting pour le calcul du Kd apparent via le logiciel ForteBio Data Analysis.

Le résultat type que l'on obtient est présenté en Figure 38 ci-dessus pour le vhh1. A partir de ces courbes de fitting et après soustraction des deux références, nous avons obtenu un Kd apparent de 2 nM pour le vhh1 et de 6 nM pour le vhh2. Dans le cas du vhh1, il y a une forte association ($1.51*10^5$ Ms^{-1}) et une phase de dissociation assez faible ($2.92*10^{-4}$ s^{-1}). Pour le vhh2, la phase d'association est également très forte ($2.46*10^5$ Ms^{-1}), mais la dissociation est de l'ordre de 10 fois plus importante ($1.38*10^{-3}$ s^{-1}) comparé au vhh1 (Voir Figure 39).

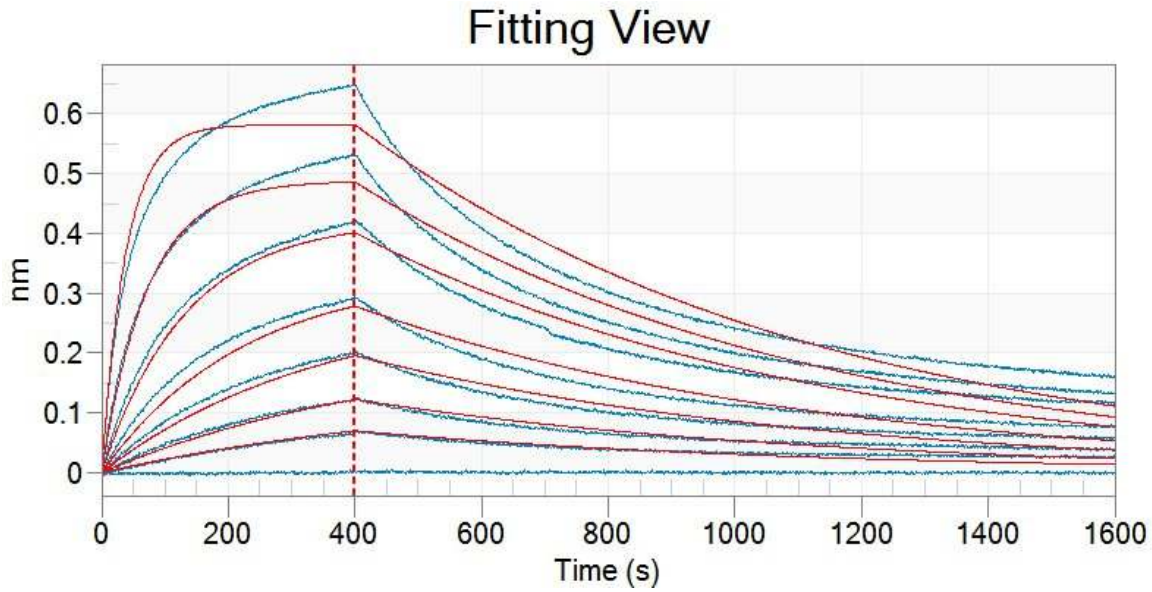


Figure 39 : Graphique de fitting des phases d'associations et de dissociation pour le vhh1.

Graphique de fitting des phases d'associations (de 0 à 400 s) et de dissociation (de 400 à 1600 s) pour le vhh2 (5 µg/ml) lors des mesures en double référent. Les courbes bleu de bas en haut vont de 1.56 nM à 100 nM (par pas de 2 fois la concentration précédente), et en rouge correspond les courbes de fitting pour le calcul du Kd apparent via le logiciel ForteBio Data Analysis.

C'est pour cette raison que nous avons choisi d'utiliser le vhh1 en priorité pour les premiers tests de cristallogenèse. Mais avant de passer à la cristallogenèse, nous souhaitions réaliser d'autres essais de formation de différents complexes avec les protéines nsp14/nsp10/vhhx, et visualiser ces interactions avec l'aide du BLI.

(ii) Essai d'assemblage de complexe nsp14/nsp10/vhhx

D'autres essais avec le BLI ont été réalisés, notamment des expériences où, toujours avec un des vhh fixé sur la sonde, nous avons associé nsp14 avant de suivre une nouvelle phase d'association avec l'ajout de nsp10 (voir Figure 40).

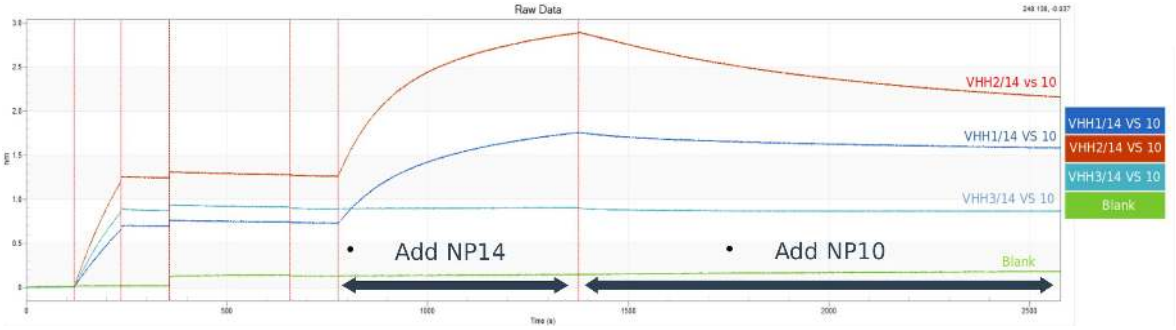


Figure 40 : Graphique de l'essai de l'association de nsp14 sur le vhh, suivi de l'ajout de nsp10

Graphique de suivi des phases d'association et de dissociation obtenues en BLI. Dans l'ordre suivant : tampon, association à un des vhh, lavage, saturation de la sonde pour éviter les liaisons non spécifiques, lavage, ajout de nsp14, ajout de nsp10.

Lors de ces essais, 50 nM de nsp14 ont été utilisés pour la phase d'association et 60 nM de nsp10 pour la phase d'association de cette dernière. On remarque ainsi, que ce soit pour le vhh1 (en bleu) ou le vhh2 (en rouge), que lorsque nsp10 est ajouté, au lieu d'observer une augmentation de la courbe qui signifierait qu'il y a une nouvelle phase d'association, on observe une très légère dissociation pour le vhh1 et une dissociation plus importante pour le vhh2.

Lorsqu'un nouvel essai est réalisé cette fois avec une phase d'association au vhh du complexe nsp14/nsp10 préformé, suivi d'une phase de dissociation (voir Figure 41), on retrouve le profil avec très peu de dissociation après l'association du complexe au vhh1, et une dissociation plus importante pour le vhh2.

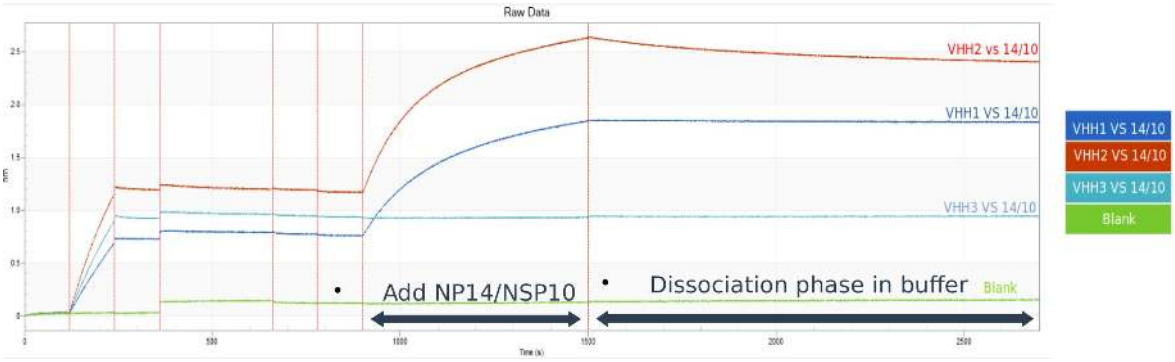


Figure 41 : Graphique de l'essai de l'association du complexe nsp14/nsp10 sur le vhh, suivi d'une phase de dissociation

Graphique de suivi des phases d'association et de dissociation obtenues en BLI. Dans l'ordre suivant : tampon, association à un des vhh, lavage, saturation de la sonde pour éviter les liaisons non spécifiques, lavage, ajout de nsp14/nsp10, phase de dissociation dans du tampon.

Ces résultats (voir Figure 41) pourraient nous laisser suggérer que le complexe tripartite nsp14/nsp10/vhhx est possible, mais que les sites de fixation du vhh1 et du vhh2 sont suffisamment proches de nsp10 pour potentiellement gêner la fixation de cette dernière si elle est ajoutée après la formation du complexe nsp14/vhh.

Lors des assemblages futurs des complexes, le complexe nsp14/nsp10 sera alors formé en premier, suivi d'un des vhh.

Au vu des nouveaux résultats avec le vhh2, nous travaillerons en priorité avec le vhh1 pour la caractérisation de ce complexe et les premiers essais de cristallogenèse.

(b) Caractérisation en SEC-MALS des différents complexes nsp14/nsp10/vhh

La caractérisation du vhh1 a été poursuivie avec l'ajout de ce dernier à la protéine nsp14 ou au complexe nsp14/nsp10. Grâce au SEC-MALS, la partie SEC permet de séparer les différents composants de notre échantillon en fonction de leur taille, tandis que la partie MALS (diffusion de la lumière multi-angle) permet d'estimer l'homogénéité et la masse moléculaire des échantillons purifiés pour les comparer.

Dans un premier temps, les protéines nsp14 (60 kDa) et vhh1 (14 kDa) ont été comparées au complexe nsp14/vhh1 (74 kDa).

Pour nsp14 seule (courbe verte ; voir Figure 42), un léger épaulement au début du pic est observé (estimé entre 103 et 90 kDa). Il s'agit peut-être d'une autre forme (dimérisation) de nsp14 lorsqu'elle est seule, car cela est toujours observé, même après purification par deux gel-filtrations consécutives. Au centre du pic de nsp14

seule, le MALS estime une masse d'environ 67 kDa pour cette partie de l'échantillon, soit une surestimation de 7 kDa au-dessus de la masse moléculaire réelle de la protéine nsp14, probablement liée au léger épaulement observé précédemment.

Pour l'échantillon de vhh (courbe rouge ; voir Figure 42), un pic très homogène et symétrique est observé. L'estimation de sa masse moléculaire est de 13,8 kDa, soit pratiquement identique aux 14 kDa théoriques.

Concernant l'échantillon de nsp14/vhh (courbe bleue ; voir Figure 42), on constate que lorsque le vhh est ajouté à nsp14, le pic devient plus homogène et symétrique que pour nsp14 seule (courbe verte), bien qu'un léger épaulement soit toujours visible. La masse moléculaire estimée par le MALS au centre de ce pic est de 82 kDa, soit une surestimation de 8 kDa par rapport à la masse théorique (74 kDa). En prenant cela en compte, l'augmentation de la masse entre nsp14 et nsp14/vhh, due à l'ajout du vhh, reste cohérente.

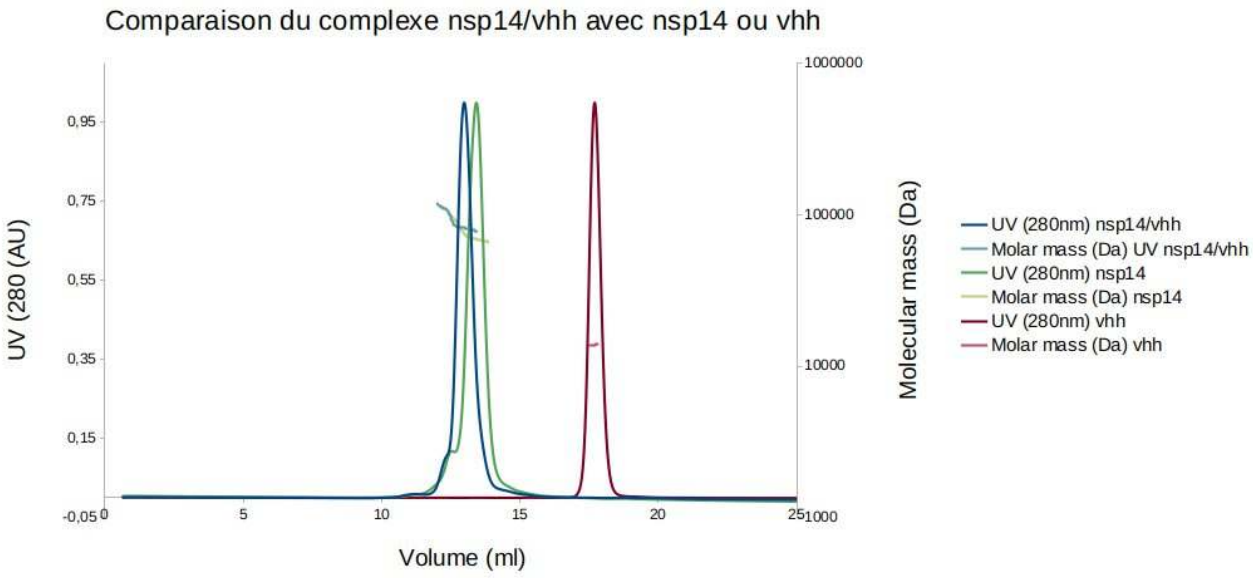


Figure 42 : Superposition des essais de SEC-MALS pour le complexe nsp14/vhh1 et les protéines nsp14 ou vhh1

Séries d'essais de SEC-MALS réalisés sur une colonne 10/300 GL S200. Les résultats ont été superposés. La courbe bleue correspond au complexe nsp14/vhh1, en vert à celle de nsp14, et en rouge à celle du vhh1.

Concernant le complexe nsp14/nsp10/vhh (courbe orange ; voir Figure 43), la partie SEC a dû être réalisée sur une colonne différente des autres échantillons. Pour cet échantillon, le pic est homogène avec un léger post-pic. L'estimation du poids moléculaire pour ce complexe est de 94 kDa, soit très proche des 91 kDa attendus théoriquement.

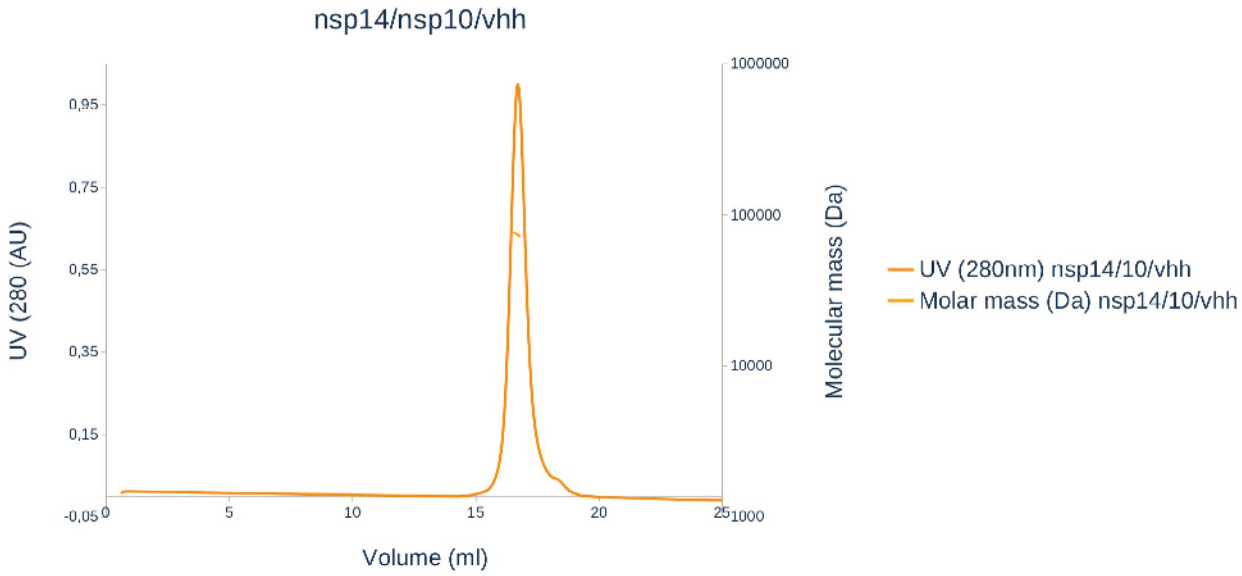


Figure 43 : Essais de SEC-MALS pour le complexe nsp14/nsp10/vhh1

Essais de SEC-MALS réalisés sur une colonne 10/300 GL Superose 6.

Ainsi, ces premiers résultats de SEC-MALS nous permettent de constater que les deux complexes nsp14/vhh et nsp14/nsp10/vhh existent en solution, et que ces échantillons sont homogènes, ce qui va nous permettre de poursuivre leur caractérisation par d'autres méthodes.

4.4.3.4. Caractérisation fonctionnelle de l'influance du vhh sur les activité ExoN et Mtase

Afin d'évaluer l'effet du vhh sur les différentes activités de nsp14, j'ai procédé à l'évaluation fonctionnelle du domaine exonucléase et méthyltransférase en présence du vhh.

Après ces permiers apperçu des intéractions du vhh1 avec avec nsp14 ou le complexe nsp14/nsp10, nous avons voulu nous intérésé aux concéquances que pourrait avoir l'ajout de ce vhh sur les activitées exonucléase et méthyltransférase.

(a) Test enzymatique ExoN

Pour ces essais enzymatiques, le but est d'utiliser un hairpin d'ARN (HP_A, voir Figure 44 C) marqué d'un fluorophore pour visualiser l'activité exonucléase de différents assemblages de protéines (nsp14 seul, nsp14/vhh, nsp14/nsp10/vhh), puis d'observer les produits de dégradation de l'ARN sur gel d'urée 20 %.

Pour ce faire, j'ai d'abord reproduit le ratio nsp14/nsp10 de 1:4, précédemment présenté comme optimal (Bouvet, Imbert, *et al*. 2012) pour visualiser cette activité. Plusieurs autres optimisations ont aussi été réalisées pour trouver la meilleure concentration du complexe nsp14/nsp10 avec mes protéines purifiées, ainsi que la quantité d'ions Mg^{2+}. Des contrôles, positifs (nsp14/nsp10) et négatifs (ajout de 5 mM d'EDTA), ont également été réalisés lors de chaque essai. Des échantillons sont prélevés au cours du temps jusqu'à 1 h, et les résultats de ces gels d'activité sont présentés dans la Figure 44.

Lors des premiers essais (voir Figure 44 A), on observe que l'ajout du vhh seul à nsp14 (ratio 1:4) ne suffit pas à nsp14 pour retrouver son activité ExoN comparée au contrôle positif nsp14/nsp10. L'activité ExoN de nsp14 seul a également été testée dans les mêmes conditions pour être comparée à nsp14/vhh, mais aucun produit de dégradation n'est observé (voir Figure 44 A, colonne 3 comparée à la colonne 4). Il faut multiplier par 35 la concentration de nsp14 pour obtenir un profil de dégradation sur gel similaire à celui observé pour le complexe nsp14/nsp10 (Bouvet, Imbert, *et al*. 2012).

D'autres essais avec le complexe nsp14/vhh, avec des concentrations jusqu'à 40 fois supérieures à nos conditions contrôle positif, ont également été réalisés, mais aucune différence ni produit de dégradation n'a été observé sur gel comparé à l'activité ExoN plus faible de nsp14 seul (gels non présentés).

Ces premiers résultats suggèrent que l'interaction du vhh, à lui seul, ne permet pas au domaine ExoN de nsp14 de retrouver une activité similaire à celle qu'apporte nsp10. Le vhh seul ne semble donc pas permettre d'augmenter l'activité ExoN de nsp14 seule.

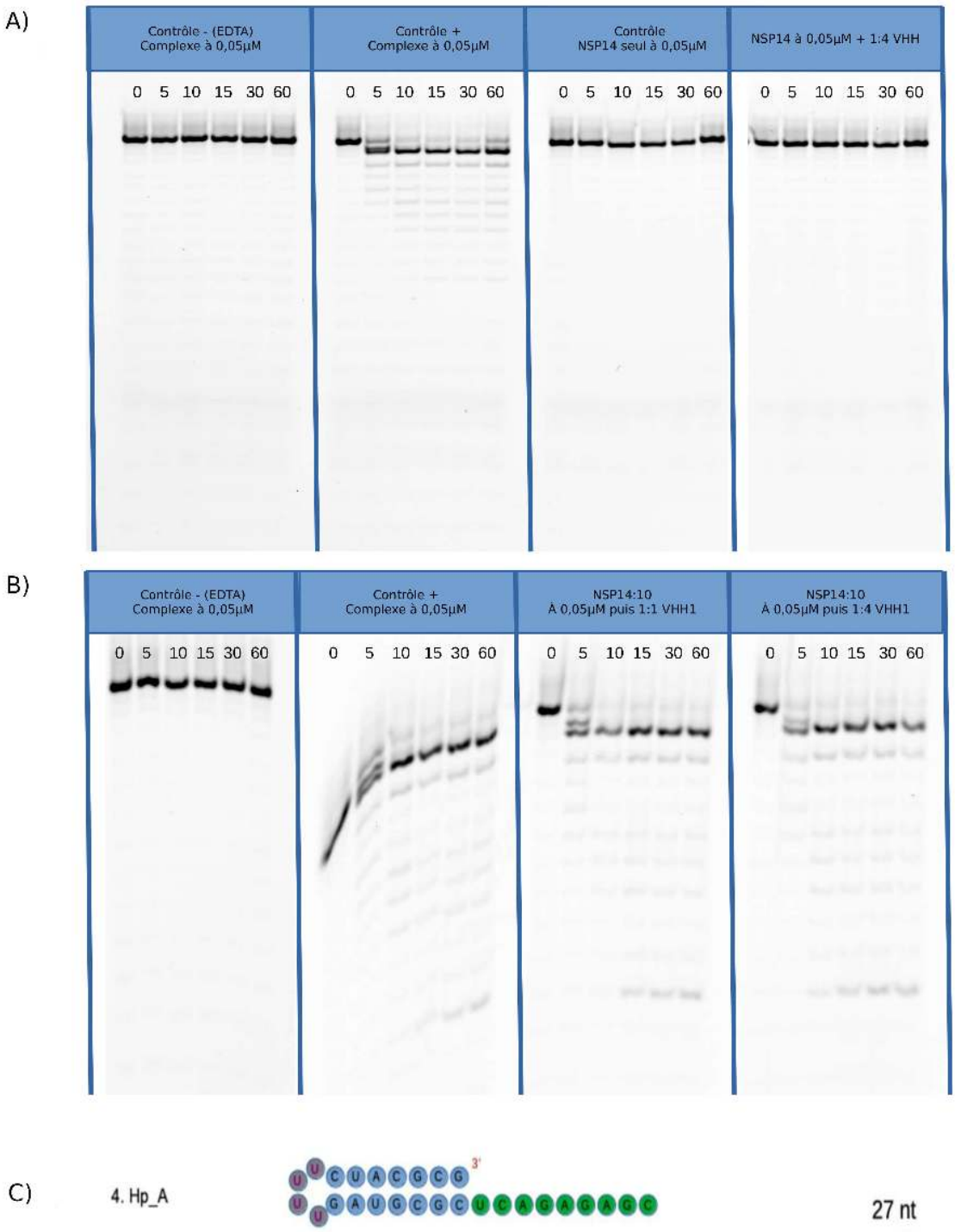


Figure 44 : Panel de gels urée 20% pour la caractérisations de l'activité ExoN

A) Test de l'activité ExoN du complexe nsp14:vhh (1:4). B) Test de l'activité ExoN du complexe nsp14:nsp10 1:4 pré-assemblé puis ajout de vhh (1:1 ou 1:4). Les chiffres au dessus des bandes corresponds au différentes temps d'activité avant le préfévement d'un échantillon, 0, 5, 10, 15, 30, 60 minutes. C) Structure du Hairpin_A (HP_A) de 27 nucléotides.

D’autres essais ont été réalisés (voir Figure 44 B), cette fois sur le complexe nsp14/nsp10 préformé. Le vhh a ensuite été ajouté en ratio 1:1 ou 1:4 (complexe nsp14/nsp10 : vhh). Dans ces conditions, il ne semble pas y avoir de différence dans la quantité de produit de dégradation de l’activité ExoN (voir Figure 44 B) comparée à la condition contrôle positif.

Un essai avec une concentration 20 fois plus élevée du complexe nsp14/nsp10 : vhh a également été réalisé, sans influence visible sur l’activité ExoN comparée à l’activité standard (gels non présentés). Le vhh ne semble donc pas influencer l’activité ExoN du complexe nsp14/nsp10 lorsque ce dernier est préassemblé.

Des essais où nsp10 est ajouté dans un second temps, après la formation du complexe nsp14/vhh, ont également été réalisés pour évaluer l’influence du vhh sur l’assemblage de nsp10 sur le complexe nsp14/vhh, et voir si l’activité ExoN pouvait être restaurée. Lors de ces essais, les contrôles positifs n’étaient pas suffisamment clairs pour pouvoir interpréter correctement les résultats, et devront être renouvelés.

Pour conclure, ces différents essais nous permettent de constater que le vhh ne semble pas permettre à lui seul de restorer une activité ExoN similaire à celle du complexe nsp14/nsp10. Concernant l’influence du vhh sur le complexe nsp14/nsp10, le vhh ne semble ni détériorer ni améliorer l’activité ExoN lorsque le complexe est préassemblé. Cependant, l’étude devra être finalisée avec un essai où nsp10 est ajouté après la formation du complexe nsp14/vhh pour être complète.

(b)Test enzymatique MTase

Pour la caractérisation de l'activité N7-méthyltransférase (N7-MTase) de nsp14 en présence de vhh, un ARN double brin synthétique, servant de coiffe analogue, le GpppAC4, a été utilisé.

La concentration finale de nsp14 est de 10 nM, et de 40 nM pour nsp10 ou vhh lorsqu'ils sont ajoutés, afin de maintenir le ratio 1:4. Lors de ces essais, 0,1 µM de SAM tritié (SAM-3H) est utilisé pour suivre l'activité de méthylation en position N7 sur la coiffe analogue, ainsi que 2 µM de SAM nécessaires pour la réaction. Ces essais ont été réalisés par filter binding assay (FBA), comme décrit précédemment (Ahmed-Belkacem *et al*. 2022). Lors de cette filtration, seules les coiffes sont retenues sur la membrane, tandis que les SAM sont éliminés des échantillons. Un comptage de la radioactivité grâce au compteur à scintillation liquide Wallac MicroBeta TriLux (Perkin Elmer) permet de mesurer la quantité de radioactivité et donc la méthylation réalisée lors de nos différents essais.

Lors de ces essais (voir Figure 45), l'activité N7-MTase de nsp14 après 30 minutes a été utilisée pour comparer et relativiser l'activité des autres assemblages de protéines testés, et calculer ainsi le pourcentage d'activité relatif. Chaque essai a été réalisé en triplicat.

Le standard (voir Figure 45 A, courbe bleue) montre le profil d'activité N7-MTase de 0 à 30 minutes du domaine C-terminal de nsp14, avec un pic d'activité à 100 % observé à 30 minutes.

L'ajout de nsp10 à nsp14 (courbe rouge) ne semble pas modifier l'activité précédemment observée pour nsp14 seule (courbe bleue ; voir Figure 45 A), et un profil similaire est obtenu au cours des différents temps de l'essai.

Dans le cas où le vhh est ajouté après la formation du complexe nsp14/nsp10 (complexe nsp14/nsp10/vhh ; courbe verte ; voir Figure 45 B), le profil d'activité reste très proche de celui de nsp14 seul jusqu'au maximum de l'activité attendu à 30 minutes, où une différence d'environ 10 % au-dessus de la référence est observée.

Pour ces deux premiers essais, l'ajout de nsp10 seul, ou l'ajout de nsp10 suivi du vhh, ne semble pas vraiment modifier l'activité MTase de nsp14.

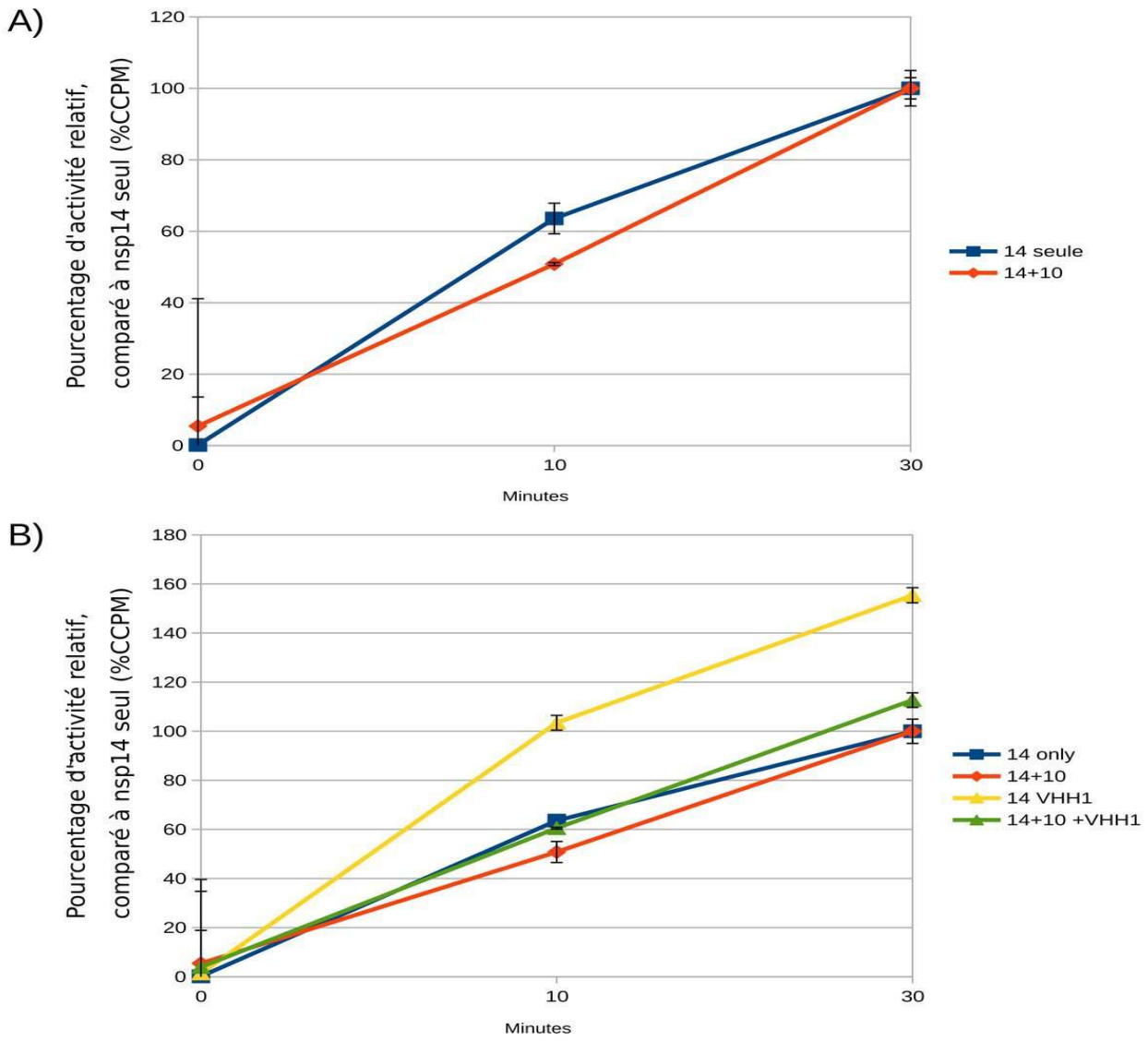


Figure 45 : Courbes normalisées des caractérisations de l'activité MTase

Courbes de l'activité MTase de différents assemblages de protéines, indiquée par le %CCPM (corrected counts per minute), normalisées à 100 % par rapport à l'activité de nsp14 à 30 minutes. Les mesures ont été prises à 0, 10 et 30 minutes.

Pour les essais suivants, nous avons observé l'influence de l'ajout du vhh directement après nsp14. Pour ce troisième essai (courbe jaune), une augmentation d'environ 50 % de l'activité N7-MTase est observée comparée au standard nsp14 seul (courbe bleue ; voir Figure 45 B).

Cela pourrait venir du fait que le vhh, en se fixant sur nsp14, favorise une certaine conformation ou dynamique globale de la protéine (et/ou de la coiffe analogue), permettant *in fine* d'obtenir un meilleur rendement d'activité du domaine Mtase.

Cette différence d'activité entre les complexes nsp14/nsp10/vhh (courbe verte) et nsp14/vhh (courbe jaune) reste surprenante. Cela pourrait nous laisser penser que l'ajout de nsp10 réintroduit du mouvement et crée une autre dynamique au sein du complexe tripartite, qui n'est plus aussi favorable pour l'activité N7-MTase comparé au complexe nsp14/vhh.

Dans la partie suivante sur la caractérisation structurale, nous verrons que le complexe tripartite nsp14/nsp10/vhh existe en solution.

4.4.3.5. Caractérisation structurale des intéractions du vhh avec nsp14 et le complexe nsp14/nsp10

L'étude a été poursuivie en s'intéressant maintenant aux conséquences structurales de l'ajout du vhh sur nsp14 et dans le complexe nsp14/nsp10. J'ai ainsi débuté cette caractérisation d'abord *via* l'utilisation de la diffusion aux petits angles (SAXS) pour obtenir des données structurales des différents complexes en solution, puis poursuivi avec la production de cristaux de ces complexes afin d'obtenir leur structure par diffraction aux rayons X.

(a) Caractérisation en SEC-SAXS

Nous avons ainsi choisi d'étudier la structure adoptée par les quatre protéines/complexes (nsp14, nsp14/vhh, nsp14/nsp10, nsp14/nsp10/vhh) en solution en utilisant la technique de SEC-SAXS.

Pour cela, les différentes protéines ont été purifiées séparément, puis les complexes ont été assemblés directement sur place au Synchrotron (Soleil, ligne de lumière SWING) avant d'être injectés sur la colonne BioSec 3-300 pour la partie SEC, permettant de séparer les différentes formes par exclusion de taille, avant que l'élution ne soit soumise à la diffraction.

Théoriquement, en se basant uniquement sur la masse moléculaire (kDa) des protéines, nous devrions avoir : nsp14 à environ 60 kDa, nsp14/vhh à environ 74 kDa, nsp14/nsp10 à environ 77 kDa, et un complexe tripartite nsp14/nsp10/vhh d'environ 91 kDa.

Expérimentalement, en utilisant l'optimisation bayésienne des masses moléculaires mesurées (Bayes), les résultats sont très proches, avec respectivement 58 kDa pour nsp14 (60 kDa) et 76 kDa pour nsp14/vhh (74 kDa).

Concernant les deux autres complexes, en se basant sur l'estimation de la masse moléculaire de ces deux complexes en fonction de la forme et de la taille des particules (Shape and Size (S&S)), cette méthode permet d'avoir une meilleure estimation pour des complexes.
Pour le complexe nsp14/nsp10, on obtient 66 kDa, soit une différence d'environ 15 % (11 kDa), mais cela reste une augmentation de la masse estimée comparée à nsp14 seul (60 kDa).

Pour le complexe tripartite nsp14/nsp10/vhh, on obtient une estimation de 93 kDa, ce qui correspond à la valeur théorique de masse moléculaire calculée pour les trois protéines (91 kDa).

Cette différence pour le complexe nsp14/nsp10 pourrait s'expliquer par le fait que nsp10 vient s'adsorber "dans la main" de nsp14, créant une zone plus compacte à cette interface et réduisant ainsi la surface globale de la particule, tant en termes de temps de rétention pour la partie SEC que pour la quantité de rayons X diffusée pour la partie SAXS.

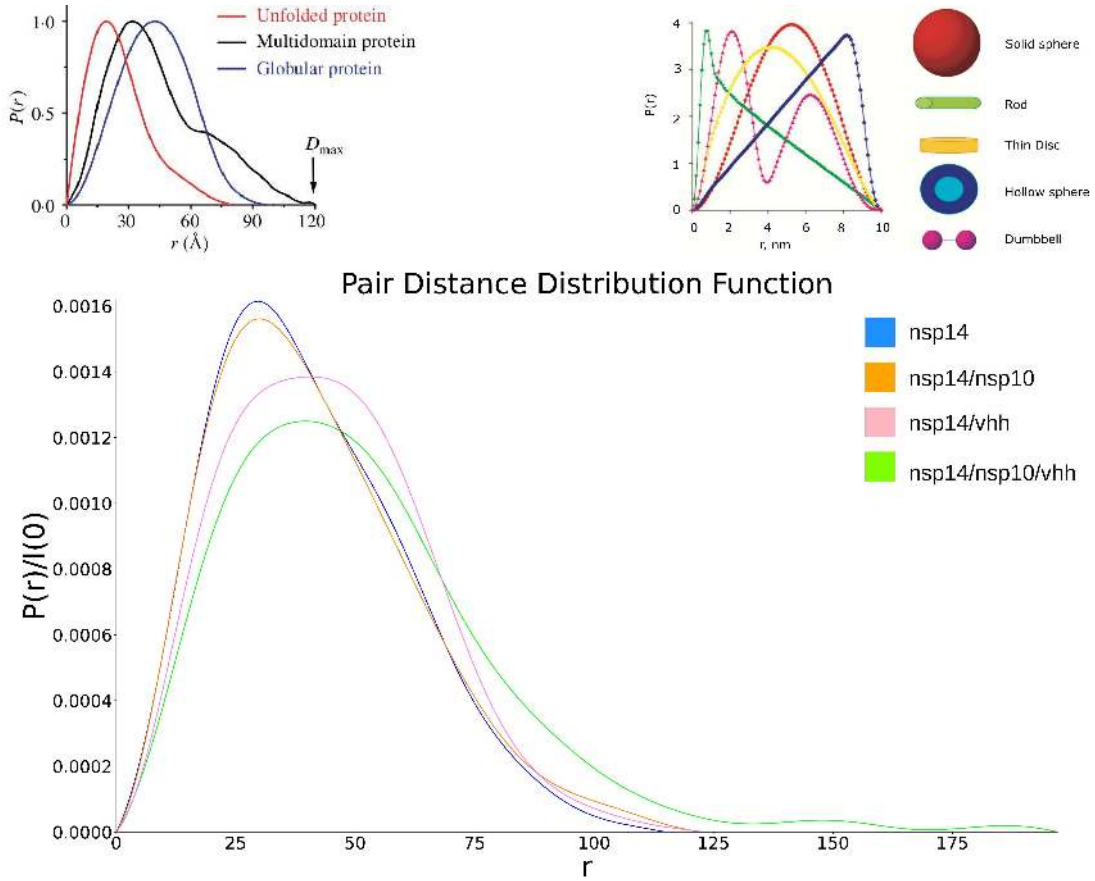


Figure 46 : Superposition des Fonction de distribution des distances de paires (P(r)) de différents complexes avec nsp14

Superposition des Fonction de distribution des distances de paires (P(r)) (Pair Distance Distribution Function) de différents complexes avec nsp14. En abscisse on retrouve la distance entre les pair de points (r) en Å, et en ordonné, (P(r)/I(0) représente la proportion de pair de point (r) normalisée. En haut (à gauche et à droite) ce trouve des exemples type de « Pair Distance Distribution Function » en fonction de l'état de notre protéine ou complexe. Ces distributions des distances de paires (P(r)), fournissent des informations sur la taille et la forme globale des complexes, où un décalage ou une expansion du pic peut indiquer des changements conformationnels ou d'association entre les partenaires protéiques.

En observant les rayons de giration (Rg) calculés à l'aide de l'analyse de Guinier, nsp14 et nsp14/nsp10 ont effectivement des Rg très proches (respectivement de 31 et 32), ce qui est cohérent avec les profils de « Pair Distance Distribution Function » (voir Figure 46), qui montrent une similarité importante entre nsp14 et le complexe

nsp14/nsp10, indiquant que nsp10 modifie finalement peu la forme globale de nsp14 en sa présence.

Le Dmax des deux complexes bipartites (nsp14/nsp10 et nsp14/vhh) est identique, soit 123 nm (contre 115 nm pour nsp14 seule). La différence entre nsp14/nsp10 et nsp14/vhh réside en réalité dans le profil de P(r), qui est plus large pour le complexe nsp14/vhh, se traduisant par une augmentation du Rg à 34, bien que les deux complexes bipartites aient tous deux un Dmax de 123 nm (contre 115 nm pour nsp14 seule).

Concernant le complexe nsp14/nsp10/vhh, il présente un Rg plus élevé de 38, avec un profil P(r) similaire à celui de nsp14/vhh mais plus étalé et prolongé, ce qui est classiquement observé pour les protéines multidomaines ou complexes, atteignant un Dmax de 197 nm.

Le Kratky plot (voir Figure 47) nous informe sur les propriétés structurelles et notamment sur la flexibilité générale des différentes macromolécules étudiées. En prenant le Kratky de nsp14 comme référence (courbe bleue), on observe une remontée en fin de courbe, indiquant une grande flexibilité de la protéine nsp14. Ce résultat était attendu au vu du comportement de nsp14 SARS-COV. L'ajout de nsp10 dans le complexe nsp14/nsp10 (courbe orange) réduit partiellement cette pente observée à la fin de la courbe de Kratky, suggérant une flexibilité moindre par rapport à nsp14 seule, mais toujours présente. Ceci suggere que la région N-terminale se retrouve stabilisée par nsp10.

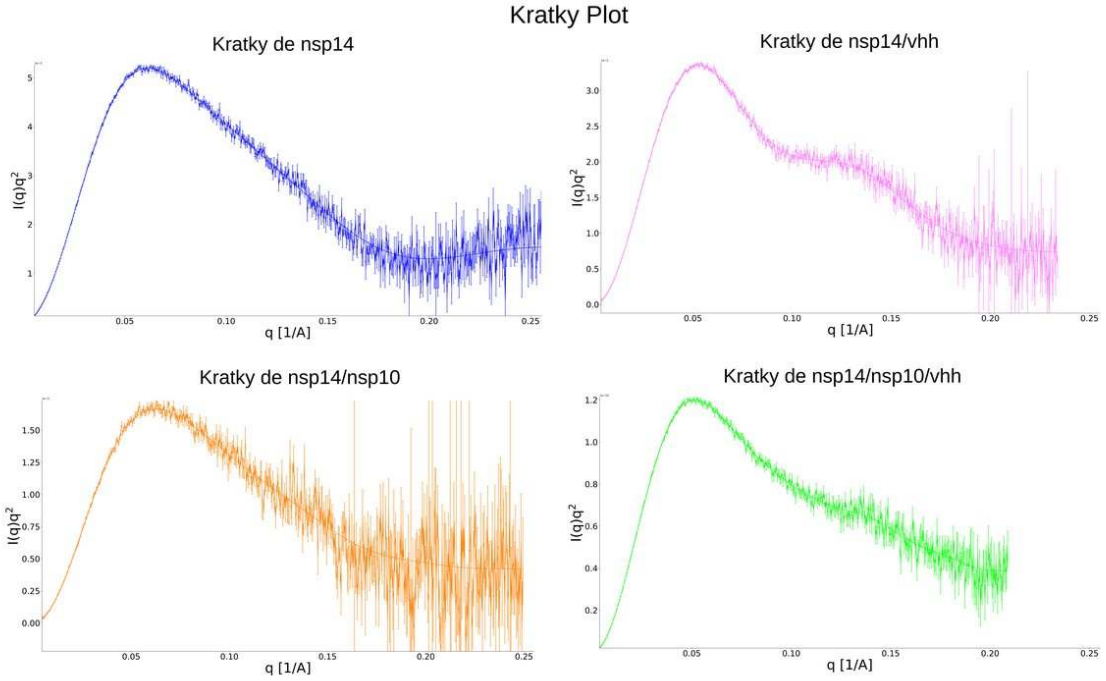


Figure 47 : Kratky Plot

Représentation des courbes de Kratky pour les différentes combinaisons de protéines et complexes étudiées. Ces courbes de Kratky fournissent des informations sur les changements de conformation et l'état de repliement global des protéines et de leurs complexes, essentiels pour comprendre leurs interactions fonctionnelles et structurales.

Si l'on ajoute cette fois le vhh à nsp14 (courbe rose), deux phénomènes sont observés : d'une part, une diminution de la pente en fin de courbe, similaire à celle observée pour nsp14/nsp10, et d'autre part, l'apparition d'un pic (vers 14 q [1/Å]), correspondant probablement à l'ajout du vhh.

Pour le complexe nsp14/nsp10/vhh, ce pic est toujours visible, mais la pente de la courbe de Kratky est encore plus marquée, traduisant une stabilisation accrue de nsp14 lorsque nsp10 et le vhh sont ajoutés ensemble.

Nous avons également utilisé le programme DAMMIF pour calculer différentes enveloppes pour les complexes contenant le vhh. Pour l'échantillon nsp14/vhh, DAMMIF génère un modèle qui correspond aux données expérimentales (voir Figure 48 A). Cette enveloppe correspond à une résolution de 16 Å. L'alignement de l'enveloppe expérimentale obtenue par DAMMIF et de l'enveloppe correspondant à la structure cristallographique de nsp14/vhh (décrite dans le paragraphe suivant) montre une correspondance de 92 %. Cette estimation est confirmée par la simulation de crysol (voir Figure 48 B).

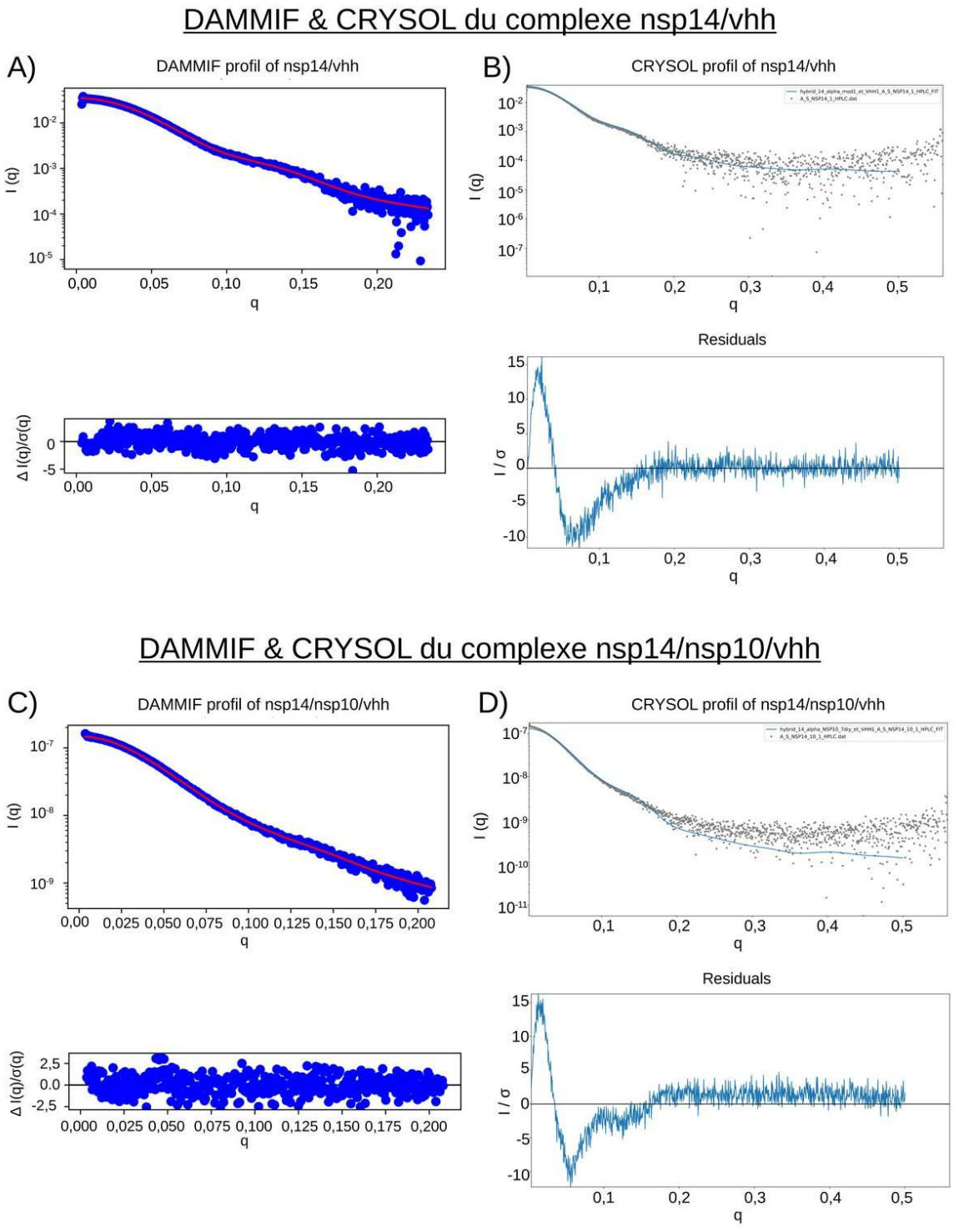


Figure 48 : Analyse par DAMMIF et CRYSOL des profils de diffusion en SAXS des complexes nsp14/vhh et nsp14/nsp10/vhh

Dans le cas du complexe tripartite nsp14/nsp10/vhh, DAMMIF génère un modèle qui correspond aux données expérimentales avec une résolution de 24 Å (voir Figure 48 C). Lorsqu'on aligne cette enveloppe avec l'enveloppe correspondant à un modèle tenant compte de la position du vhh sur nsp14 et enrichi de nsp10, on obtient une convergence des enveloppes de 90 %. La simulation de crysol confirme encore cette estimation (voir Figure 48 D).

Cette analyse en SEC-SAXS nous apporte beaucoup d'informations sur l'organisation des complexes étudiés en solution. Elle nous a notamment permis de confirmer l'existence d'un complexe tripartite nsp14/nsp10/vhh, et que nsp14/vhh est possible en solution, montrant que la liaison du vhh seul sur nsp14 contribue à rigidifier les complexes.

Grâce aux informations obtenues par les différentes approches de MALS et de SAXS, nous avons mis en œuvre la cristallisation de ces complexes, dont les résultats sont présentés dans la partie suivante.

(b)Essais cristallographique

(i) Criblage commerciaux sur tous les complexes

Après avoir identifié le vhh1 comme le meilleur candidat potentiel pour stabiliser nsp14 et potentiellement le complexe nsp14/nsp10, des criblages de conditions de cristallogenèse ont été effectués. Au total, 5 criblages de tampons ont été utilisés sur les complexes de protéines nsp14/nsp10, nsp14/vhh1, et nsp14/nsp10/vhh1.

Quelques conditions (notamment : Proplex E11, D7, C7 ; JCSG Plus C9, C3, H8, H11 ; PACT Premier B3, B5, B11, E2, F1) ont permis d'obtenir des cristaux pour les boîtes contenant nsp14/vhh et nsp14/nsp10/vhh, mais aucune condition n'a permis d'obtenir des cristaux pour nsp14/nsp10 seul.

Les premiers cristaux obtenus proviennent des plaques où de criblages du complexe tripartite (nsp14/nsp10/vhh). Les premiers premiers cristaux ont été testé au synchrotron Soleil et les essais de diffraction de ces cristaux ont permis d'obtenir une structure à 3,6 Å de résolution. Cette dernière nous a permis de vérifier comment le complexe était formé. Il a ainsi été observé que seules nsp14 et une densité électronique correspondant au vhh étaient visibles, mais pas celle de nsp10, ce qui indique que nsp10 a été chassée du complexe lors de la cristallogenèse.

Je me suis donc concentré sur l'optimisation de ces conditions de cristallogenèse en utilisant juste nsp14/vhh ce qui m'a permis d'obtenir les cristaux du complexe nsp14/vhh. Un effort particulier a été porté sur l'optimisation des différents tampons, à la fois des tampons de cristallisation pour augmenter la taille des cristaux et améliorer leur forme, mais aussi des tampons de purification de la protéine. En effet, il a été observé que le tampon de purification de nsp14 avait une grande incidence sur la diffraction finale des cristaux, notamment en raison de son pH (7) et de sa concentration en ions Ca^{2+} (5 mM). Une condition de cristallogenèse dérivée de PACT Premier F1 a ainsi été obtenue.

Cette dernière condition nous a permis d'obtenir la structure finale du complexe nsp14/vhh à une résolution de 2,53 Å, qui sera présentée dans la partie suivante.

En parallèle, d'autres criblages de conditions de cristallogenèse ont été réalisés, mais cette fois-ci avec le vhh2 (nsp14/vhh2 ou nsp14/nsp10/vhh2). Cependant, aucune condition permettant de cristalliser ces complexes avec ce vhh n'a été obtenue pour le moment.

(ii) Structure nsp14/vhh

La structure cristallographique du complexe nsp14/vhh a ainsi été résolue à 2,53 Å (PDB : 9H01 ; caractéristiques résumées dans le Tableau 6 et présentées en Figure 49).

Data processing	
	nsp14 of SARS-CoV-2 in complex with a camelid nanobody
Wavelength (Å)	0.97856
Space group	P212121
– a, b, c (Å)	55.526, 80.832, 138.160
– α, β, γ (°)	90.00, 90.00, 90.00
Resolution range (Å)	51.52 - 2.53 (2.58 - 2.53)
Total No. of reflections	255394 (11704)
No. of unique reflections	21341 (1021)
Completeness (%)	99.9 (98.5)
Multiplicity	12.0 (11.5)
I/σ(I)	10.2 (1.3)
R meas	0.148 (2.1)
R merge	0.142 (2.0)
R-pim	0.43 (0.6)
CC1/2	0.999 (0.8)
Wilson B-factor (Å)	60.4
Structure solution & refinement	
No. of reflections, working set	21338 (406)
No. of reflections, test set (%)	1057 (21), 5.0
R_{work}	0.24
R_{free}	0.29
No. of non-H atoms	
-Protein	4614
-Ligand (Metal-ions)	3
-Water	54
Protein residues	Total 592 : Nsp14 (A) 467 / VHH (B) 125
R.M.S. deviations	
-Bonds (Å)	0.01
-Angles (°)	1.52
Average B-factors (Å²)	
-Protein	70.78
-Ligand	64.10
-Water	62.24
Ramachandran Plot	
-Favoured (%)	93.46
-Allowed (%)	5.51
-Outliers (%)	1.03
Clashcore	6.18
PDB code	9H01

Tableau 6 : Données cristallographiques et statistiques pour la nsp14 du SARS-CoV-2 en complexe avec un nanocorps de camélidé

Tableau récapitulatif des statistiques des données de la structure. Les statistiques générales sont présentées en premier et les statistiques pour la haute résolution sont présentées entre parenthèses.

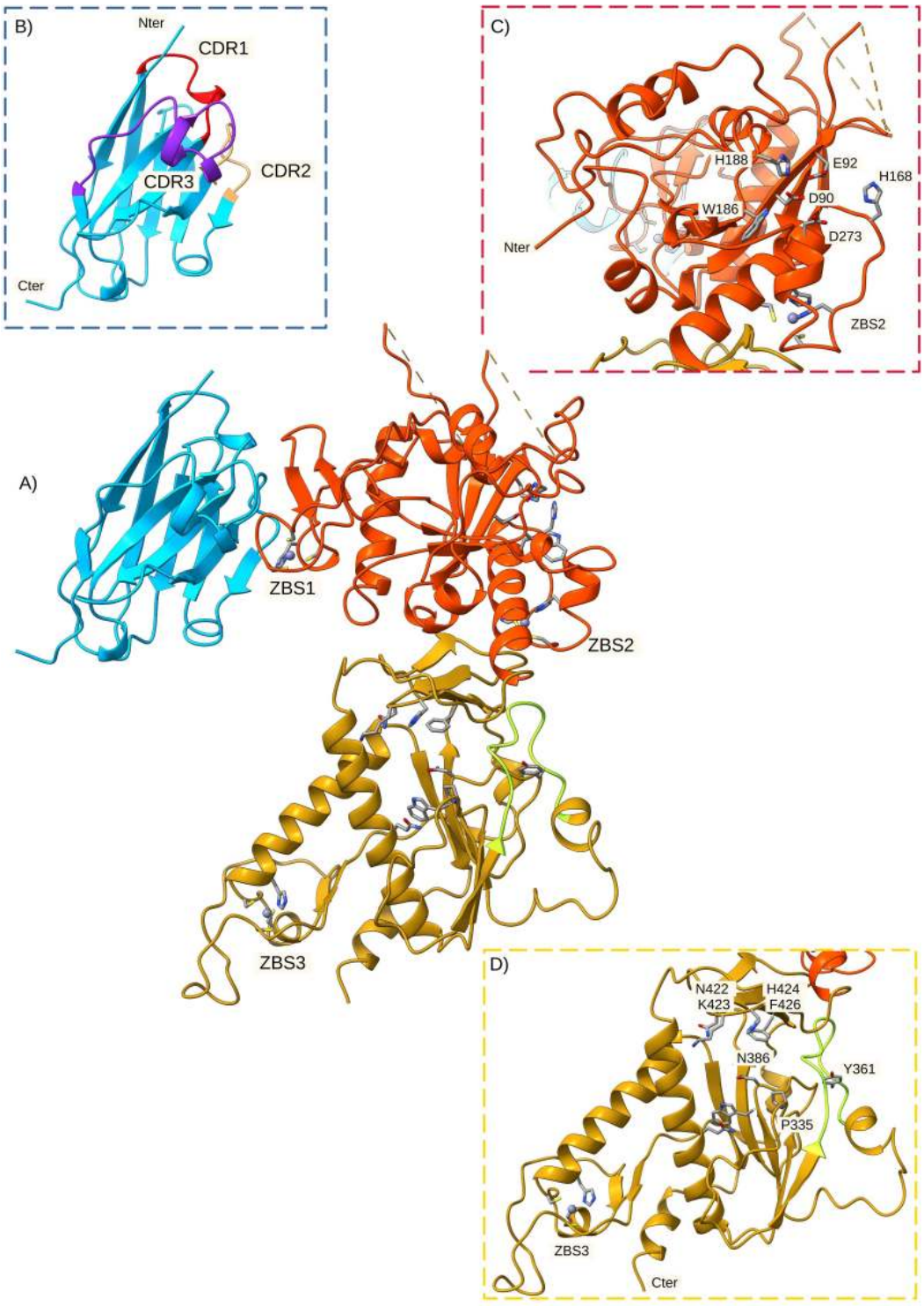


Figure 49 : Structure de nsp14 stabilisée par un vhh (PDB : 9H01)

A) Structure cristallographique du complexe nsp14/vhh (PDB : 9h01) avec les acides aminés des deux sites catalytiques ExoN (orange) et MTase (jaune), ainsi que les 3 doigts de zinc apparents. Le vhh stabilisant nsp14 est visible en bleu. B) Annotation des différents CDR (complementarity-determining regions) du vhh. C) Zoom sur le domaine ExoN de nsp14 avec les acides aminés catalytiques et ceux qui stabilisent l'ARN. D) Zoom sur le domaine MTase de nsp14 avec les acides aminés catalytiques

apparents, ainsi que ceux qui stabilisent le SAM/SAH et l'ARN, de même que la boucle (en vert) qui permet l'entrée et la sortie du SAM/SAH.

Cette structure est constituée d'un hétérodimère composé de nsp14 de SARS-CoV-2 (voir Figure 49 A ; en orange le domaine ExoN et en jaune le domaine N7-MTase) et d'un vhh (en bleu). La protéine nsp14 se présente sous une forme allongée où les domaines ExoN et MTase sont alignés. Aucune déformation au niveau de la région charnière n'est observée. Le vhh interagit par ses régions déterminant la complémentarité (CDR) au niveau du premier doigt de zinc (ZBS1) du domaine ExoN de nsp14, opposé au site catalytique, et ne semble pas interagir directement avec le domaine MTase. L'ion zinc du ZBS1 est présent, et on peut observer les deux autres doigts de zinc de nsp14 (ZBS2 et ZBS3) sur la structure, présents avec leurs ions également, participant à la bonne conformation structurale de l'ensemble de nsp14.

Concernant le vhh lui-même, il adopte une structure compacte en feuillets β (β-sandwich) composée de 9 brins connectés par des boucles flexibles, et est similaire au domaine VH des anticorps conventionnels. Le vhh se lie via trois régions hydrophobes situées dans des boucles de surface correspondant aux régions déterminant la complémentarité (CDR). Dans le cas du complexe présenté ici, il semble que la reconnaissance de nsp14 par le vhh ait lieu principalement via les CDR2 et CDR3, qui sont très proches de nsp14, tandis que le CDR1 semble plus éloigné (voir Figure 49 et Figure 50).

Au niveau de la partie N-terminale de nsp14, nsp14 est constitué d'un sous-domaine correspond à la partie flexible de nsp14 qui se déplie pour enserrer nsp10.

Ici, en l'absence de nsp10, une grande partie de ce sous-domaine flexible n'est pas résolue (de A1 à P24, de A95 à A103 et de V125 à K155),(voir Figure 49). En revanche le cœur catalytique de l'ExoN est parfaitement résolu. La triade catalytique E92, D90 et D273, encadrée par les deux histidines H188 et H268, adopte la conformation connue (voir Figure 49 C) comme dans la structure du domaine ExoN/nsp10 précédemment décrite (PDB : 7DIY ; Lin *et al*. 2021). Cependant, on s'attendait à observer au site catalytique DVE, la présence de deux ions Mg^{2+}. Leur absence indique que le vhh ne compense pas la présence de nsp10 pour stabiliser ces ions.

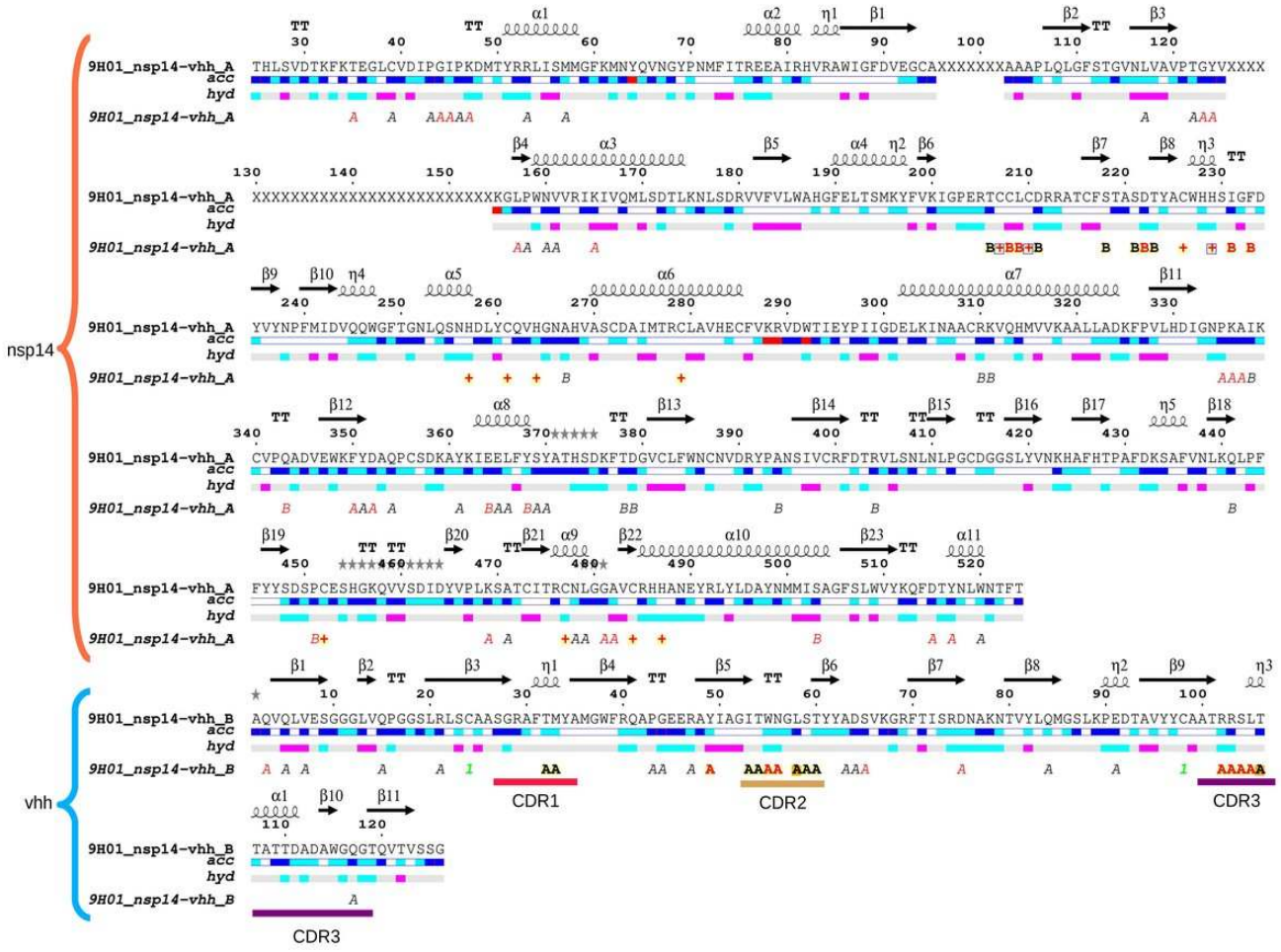


Figure 50 : Représentation ENDscript de la stucture nsp14/vhh

Représentation de la structure du complexe nsp14/vhh sous forme ENDScript. La chaîne A correspond à nsp14 et la chaîne B au vhh, suivies de leurs séquences respectives numérotées. Au-dessus de la ligne correspondant à la chaîne (A ou B) sont représentées les structures secondaires : les hélices-α (représentées par des boucles), les brins-β (représentés par des flèches), un tour d'hélice-α (lettres TTT) et un tour de brin-β (lettres TT). La ligne en dessous de la séquence "acc" indique l'accessibilité relative calculée (blanc = inaccessible (0,1 < A), le cyan est intermédiaire (0,1 ≤ A ≤ 0,4), le bleu est accessible (0,4 < A ≤ 1)). La ligne "hyd" montre la nature hydrophobe calculée (le rose est hydrophobe (H > 1,5), le gris est intermédiaire (-1,5 ≤ H ≤ 1,5), et le cyan est hydrophile (H < 1,5)). La dernière ligne, où apparaissent principalement les lettres A et B, correspond aux interactions de cette chaîne (A ou B) avec les autres chaînes (A ou B) de la même unité asymétrique (lettres écrites droites) ou avec celles des unités asymétriques voisines (lettres en italique). Les lettres en rouge indiquent des distances de <3,2 Å, et celles en noir des distances de 3,2-5,0 Å. Pour les contacts non cristallographiques (interaction protéine-protéine), les lettres sont surlignées en jaune. Les ions de zinc correspondent au "+" et les "1" aux ponts disulfures. Les 3 CDR du vhh sont représentés par un trait de couleur.

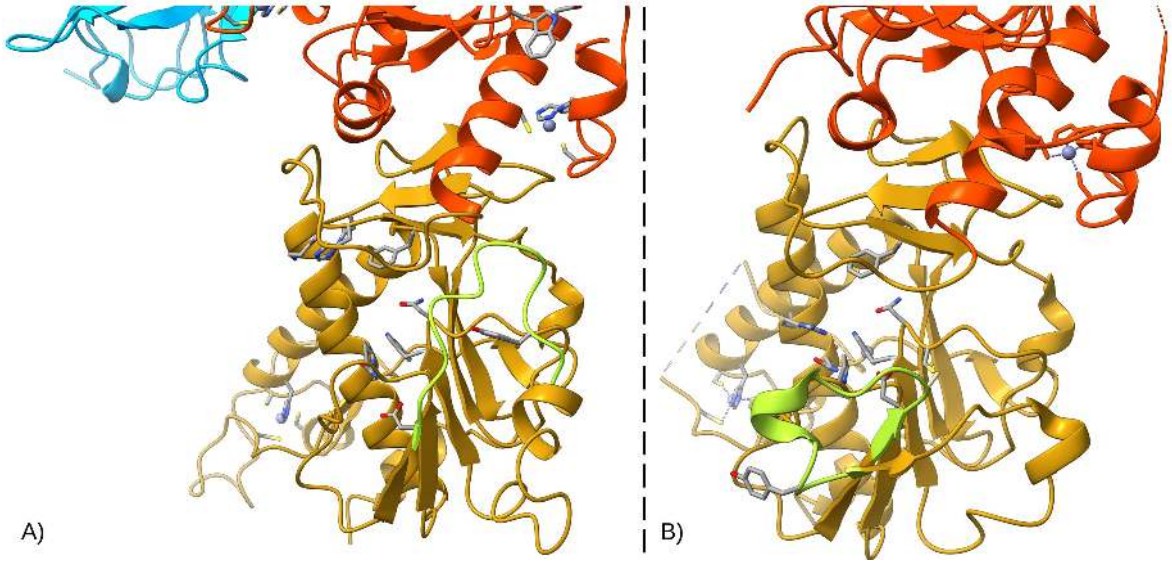


Figure 51 : Comparaison des domaines MTase de nsp14 avec la "poche à SAM" dans ses deux configurations

A) Présentation de notre structure (PDB : 9H01) avec la "poche à SAM" en position fermé.

B) Présentation de notre structure (PDB : 7R2V) avec la "poche à SAM" en position ouverte.

Pour le domaine MTase, sa structure mime un domaine méthyltransférase ; le domaine catalytique est coiffé par la région charnière, et sa conformation semble conforme aux autres structures décrites précédemment (PDB : 7R2V ; Imprachim, Yosaatmadja, et Newman 2023). Le site catalytique est vide, et ni le SAM ni son produit SAH ne sont visibles. La « poche à SAM », constituée d'une boucle flexible allant de la tyrosine Y351 à l'isoleucine I363, est dans une configuration fermée. Cette boucle, représentée en vert (Voir Figure 49 D), est en position haute (comparaison avec une poche ouverte, voir Figure 51 A et B). La tyrosine Y361, qui permet d'extraire le SAH, et la proline P335 (qui stabilise le ribose du SAM/SAH) sont observées dans la même conformation, et les autres acides aminés impliqués dans la stabilisation de la coiffe GpppN, permettant l'attaque nucléophile du SAM pour le transfert du méthyle et la formation finale du 7MeGpppN, sont également résolus et dans la conformation permettant ce type de réaction (Ogando *et al*. 2021).

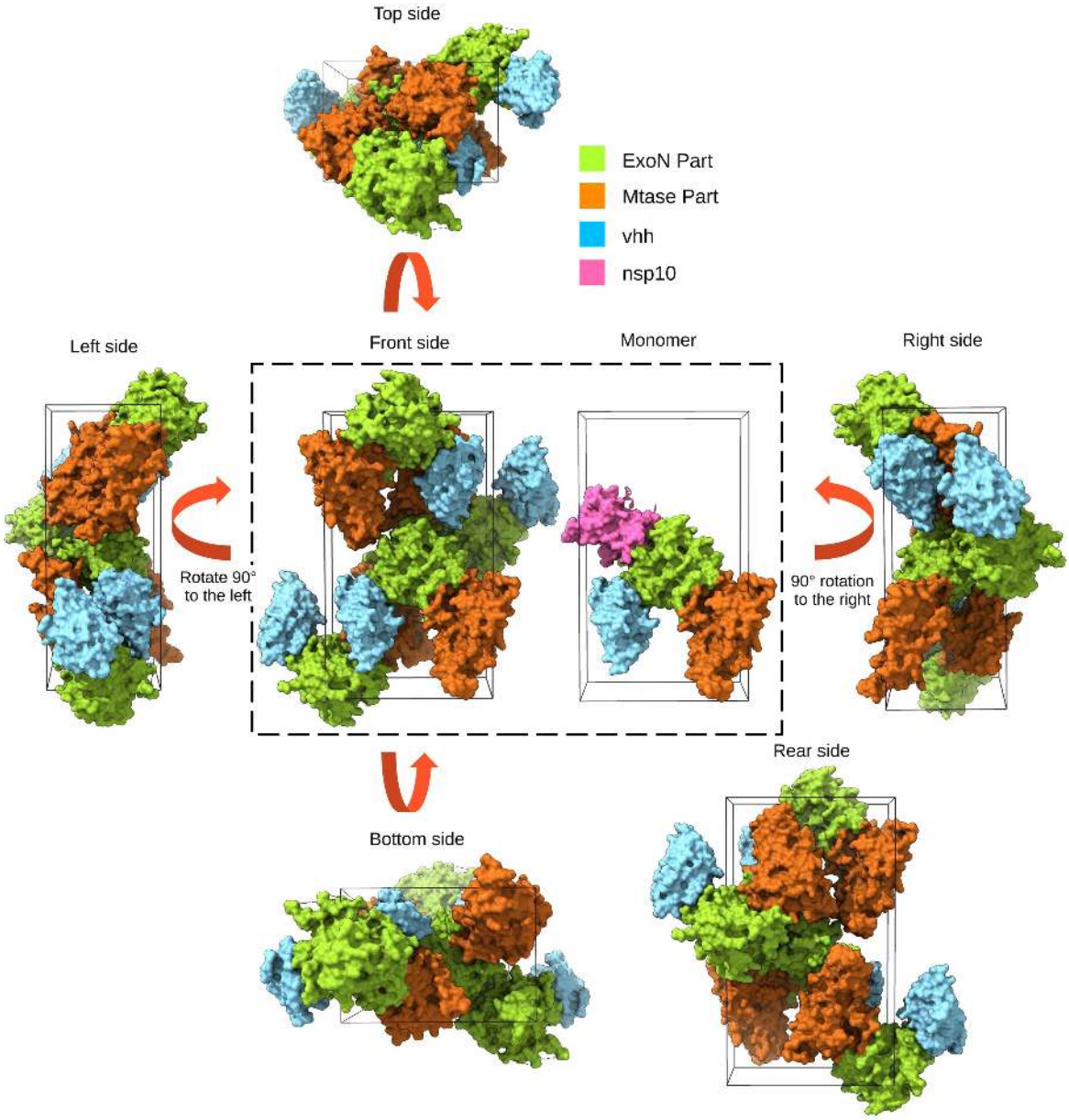


Figure 52 : Représentation de la symétrie cristallographique, ainsi que de l'emplacement théorique de nsp10

Ce schéma permet d'observer la composition de « la maille élémentaire» qui forme la symétrie du cristal. Dans l'encadré noir, à droite, nous pouvons observer le complexe nsp14/vhh (nsp14 en vert, vhh en bleu), ainsi que la position de nsp10 (en rose), si elle avait été présente dans ce complexe. À gauche de cet encadré, est représentée « la boîte » avec les quatre complexes nsp14/vhh.

Si nous nous attardons maintenant sur l'analyse de l'empilement cristallin (voir Figure 52), on peut s'apercevoir qu'il est très compact. Les deux régions du domaine MTase (en orange) sont extrêmement proches du domaine ExoN du complexe suivant (en vert) dans l'empilement cristallin (voir Figure 52, encadré central). Cela est

également visible sur la représentation ENDScript (où les lettres en italique représentent les contacts avec les autres unités asymétriques du cristal ; voir Figure 49 & Figure 50). On observe ainsi que la région ExoN allant de T5 à M57 (au-dessus du site ExoN) est en contact avec la région de K469 à A520 (ZBS3) du domaine MTase d'un premier complexe de l'unité asymétrique. De plus, la région allant de L117 à K165 du domaine ExoN (qui comporte une des deux boucles non résolues) est en contact avec la région de P335 à A371 (région de la poche à SAM) du domaine MTase.

Par ailleurs, on peut également se rendre compte du rôle que peut jouer le vhh dans l'organisation de l'empilement cristallin et la stabilisation du domaine N7-MTase de nsp14, notamment dans la région allant de R310 à S503, mais cette fois de façon indirecte (voir Figure 52). Ainsi, chaque complexe nsp14/vhh forme un U et s'imbrique en chaîne, stabilisant l'empilement cristallin et permettant ce maillage compact qui contribue à la stabilisation générale de l'ensemble du complexe. Concernant le vhh dans ce maillage, le CDR1 semble peu impliqué dans l'interaction avec nsp14, avec seulement deux résidus à des distances de 3,2-5,0 Å (voir Figure 49 & Figure 50). Cela pourrait suggérer une influence de ce maillage sur le positionnement global du vhh, entraînant un léger déplacement de ce dernier au sein de l'unité asymétrique.

Dans ce contexte globale, nous pouvons mieux comprendre pourquoi le complexe nsp14/nsp10/vhh est observé en solution (purifié sur gel-filtration, caractérisé en SEC-MALS et en SEC-SAXS), mais que lors de la cristallogenèse, nsp10 est exclue du complexe tripartite. Cela ne devrait pas empêcher l'utilisation des cristaux pour de futurs criblages d'inhibiteurs potentiels, tant au niveau du domaine MTase que du domaine ExoN, même si pour ce dernier, nous devrons ajouter les ions Mg^{2+} après la formation du cristal pour remplacer les ions Ca^{2+} nécessaires à la formation du cristal. Ce processus sera décrit dans la partie suivante.

Par ailleurs, après le début de mes travaux, une autre structure de nsp14 tronquée des 7 premiers acides aminés (PDB : 7QGI) a été publiée (Imprachim, Yosaatmadja, et Newman 2023). Cette structure est semblable à celle décrite ici, sans gain de résolution ni amélioration de la flexibilité. Notre stratégie quant à elle, utilise des ions Ca^{2+} qu'il faut retirer pour les remplacer par des ions Mg^{2+} lorsque les composés à co-cristalliser le nécessitent.

Ainsi, ces deux approches ont été mises en œuvre au sein de notre laboratoire en parallèle par un autre membre de l'équipe, lors des premiers essais de criblages d'inhibiteurs potentiels, qui seront décrits dans la prochaine partie.

(c) Premiers essais de criblage d'inhibiteurs

Après avoir obtenu les cristaux de nsp14/vhh, nous avons voulu débuter différents essais de co-cristallisations ou de trempages des cristaux avec des composés ciblant la MTase, analogues du SAM (Hernández *et al*. 2022), et d'autres ciblant l'ExoN, chélateurs des ions Mg^{2+} du site actif (Hausdorff *et al*. 2023).

Concernant les analogues du SAM, du fait que "la poche" à SAM soit en position fermée après la cristallogenèse, la co-cristallisation a été privilégiée pour ces inhibiteurs, et quelques essais de trempage ont également été réalisés. Plusieurs concentrations de composés ont été utilisées (1.5 à 20 de composés pour la co-cristallisation, et 1mM de composé dans le puit pour les essais de trampage), mais les structures obtenues lors des premiers essais de diffraction ne comportaient pas de composés.

Pour les chélateurs d'ions magnésium du site actif de l'ExoN, deux étapes intermédiaires ont été ajoutées avant de procéder au trempage des cristaux avec les composés. Ainsi, après la formation des cristaux, un ajout de 5 mM d'EDTA pour chélater les ions Ca^{2+} de la goutte où les cristaux ont poussé a été réalisé. Ensuite, les cristaux sont transférés dans une nouvelle goutte avec les mêmes conditions de tampons qu'auparavant, et avec un ajout de 5 mM d'ions Mg^{2+} ainsi que les composés à tester (1 mM de composé dans le puits). Les cristaux testés supportaient très bien les changements de tampons et d'ions ; toutefois, les premiers essais réalisés n'ont pas permis pour le moment d'obtenir de co-structures avec ces composés.

En parallèle, des essais de co-cristallisation et de trempage ont été réalisés avec la nsp14 tronquée des 7 premiers acides aminés (Imprachim, Yosaatmadja, et Newman 2023), mais les premiers cristaux diffractés n'ont pas non plus permis d'obtenir de co-structure avec ces composés.

4.4.4. Conclusions

La problématique de départ, qui consistait à stabiliser nsp14 pour pouvoir la cristalliser, a pu être résolue grâce à l'obtention de trois familles de vhh capables de la stabiliser. Après une première analyse biophysique, seul le vhh1 a été retenu pour poursuivre l'étude structurale et fonctionnelle des interactions de ce vhh avec nsp14.

La caractérisation en SEC-MALS et en SEC-SAXS a confirmé que le complexe nsp14/nsp10 stabilisé par le vhh1 était viable en solution.

Les essais des deux activités enzymatiques de nsp14 en présence du vhh n'ont pas permis de relever de différences d'activité pour le domaine ExoN pour le moment (même si un dernier test d'assemblage du complexe dans l'ordre nsp14/vhh puis nsp10 doit être refait). Pour le domaine N7-MTase, une augmentation de l'activité a pu être observée lorsque le vhh était ajouté directement sur nsp14.

Lors de la cristallogenèse, l'empilement cristallin exclut nsp10 du complexe tripartite. Le site catalytique de l'ExoN est correctement formé, bien qu'aucun ion ne soit visible dans le site catalytique de la structure apo. Ces conditions de cristallisation permettent d'obtenir ces cristaux, qui résistent au changement d'ions et peuvent être utilisés pour réaliser des trempages d'inhibiteurs potentiels.

Pour le site N7-MTase, en raison de la position fermée de la « poche à SAM », les expériences de trempage ont échoué et l'ajout de molécules devra certainement se faire dès la cristallogenèse, en co-cristallisation, pour permettre la caractérisation. Cette approche semble raisonnable du fait du nombre limité de contacts de la boucle avec des régions rigides.

Ainsi, les deux sites catalytiques de cette structure de nsp14 stabilisée par un vhh sont correctement formés, conformes aux descriptions précédentes dans la littérature (Lin *et al*. 2021 ; Ogando *et al*. 2021), et nous permettent de présenter ici une nouvelle stratégie cristallographique pour le criblage d'inhibiteurs potentiels.

4.4.5. Perspectives

Une partie importante de la caractérisation du vhh2 a également été réalisée. Bien que ce dernier présente une constante de dissociation plus élevée que le vhh1, il est possible qu'une des sous-familles de ce dernier soit plus favorable pour l'utilisation souhaitée. Si un autre vhh est identifié, on pourrait envisager de réaliser d'autres essais de cristallogenèse avec ce nouveau variant du vhh2, voire d'essayer de combiner cette nouvelle sous-famille avec le vhh1 pour stabiliser le complexe nsp14/nsp10. Cela peut également s'appliquer aux sous-familles de la troisième famille de vhh identifiée.

Les vhh sont également couramment utilisés pour augmenter la taille d'un complexe de protéines afin de le rendre observable en microscopie électronique. Dans le cas du complexe nsp14/nsp10 (environ 77 kDa), un vhh de 14 kDa ne suffirait pas à augmenter suffisamment la taille de notre objet. Ainsi, la transformation du vhh1 en Megabody, ajoutant ainsi environ 56 ou 100 kDa selon les modèles choisis (Uchański *et al*. 2021), permettrait de dépasser le seuil symbolique des 100 kDa, facilitant ainsi le travail en Cryo-EM (Harrison *et al*. 2023). Cependant, cette solution est plus contraignante que la cristallographie aux rayons X et ne peut pas être appliquée pour des criblages de composés à grande échelle ; elle serait plutôt utilisée après avoir identifié une cible prometteuse, compte tenu du temps et du coût élevés de cette méthode.

L'obtention de trois familles de vhh spécifiques de la nsp14 de SARS-CoV-2 nous ouvre également d'autres perspectives d'application pour ces derniers. Un cas d'usage serait d'utiliser ces familles de vhh et leurs interactions protéine-protéine comme outils pour différents types de diagnostics (détection du virus SARS-CoV-2 sur un domaine peu variable de ses protéines).

Il serait également intéressant de savoir si ces familles de vhh perturbent ou non la formation et le fonctionnement du RTC de SARS-CoV-2. Par exemple, si un des vhh peut se fixer au RTC sans perturber son activité, ces interactions protéine-protéine pourraient être utilisées en microscopie optique à fluorescence en couplant ce vhh avec un fluorophore tel que la GFP (Rothbauer *et al*. 2006), pour les utiliser comme outils de suivi du RTC dans des cellules infectées.

À l'inverse, un vhh inhibant la formation du RTC pourrait être utilisé comme protéine thérapeutique, soit directement, soit en n'utilisant que les CDR du vhh fusionnés à une autre protéine, afin de faciliter leur entrée et passage à travers la membrane plasmique.

5. Conclusion générale

Ce projet de thèse s'est intéressé à plusieurs aspects de la réplication des virus à long génome d'ARN.

Une part importante de ce projet a été consacrée aux coronavirus et à l'étude de leurs RTC. Cela a débuté par la recherche de techniques pour la production des enzymes du RTC des virus SARS-CoV, puis SARS-CoV-2, et notamment des protéines nsp12, pour une étude structurale et fonctionnelle complète in vitro. Cela a également permis le criblage d'inhibiteurs potentiels de ce RTC et la découverte de l'analogue nucléotidique (NA) AT-9010. L'étude plus large de ce dernier a mené à plusieurs découvertes, notamment que le virus MERS semblait moins sensible à cet NA, ce qui nous a conduits à vouloir étudier le RTC de MERS.

Malgré de nombreuses similarités avec les deux SARS, une divergence évolutive existe entre ces virus. Le virus MERS semble présenter des particularités qui compliquent son étude. De nombreux efforts ont été mis en place pour l'étudier et notamment pour travailler sur les protéines nsp12, nsp7 et nsp8 de ce virus de manière isolée. Ce travail est toujours en cours aujourd'hui, et des pistes prometteuses sont à l'étude.

La recherche sur le cas particulier du MERS contribue à l'élargissement de nos connaissances générales des virus à ARN(+). Elle nous aiderait également à mieux comprendre les mécanismes partagés avec d'autres familles de virus, et pas seulement chez les coronavirus. Cela facilite les parallèles que l'on peut établir d'une famille de virus à l'autre. Nous devons ainsi élargir nos champs de connaissances et les virus que nous choisissons d'étudier.

L'étude du virus WB en est une illustration. Chez ce virus de poisson, nous avons retrouvé une N7-MTase localisée dans l'ORF1a, alors que ce type d'enzyme est habituellement retrouvé dans l'ORF1b chez d'autres virus. Cela nous permet d'élargir encore notre vision et nos hypothèses sur les possibles événements évolutifs qui ont pu avoir lieu, tant au niveau du génome dans son intégralité qu'au niveau de la séquence protéique elle-même.

Un autre exemple est la découverte que l'analogue de nucléotide AT-9010, qui agit habituellement comme un terminateur de chaîne, semble, dans le cas du virus de la dengue, l'inhiber de plusieurs façons et agir sur plusieurs enzymes.

Plus largement, la pandémie de SARS-CoV-2 et la détérioration générale des niches écologiques nous rappellent que cela augmente le risque de nouvelles pandémies à l'avenir.

Le fait que plusieurs antiviraux (AT-9010, Remdesivir, etc.) commencent à être étudiés et certifiés contre plusieurs virus ouvre la voie à l'étude des synergies entre ces composés, qui commencent également à être décrites. Cela pourrait ainsi augmenter l'efficacité globale des traitements, tout en limitant l'apparition potentielle de variants de résistance.

En conclusion, ce travail offre une compréhension large de ces virus, contribuant à renforcer nos connaissances communes en virologie ainsi que les liens qui peuvent être établis entre les familles de virus. Cela, combiné à la découverte et à l'étude de nouveaux antiviraux, pourrait potentiellement nous amener un jour à constituer une chimiothèque de composés semblable à celle des antibiotiques. Une telle chimiothèque, combinée à nos connaissances, pourrait ainsi nous aider à nous prémunir en cas d'émergence de nouveaux virus.

6. Bibliographie

Abergel, Chantal, Matthieu Legendre, et Jean-Michel Claverie. 2015. « The Rapidly Expanding Universe of Giant Viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus ». Édité par Alain Filloux. *FEMS Microbiology Reviews* 39 (6): 779-96. https://doi.org/10.1093/femsre/fuv037.

Ahmed-Belkacem, Rostom, Marcel Hausdorff, Adrien Delpal, Priscila Sutto-Ortiz, Agathe M. G. Colmant, Franck Touret, Natacha S. Ogando, *et al*. 2022. « Potent Inhibition of SARS-CoV-2 Nsp14 *N* 7-Methyltransferase by Sulfonamide-Based Bisubstrate Analogues ». *Journal of Medicinal Chemistry* 65 (8): 6231-49. https://doi.org/10.1021/acs.jmedchem.2c00120.

Ali, Karim, Tanweer Azher, Mahin Baqi, Alexandra Binnie, Sergio Borgia, François M. Carrier, Yiorgos Alexandroa Cavayas, *et al*. 2022. « Remdesivir for the Treatment of Patients in Hospital with COVID-19 in Canada: A Randomized Controlled Trial ». *Canadian Medical Association Journal* 194 (7): E242-51. https://doi.org/10.1503/cmaj.211698.

Amani, Behnam, et Bahman Amani. s. d. « Efficacy and Safety of Nirmatrelvir/Ritonavir (Paxlovid) for COVID-19: A Rapid Review and Meta-analysis ». https://doi.org/10.1002/jmv.28441.

Angelini, Megan M., Marzieh Akhlaghpour, Benjamin W. Neuman, et Michael J. Buchmeier. 2013. « Severe Acute Respiratory Syndrome Coronavirus Nonstructural Proteins 3, 4, and 6 Induce Double-Membrane Vesicles ». Édité par Anne Moscona. *mBio* 4 (4): e00524-13. https://doi.org/10.1128/mBio.00524-13.

Atkins, John F., Gary Loughran, Pramod R. Bhatt, Andrew E. Firth, et Pavel V. Baranov. 2016. « Ribosomal Frameshifting and Transcriptional Slippage: From Genetic Steganography and Cryptography to Adventitious Use ». *Nucleic Acids Research*, juillet, gkw530. https://doi.org/10.1093/nar/gkw530.

Bakheit, Ahmed H., Hany Darwish, Ibrahim A. Darwish, et Ahmed I. Al-Ghusn. 2023. « Remdesivir ». In *Profiles of Drug Substances, Excipients and Related Methodology*, 48:71-108. Elsevier. https://doi.org/10.1016/bs.podrm.2022.11.003.

Baltimore, D. 1971. « Expression of Animal Virus Genomes ». *Bacteriological Reviews* 35 (3): 235-41. https://doi.org/10.1128/br.35.3.235-241.1971.

Barthelemy, Roxane-Marie, Taichiro Goto, et Eric Faure. 2019. « Serendipitous Discovery in a Marine Invertebrate (Phylum Chaetognatha) of the Longest Giant Viruses Reported till Date » 3 (1). https://amu.hal.science/hal-02010229.

Bawden, F. C., N. W. Pirie, J. D. Bernal, et I. Fankuchen. 1936. « Liquid Crystalline Substances from Virus-Infected Plants ». *Nature* 138 (3503): 1051-52. https://doi.org/10.1038/1381051a0.

Bedwell, Gregory J, et Alan N Engelman. 2021. « Factors That Mold the Nuclear Landscape of HIV-1 Integration ». *Nucleic Acids Research* 49 (2): 621-35. https://doi.org/10.1093/nar/gkaa1207.

Beijerinck M.W., Über ein contagium vivum fluidum als Ursache der Fleckenkrankheit der Tabaksblätter, Verhandelingen der Koninklyke akademie van Wettenschappen te Amsterdam (traduction anglaise : Concerning a contagium vivum fluidum as cause of the spot disease of tobacco leaves), in : Johnson J. (Ed.), Phytopathological Classics n° 7, American Phytopathological Society, Saint Paul, USA, 1942, pp. 33–52.

Belouzard, Sandrine, Victor C. Chu, et Gary R. Whittaker. 2009. « Activation of the SARS Coronavirus Spike Protein via Sequential Proteolytic Cleavage at Two Distinct Sites ». *Proceedings of the National Academy of Sciences* 106 (14): 5871-76. https://doi.org/10.1073/pnas.0809524106.

Bhatt, Pramod R., Alain Scaiola, Gary Loughran, Marc Leibundgut, Annika Kratzel, Romane Meurs, René Dreos, *et al*. 2021. « Structural Basis of Ribosomal Frameshifting during Translation of the SARS-CoV-2 RNA Genome ». *Science* 372 (6548): 1306-13. https://doi.org/10.1126/science.abf3546.

Birtles, Rj, Tj Rowbotham, C Storey, Tj Marrie, et D Raoult. 1997. « Chlamydia-like Obligate Parasite of Free-Living Amoebae ». *The Lancet* 349 (9056): 925-26. https://doi.org/10.1016/S0140-6736(05)62701-8.

Biswal, Mahamaya, Stephen Diggs, Duo Xu, Nelli Khudaverdyan, Jiuwei Lu, Jian Fang, Gregor Blaha, Rong Hai, et Jikui Song. 2021. « Two Conserved Oligomer Interfaces of NSP7 and NSP8 Underpin the Dynamic Assembly of SARS-CoV-2 RdRP ». *Nucleic Acids Research* 49 (10): 5956-66. https://doi.org/10.1093/nar/gkab370.

Bordería, Antonio V., Kathryn Rozen-Gagnon, et Marco Vignuzzi. 2015. « Fidelity Variants and RNA Quasispecies ». In *Quasispecies: From Theory to Experimental Systems*, édité par Esteban Domingo et Peter Schuster, 392:303-22. Current Topics in Microbiology and Immunology. Cham: Springer International Publishing. https://doi.org/10.1007/82_2015_483.

Bornstein, Paul, et Gary Balian. 1977. « [14] Cleavage at AsnGly Bonds with Hydroxylamine ». In *Methods in Enzymology*, 47:132-45. Elsevier. https://doi.org/10.1016/0076-6879(77)47016-2.

Bouvet, Mickaël, François Ferron, Isabelle Imbert, Laure Gluais, Barbara Selisko, Bruno Coutard, Bruno Canard, et Etienne Decroly. 2012. « Stratégies de formation de la structure coiffe chez les virus à ARN ». *médecine/sciences* 28 (4): 423-29. https://doi.org/10.1051/medsci/2012284021.

Bouvet, Mickaël, Isabelle Imbert, Lorenzo Subissi, Laure Gluais, Bruno Canard, et Etienne Decroly. 2012. « RNA 3'-End Mismatch Excision by the Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein Nsp10/Nsp14 Exoribonuclease Complex ». *Proceedings of the National Academy of Sciences* 109 (24): 9372-77. https://doi.org/10.1073/pnas.1201130109.

Brant, Ayslan Castro, Wei Tian, Vladimir Majerciak, Wei Yang, et Zhi-Ming Zheng. 2021. « SARS-CoV-2: From Its Discovery to Genome Structure, Transcription, and Replication ». *Cell & Bioscience* 11 (1): 136. https://doi.org/10.1186/s13578-021-00643-z.

Cavanagh D. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. Arch Virol. 1997;142(3):629-33. PMID: 9349308.

Campagnola, Grace, Seth McDonald, Stéphanie Beaucourt, Marco Vignuzzi, et Olve B. Peersen. 2015. « Structure-Function Relationships Underlying the Replication Fidelity of Viral RNA-Dependent RNA Polymerases ». Édité par K. Kirkegaard. *Journal of Virology* 89 (1): 275-86. https://doi.org/10.1128/JVI.01574-14.

Cao, Yu-chen, Qi-xin Deng, et Shi-xue Dai. 2020. « Remdesivir for Severe Acute Respiratory Syndrome Coronavirus 2 Causing COVID-19: An Evaluation of the Evidence ». *Travel Medicine and Infectious Disease* 35 (mai):101647. https://doi.org/10.1016/j.tmaid.2020.101647.

Castro, Christian, Eric D Smidansky, Jamie J Arnold, Kenneth R Maksimchuk, Ibrahim Moustafa, Akira Uchida, Matthias Götte, William Konigsberg, et Craig E Cameron. 2009. « Nucleic Acid Polymerases Use a General Acid for Nucleotidyl Transfer ».

Nature Structural & Molecular Biology 16 (2): 212-18. https://doi.org/10.1038/nsmb.1540.

Černý, Jiří, Barbora Černá Bolfíková, James J. Valdés, Libor Grubhoffer, et Daniel Růžek. 2014. « Evolution of Tertiary Structure of Viral RNA Dependent Polymerases ». Édité par Chandravanu Dash. *PLoS ONE* 9 (5): e96070. https://doi.org/10.1371/journal.pone.0096070.

Chamberland C., sur un filtre donnant de l'eau physiologiquement pure, C. R. Acad. Sci. Paris 99 (1884) 247–248.)

Chen, James, Brandon Malone, Eliza Llewellyn, Michael Grasso, Patrick M.M. Shelton, Paul Dominic B. Olinares, Kashyap Maruthi, *et al*. 2020. « Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex ». *Cell* 182 (6): 1560-1573.e13. https://doi.org/10.1016/j.cell.2020.07.033.

Chen, Yu, Hui Cai, Ji'an Pan, Nian Xiang, Po Tien, Tero Ahola, et Deyin Guo. 2009. « Functional Screen Reveals SARS Coronavirus Nonstructural Protein Nsp14 as a Novel Cap N7 Methyltransferase ». *Proceedings of the National Academy of Sciences* 106 (9): 3484-89. https://doi.org/10.1073/pnas.0808790106.

Chrisman, Brianna Sierra, Kelley Paskov, Nate. Stockham, Kevin Tabatabaei, Jae-Yoon Jung, Peter Washington, Maya Varma, Min Woo Sun, Sepideh Maleki, et Dennis P. Wall. 2021. « Indels in SARS-CoV-2 Occur at Template-Switching Hotspots ». *BioData Mining* 14 (1): 20. https://doi.org/10.1186/s13040-021-00251-0.

Claverie, Jean-Michel. 2006. « Viruses Take Center Stage in Cellular Evolution ». *Genome Biology* 7 (6): 110. https://doi.org/10.1186/gb-2006-7-6-110.

Claverie, Jean-Michel, et Chantal Abergel. 2009. « Mimivirus and Its Virophage ». *Annual Review of Genetics* 43 (1): 49-66. https://doi.org/10.1146/annurev-genet-102108-134255.

Claverie, Jean-Michel. 2013. « Open Questions About Giant Viruses ». In *Advances in Virus Research*, 85:25-56. Elsevier. https://doi.org/10.1016/B978-0-12-408116-1.00002-1.

Claverie, Jean-Michel, Chantal Abergel, et Matthieu Legendre. 2018. « Des virus géants créateurs de leurs propres gènes ? » *médecine/sciences* 34 (12): 1087-91. https://doi.org/10.1051/medsci/2018300.

Colson, Philippe, Natalya Yutin, Svetlana A. Shabalina, Catherine Robert, Ghislain Fournous, Bernard La Scola, Didier Raoult, et Eugene V. Koonin. 2011. « Viruses with More Than 1,000 Genes: Mamavirus, a New Acanthamoeba Polyphagamimivirus Strain, and Reannotation of Mimivirus Genes ». *Genome Biology and Evolution* 3 (janvier):737-42. https://doi.org/10.1093/gbe/evr048.

Decroly, Etienne, Claire Debarnot, François Ferron, Mickael Bouvet, Bruno Coutard, Isabelle Imbert, Laure Gluais, *et al*. 2011. « Crystal Structure and Functional Analysis of the SARS-Coronavirus RNA Cap 2′-O-Methyltransferase Nsp10/Nsp16 Complex ». Édité par Félix A. Rey. *PLoS Pathogens* 7 (5): e1002059. https://doi.org/10.1371/journal.ppat.1002059.

Decroly, Etienne, François Ferron, Julien Lescar, et Bruno Canard. 2012. « Conventional and Unconventional Mechanisms for Capping Viral mRNA ». *Nature Reviews Microbiology* 10 (1): 51-65. https://doi.org/10.1038/nrmicro2675.

Deng, Jian, Yi Zheng, Sheng-Nan Zheng, Mei-Ling Nan, Lulu Han, Jing Zhang, Yunyun Jin, Ji-An Pan, Chengjiang Gao, et Pei-Hui Wang. 2023. « SARS-CoV-2 NSP7 Inhibits Type I and III IFN Production by Targeting the RIG-I/MDA5, TRIF, and STING Signaling Pathways ». *Journal of Medical Virology* 95 (3): e28561. https://doi.org/10.1002/jmv.28561.

Di, Han, Joseph C. Madden, Esther K. Morantz, Hsin-Yao Tang, Rachel L. Graham, Ralph S. Baric, et Margo A. Brinton. 2017. « Expanded Subgenomic mRNA Transcriptome and

Coding Capacity of a Nidovirus ». *Proceedings of the National Academy of Sciences* 114 (42). https://doi.org/10.1073/pnas.1706696114.

Dolja, Valerian V., et Eugene V. Koonin. 2018. « Metagenomics Reshapes the Concepts of RNA Virus Evolution by Revealing Extensive Horizontal Virus Transfer ». *Virus Research* 244 (janvier):36-52. https://doi.org/10.1016/j.virusres.2017.10.020.

Duffy, Siobain, Laura A. Shackelton, et Edward C. Holmes. 2008. « Rates of Evolutionary Change in Viruses: Patterns and Determinants ». *Nature Reviews Genetics* 9 (4): 267-76. https://doi.org/10.1038/nrg2323.

Fanales-Belasio, Emanuele, Mariangela Raimondo, Barbara Suligoi, et Stefano Buttò. 2010. « HIV Virology and Pathogenetic Mechanisms of Infection: A Brief Overview ». *Annali dellIstituto Superiore Di Sanità* 46 (1). https://doi.org/10.1590/S0021-25712010000100002.

Fehr, Anthony R., et Stanley Perlman. 2015. « Coronaviruses: An Overview of Their Replication and Pathogenesis ». In *Coronaviruses*, édité par Helena Jane Maier, Erica Bickerton, et Paul Britton, 1282:1-23. Methods in Molecular Biology. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-2438-7_1.

Feng, Kuan, Hui-Jiao Zhang, Yuan-Qin Min, Min Zhou, Fei Deng, Hua-Lin Wang, Pei-Qing Li, et Yun-Jia Ning. 2023. « SARS-CoV-2 NSP13 Interacts with Host IRF3, Blocking Antiviral Immune Responses ». *Journal of Medical Virology* 95 (6): e28881. https://doi.org/10.1002/jmv.28881.

Ferron, François, Humberto J Debat, Ashleigh Shannon, Etienne Decroly, et Bruno Canard. 2020. « A N7-Guanine RNA Cap Methyltransferase Signature-Sequence as a Genetic Marker of Large Genome, Non-Mammalian Tobaniviridae ». *NAR Genomics and Bioinformatics* 2 (1): lqz022. https://doi.org/10.1093/nargab/lqz022.

Ferron, François, Bhawna Sama, Etienne Decroly, et Bruno Canard. 2021. « The Enzymes for Genome Size Increase and Maintenance of Large (+)RNA Viruses ». *Trends in Biochemical Sciences* 46 (11): 866-77. https://doi.org/10.1016/j.tibs.2021.05.006.

Ferron, François, Lorenzo Subissi, Ana Theresa Silveira De Morais, Nhung Thi Tuyet Le, Marion Sevajol, Laure Gluais, Etienne Decroly, *et al*. 2018. « Structural and Molecular Basis of Mismatch Correction and Ribavirin Excision from Coronavirus RNA ». *Proceedings of the National Academy of Sciences* 115 (2). https://doi.org/10.1073/pnas.1718806115.

Forni, Diego, Rachele Cagliani, Mario Clerici, et Manuela Sironi. 2017. « Molecular Evolution of Human Coronavirus Genomes ». *Trends in Microbiology* 25 (1): 35-48. https://doi.org/10.1016/j.tim.2016.09.001.

Forterre, Patrick. 2006. « The Origin of Viruses and Their Possible Roles in Major Evolutionary Transitions ». *Virus Research* 117 (1): 5-16. https://doi.org/10.1016/j.virusres.2006.01.010.

Forterre, Patrick, et Mart Krupovic. 2012. « The Origin of Virions and Virocells: The Escape Hypothesis Revisited ». In *Viruses: Essential Agents of Life*, édité par Günther Witzany, 43-60. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-4899-6_3.

Franklin, Rosalind E., et R. G. Gosling. 1953. « Molecular Configuration in Sodium Thymonucleate ». *Nature* 171 (4356): 740-41. https://doi.org/10.1038/171740a0.

Freeman, Megan Culler, Christopher T. Peek, Michelle M. Becker, Everett Clinton Smith, et Mark R. Denison. 2014. « Coronaviruses Induce Entry-Independent, Continuous Macropinocytosis ». Édité par Michael J. Buchmeier. *mBio* 5 (4): e01340-14. https://doi.org/10.1128/mBio.01340-14.

Gadhave, Kundlik, Prateek Kumar, Ankur Kumar, Taniya Bhardwaj, Neha Garg, et Rajanish Giri. 2021. « Conformational Dynamics of 13 Amino Acids Long NSP11 of SARS-

CoV-2 under Membrane Mimetics and Different Solvent Conditions ». *Microbial Pathogenesis* 158 (septembre):105041. https://doi.org/10.1016/j.micpath.2021.105041.

Gao, Yan, Liming Yan, Yucen Huang, Fengjiang Liu, Yao Zhao, Lin Cao, Tao Wang, *et al*. 2020. « Structure of the RNA-Dependent RNA Polymerase from COVID-19 Virus ». *Science* 368 (6492): 779-82. https://doi.org/10.1126/science.abb7498.

Gierer, A., et G. Schramm. 1956. « Infectivity of Ribonucleic Acid from Tobacco Mosaic Virus ». *Nature* 177 (4511): 702-3. https://doi.org/10.1038/177702a0.

Glhoom, Shaymaa. 2024. « The Efficacy of Tixagevimab/Cilgavimab (Evusheld) in Prophylaxis and Treatment of COVID-19 in Immunocompromised Patients: A Systematic Review and Meta-Analysis ». https://doi.org/10.1186/s40001-023-01549-x.

Godwin, Patrick O., Bryan Polsonetti, Michael F. Caron, et Thomas F. Oppelt. 2024. « Remdesivir for the Treatment of COVID-19: A Narrative Review ». *Infectious Diseases and Therapy* 13 (1): 1-19. https://doi.org/10.1007/s40121-023-00900-3.

Good, Steven S., Jonna Westover, Kie Hoon Jung, Xiao-Jian Zhou, Adel Moussa, Paolo La Colla, Gabriella Collu, Bruno Canard, et Jean-Pierre Sommadossi. 2021. « AT-527, a Double Prodrug of a Guanosine Nucleotide Analog, Is a Potent Inhibitor of SARS-CoV-2 *In Vitro* and a Promising Oral Antiviral for Treatment of COVID-19 ». *Antimicrobial Agents and Chemotherapy* 65 (4): e02479-20. https://doi.org/10.1128/AAC.02479-20.

Gorbalenya, A.E., C. Lauber, et S. Siddell. 2019. « Taxonomy of Viruses ». In *Reference Module in Biomedical Sciences*, B9780128012383992377. Elsevier. https://doi.org/10.1016/B978-0-12-801238-3.99237-7.

Gorbalenya, Alexander E., Luis Enjuanes, John Ziebuhr, et Eric J. Snijder. 2006. « Nidovirales: Evolving the Largest RNA Virus Genome ». *Virus Research* 117 (1): 17-37. https://doi.org/10.1016/j.virusres.2006.01.017.

Griesemer, Sara B, Laura D Kramer, Greta A Van Slyke, Janice D Pata, David W Gohara, Craig E Cameron, *et Al*exander T Ciota. 2017. « Mutagen Resistance and Mutation Restriction of St. Louis Encephalitis Virus ». *Journal of General Virology* 98 (2): 201-11. https://doi.org/10.1099/jgv.0.000682.

Harrison, Peter J., Tereza Vecerkova, Daniel K. Clare, et Andrew Quigley. 2023. « A Review of the Approaches Used to Solve Sub-100 kDa Membrane Proteins by Cryo-Electron Microscopy ». *Journal of Structural Biology* 215 (2): 107959. https://doi.org/10.1016/j.jsb.2023.107959.

Hassan, Haitham, Jeanne Chiavaralli, Afnan Hassan, Loay Bedda, Tim Krischuns, Kuang-Yu Chen, Alice Shi Ming Li, *et al*. 2023. « Design and Synthesis of Naturally-Inspired SARS-CoV-2 Inhibitors ». *RSC Medicinal Chemistry* 14 (3): 507-19. https://doi.org/10.1039/D2MD00149G.

Hastie, Kathryn M., Christopher R. Kimberlin, Michelle A. Zandonatti, Ian J. MacRae, et Erica Ollmann Saphire. 2011. « Structure of the Lassa Virus Nucleoprotein Reveals a dsRNA-Specific 3′ to 5′ Exonuclease Activity Essential for Immune Suppression ». *Proceedings of the National Academy of Sciences* 108 (6): 2396-2401. https://doi.org/10.1073/pnas.1016404108.

Hausdorff, Marcel, Adrien Delpal, Sarah Barelier, Laura Nicollet, Bruno Canard, Franck Touret, Agathe Colmant, et al. 2023. « Structure-Guided Optimization of Adenosine Mimetics as Selective and Potent Inhibitors of Coronavirus Nsp14 N7-Methyltransferases ». *European Journal of Medicinal Chemistry* 256 (août):115474. https://doi.org/10.1016/j.ejmech.2023.115474.

Hernández, Sergio, Mikael Feracci, Carolina Trajano De Jesus, Priscila El Kazzi, Rafik Kaci, Laura Garlatti, Clemence Mondielli, *et al*. 2022. « Identification of Potent Inhibitors of

Arenavirus and SARS-CoV-2 Exoribonucleases by Fluorescence Polarization Assay ». *Antiviral Research* 204 (août):105364. https://doi.org/10.1016/j.antiviral.2022.105364.
Hilgenfeld, Rolf. 2014. « From SARS to MERS : Crystallographic Studies on Coronaviral Proteases Enable Antiviral Drug Design ». *The FEBS Journal* 281 (18): 4085-96. https://doi.org/10.1111/febs.12936.
Hurst, Kelley R., Cheri A. Koetzner, et Paul S. Masters. 2009. « Identification of In Vivo-Interacting Domains of the Murine Coronavirus Nucleocapsid Protein ». *Journal of Virology* 83 (14): 7221-34. https://doi.org/10.1128/JVI.00440-09.
Imbert, Isabelle, Jean-Claude Guillemot, Jean-Marie Bourhis, Cécile Bussetta, Bruno Coutard, Marie-Pierre Egloff, François Ferron, Alexander E Gorbalenya, et Bruno Canard. 2006. « A second, non-canonical RNA-dependent RNA polymerase in SARS Coronavirus ». *The EMBO Journal* 25 (20): 4933-42. https://doi.org/10.1038/sj.emboj.7601368.
Imprachim, Nergis, Yuliana Yosaatmadja, et Joseph A Newman. 2023. « Crystal Structures and Fragment criblage of SARS-CoV-2 NSP14 Reveal Details of Exoribonuclease Activation and mRNA Capping and Provide Starting Points for Antiviral Drug Development ». *Nucleic Acids Research* 51 (1): 475-87. https://doi.org/10.1093/nar/gkac1207.
Irigoyen, Nerea, Andrew E. Firth, Joshua D. Jones, Betty Y.-W. Chung, Stuart G. Siddell, et Ian Brierley. 2016. « High-Resolution Analysis of Coronavirus Gene Expression by RNA Sequencing and Ribosome Profiling ». Édité par Matthew B. Frieman. *PLOS Pathogens* 12 (2): e1005473. https://doi.org/10.1371/journal.ppat.1005473.
Ivanovski D., Über die Mosaikkrankheit der Tabakspflanze, St. Petersb. Acad. Imp. Sci. Bul. 35 (1892) 67-70 (traduction anglaise : Concerning the mosaic disease of tobacco plant), in : Johnson J. (Ed.), Phytopathological Classics n° 7, American Phytopathological Society, Saint Paul, USA, 1942, pp. 27–30.)
Jackson, Cody B., Michael Farzan, Bing Chen, et Hyeryun Choe. 2022. « Mechanisms of SARS-CoV-2 Entry into Cells ». *Nature Reviews Molecular Cell Biology* 23 (1): 3-20. https://doi.org/10.1038/s41580-021-00418-x.
Ji, Mingming, Meng Li, Long Sun, Hongyu Zhao, Ying Li, Lulu Zhou, Zhenni Yang, *et al*. 2022. « VMP1 and TMEM41B Are Essential for DMV Formation during β-Coronavirus Infection ». *Journal of Cell Biology* 221 (6): e202112081. https://doi.org/10.1083/jcb.202112081.
Kausche, G. A., E. Pfankuch, et H. Ruska. 1939. « Die Sichtbarmachung von pflanzlichem Virus im Übermikroskop ». *Die Naturwissenschaften* 27 (18): 292-99. https://doi.org/10.1007/BF01493353.
Knoops, Kèvin, Marjolein Kikkert, Sjoerd H. E. Van Den Worm, Jessika C Zevenhoven-Dobbe, Yvonne Van Der Meer, Abraham J Koster, A. Mieke Mommaas, et Eric J Snijder. 2008. « SARS-Coronavirus Replication Is Supported by a Reticulovesicular Network of Modified Endoplasmic Reticulum ». Édité par Michael Emerman. *PLoS Biology* 6 (9): e226. https://doi.org/10.1371/journal.pbio.0060226.
Koonin, Eugene V, et Valerian V Dolja. 2013. « A Virocentric Perspective on the Evolution of Life ». *Current Opinion in Virology* 3 (5): 546-57. https://doi.org/10.1016/j.coviro.2013.06.008.
Koonin, Eugene V., Valerian V. Dolja, Mart Krupovic, Arvind Varsani, Yuri I. Wolf, Natalya Yutin, F. Murilo Zerbini, et Jens H. Kuhn. 2020. « Global Organization and Proposed Megataxonomy of the Virus World ». *Microbiology and Molecular Biology Reviews* 84 (2): e00061-19. https://doi.org/10.1128/MMBR.00061-19.
Krupovic, Mart, et Eugene V. Koonin. 2017. « Multiple Origins of Viral Capsid Proteins from Cellular Ancestors ». *Proceedings of the National Academy of Sciences* 114 (12). https://doi.org/10.1073/pnas.1621061114.

Kuhn, Jens H., Yuri I. Wolf, Mart Krupovic, Yong-Zhen Zhang, Piet Maes, Valerian V. Dolja, et Eugene V. Koonin. 2019. « Classify Viruses — the Gain Is Worth the Pain ». *Nature* 566 (7744): 318-20. https://doi.org/10.1038/d41586-019-00599-8.

La Scola, Bernard, Christelle Desnues, Isabelle Pagnier, Catherine Robert, Lina Barrassi, Ghislain Fournous, Michèle Merchat, *et al*. 2008. « The Virophage as a Unique Parasite of the Giant Mimivirus ». *Nature* 455 (7209): 100-104. https://doi.org/10.1038/nature07218.

Lai, Michael M.C., et David Cavanagh. 1997. « The Molecular Biology of Coronaviruses ». In *Advances in Virus Research*, 48:1-100. Elsevier. https://doi.org/10.1016/S0065-3527(08)60286-9.

Lai, Vicky C. H., C. Cheng Kao, Eric Ferrari, Justin Park, Annette S. Uss, Jacquelyn Wright-Minogue, Zhi Hong, et Johnson Y. N. Lau. 1999. « Mutational Analysis of Bovine Viral Diarrhea Virus RNA-Dependent RNA Polymerase ». *Journal of Virology* 73 (12): 10129-36. https://doi.org/10.1128/jvi.73.12.10129-10136.1999.

Le Poder, Sophie. 2021. « Les différents Coronavirus connus chez les animaux ; risques de zoonoses ? » *Revue Neurologique* 177 (avril):S136. https://doi.org/10.1016/j.neurol.2021.02.014.

Legendre, Matthieu, Defne Arslan, Chantal Abergel, et Jean-Michel Claverie. 2012. « Genomics of Megavirus and the Elusive Fourth Domain of Life ». *Communicative & Integrative Biology* 5 (1): 102-6. https://doi.org/10.4161/cib.18624.

Legendre, Matthieu, Elisabeth Fabre, Olivier Poirot, Sandra Jeudy, Audrey Lartigue, Jean-Marie Alempic, Laure Beucher, *et al*. 2018. « Diversity and Evolution of the Emerging Pandoraviridae Family ». *Nature Communications* 9 (1): 2285. https://doi.org/10.1038/s41467-018-04698-4.

Lehmann, Kathleen C., Anastasia Gulyaeva, Jessika C. Zevenhoven-Dobbe, George M. C. Janssen, Mark Ruben, Hermen S. Overkleeft, Peter A. van Veelen, *et al*. 2015. « Discovery of an Essential Nucleotidylating Activity Associated with a Newly Delineated Conserved Domain in the RNA Polymerase-Containing Protein of All Nidovirus es ». *Nucleic Acids Research* 43 (17): 8416-34. https://doi.org/10.1093/nar/gkv838.

Li, Aixin, Bei Zhang, Kaitao Zhao, Zhinang Yin, Yan Teng, Lu Zhang, Zaichao Xu, Kaiwei Liang, Xiaoming Cheng, et Yuchen Xia. 2023. « SARS-CoV-2 Nsp13 Restricts Episomal DNA Transcription without Affecting Chromosomal DNA ». Édité par Kanta Subbarao. *Journal of Virology* 97 (7): e00512-23. https://doi.org/10.1128/jvi.00512-23.

Li, Guangdi, Yali Wang, et Erik De Clercq. 2022. « Approved HIV Reverse Transcriptase Inhibitors in the Past Decade ». *Acta Pharmaceutica Sinica B* 12 (4): 1567-90. https://doi.org/10.1016/j.apsb.2021.11.009.

Li, Shiqian, Dai Xiao, Luwen Zhang, Rui Chen, Daili Song, Yiping Wen, Rui Wu, *et al*. 2023. « Porcine DeltaCoronavirus Enters ST Cells by Clathrin-Mediated Endocytosis and Does Not Require Rab5, Rab7, or Rab11 ». Édité par Daniela S. Rajao. *Microbiology Spectrum* 11 (6): e02553-23. https://doi.org/10.1128/spectrum.02553-23.

Liao, Ying, Huan Wang, Huiyu Liao, Yingjie Sun, Lei Tan, Cuiping Song, Xusheng Qiu, et Chan Ding. 2024. « Classification, Replication, and Transcription of Nidovirales ». *Frontiers in Microbiology* 14 (janvier):1291761. https://doi.org/10.3389/fmicb.2023.1291761.

Lin, Sheng, Hua Chen, Zimin Chen, Fanli Yang, Fei Ye, Yue Zheng, Jing Yang, *et al*. 2021. « Crystal Structure of SARS-CoV-2 Nsp10 Bound to Nsp14-ExoN Domain Reveals an Exoribonuclease with Both Structural and Functional Integrity ». *Nucleic Acids Research* 49 (9): 5382-92. https://doi.org/10.1093/nar/gkab320.

Liu, Chang, Wei Shi, Scott T. Becker, David G. Schatz, Bin Liu, et Yang Yang. 2021. « Structural Basis of Mismatch Recognition by a SARS-CoV-2 Proofreading Enzyme ». *Science*, juillet, eabi9310. https://doi.org/10.1126/science.abi9310.

Ma, Yuanyuan, Lijie Wu, Neil Shaw, Yan Gao, Jin Wang, Yuna Sun, Zhiyong Lou, Liming Yan, Rongguang Zhang, et Zihe Rao. 2015. « Structural Basis and Functional Analysis of the SARS Coronavirus Nsp14–Nsp10 Complex ». *Proceedings of the National Academy of Sciences* 112 (30): 9436-41. https://doi.org/10.1073/pnas.1508686112.

Lustig, A, et A J Levine. 1992. « One Hundred Years of Virology ». *Journal of Virology* 66 (8): 4629-31. https://doi.org/10.1128/jvi.66.8.4629-4631.1992.

Luytjes, Willem, Peter J. Bredenbeek, Ans F.H. Noten, Marian C. Horzinek, et Willy J.M. Spaan. 1988. « Sequence of Mouse Hepatitis Virus A59 mRNA 2: Indications for RNA Recombination between Coronaviruses and Influenza C Virus ». *Virology* 166 (2): 415-22. https://doi.org/10.1016/0042-6822(88)90512-0.

Lwoff, A. 1957. « The Concept of Virus ». *Microbiology* 17 (2): 239-53. https://doi.org/10.1099/00221287-17-2-239.

Lwoff, André, et Paul Tournier. 1966. « The Classification of Viruses ». *Annual Review of Microbiology* 20 (1): 45-74. https://doi.org/10.1146/annurev.mi.20.100166.000401.

Malone, Brandon, Nadya Urakova, Eric J. Snijder, et Elizabeth A. Campbell. 2022. « Structures and Functions of Coronavirus Replication–Transcription Complexes and Their Relevance for SARS-CoV-2 Drug Design ». *Nature Reviews Molecular Cell Biology* 23 (1): 21-39. https://doi.org/10.1038/s41580-021-00432-z.

Martínez-Sobrido, Luis, Benson Yee Hin Cheng, et Juan Carlos De La Torre. 2016. « Reverse Genetics Approaches to Control Arenavirus ». In *Vaccine Design*, édité par Sunil Thomas, 1403:313-51. Methods in Molecular Biology. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4939-3387-7_17.

Mayer A., Über die Mosaikkrankheit des Tabaks, die Landwirtschaftlichen Versuchs-Stationen 32 (1886) 451-467 (traduction anglaise: Concerning the mosaic disease of tobacco), in : Johnson J. (Ed.), Phytopathological Classics n° 7, American Phytopathological Society, Saint Paul, USA, 1942, pp. 11–24)

Meng, Tao, et Jimmy Kwang. 2014. « Attenuation of Human Enterovirus 71 High-Replication-Fidelity Variants in AG129 Mice ». Édité par K. Kirkegaard. *Journal of Virology* 88 (10): 5803-15. https://doi.org/10.1128/jvi.00289-14.

Mielech, Anna M., Yafang Chen, Andrew D. Mesecar, et Susan C. Baker. 2014. « Nidovirus Papain-like Proteases: Multifunctional Enzymes with Protease, Deubiquitinating and deISGylating Activities ». *Virus Research* 194 (décembre):184-90. https://doi.org/10.1016/j.virusres.2014.01.025.

Millet, Jean Kaoru, et Gary R. Whittaker. 2014. « Host Cell Entry of Middle East Respiratory Syndrome Coronavirus after Two-Step, Furin-Mediated Activation of the Spike Protein ». *Proceedings of the National Academy of Sciences* 111 (42): 15214-19. https://doi.org/10.1073/pnas.1407087111.

Naeli, Parisa, Xu Zhang, Patric Harris Snell, Susanta Chatterjee, Muhammad Kamran, Reese Jalal Ladak, Nick Orr, Thomas Duchaine, Nahum Sonenberg, et Seyed Mehdi Jafarnejad. 2023. « The SARS-CoV-2 Protein NSP2 Enhances microRNA-Mediated Translational Repression ». *Journal of Cell Science* 136 (19): jcs261286. https://doi.org/10.1242/jcs.261286.

Narayanan, Krishna, Sydney I. Ramirez, Kumari G. Lokugamage, et Shinji Makino. 2015. « Coronavirus Nonstructural Protein 1: Common and Distinct Functions in the Regulation of Host and Viral Gene Expression ». *Virus Research* 202 (avril):89-100. https://doi.org/10.1016/j.virusres.2014.11.019.

Neuman, Benjamin W., Megan M. Angelini, et Michael J. Buchmeier. 2014. « Does Form Meet Function in the Coronavirus Replicative Organelle? » *Trends in Microbiology* 22 (11): 642-47. https://doi.org/10.1016/j.tim.2014.06.003.

Neuman, Benjamin W., Gabriella Kiss, Andreas H. Kunding, David Bhella, M. Fazil Baksh, Stephen Connelly, Ben Droese, *et al*. 2011. « A Structural Analysis of M Protein in Coronavirus Assembly and Morphology ». *Journal of Structural Biology* 174 (1): 11-22. https://doi.org/10.1016/j.jsb.2010.11.021.

Nishihara, Kazuyo, Masaaki Kanemori, Hideki Yanagi, et Takashi Yura. 2000. « Overexpression of Trigger Factor Prevents Aggregation of Recombinant Proteins in *Escherichia Coli* ». *Applied and Environmental Microbiology* 66 (3): 884-89. https://doi.org/10.1128/AEM.66.3.884-889.2000.

Ogando, Natacha S., Francois Ferron, Etienne Decroly, Bruno Canard, Clara C. Posthuma, et Eric J. Snijder. 2019. « The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity ». *Frontiers in Microbiology* 10 (août):1813. https://doi.org/10.3389/fmicb.2019.01813.

Ogando, Natacha S., Priscila El Kazzi, Jessika C. Zevenhoven-Dobbe, Brenda W. Bontes, Alice Decombe, Clara C. Posthuma, Volker Thiel, *et al*. 2021. « Structure–Function Analysis of the Nsp14 N7–Guanine Methyltransferase Reveals an Essential Role in *Betacoronavirus* Replication ». *Proceedings of the National Academy of Sciences* 118 (49): e2108709118. https://doi.org/10.1073/pnas.2108709118.

Ogino, Tomoaki, et Amiya K. Banerjee. 2007. « Unconventional Mechanism of mRNA Capping by the RNA-Dependent RNA Polymerase of Vesicular Stomatitis Virus ». *Molecular Cell* 25 (1): 85-97. https://doi.org/10.1016/j.molcel.2006.11.013.

Ogino, Tomoaki, et Todd J. Green. 2019. « RNA Synthesis and Capping by Non-segmented Negative Strand RNA Viral Polymerases: Lessons From a Prototypic Virus ». *Frontiers in Microbiology* 10 (juillet):1490. https://doi.org/10.3389/fmicb.2019.01490.

Örd, Mihkel, Ilona Faustova, et Mart Loog. 2020. « The Sequence at Spike S1/S2 Site Enables Cleavage by Furin and Phospho-Regulation in SARS-CoV2 but Not in SARS-CoV1 or MERS-CoV ». *Scientific Reports* 10 (1): 16944. https://doi.org/10.1038/s41598-020-74101-0.

Orthwein, Tim, Luciano Huergo, Karl Forchhammer, et Khaled Selim. 2021. « Kinetic Analysis of a Protein-Protein Complex to Determine Its Dissociation Constant (KD) and the Effective Concentration (EC50) of an Interplaying Effector Molecule Using Bio-Layer Interferometry ». *BIO-PROTOCOL* 11 (17). https://doi.org/10.21769/BioProtoc.4152.

Otter, Clayton J., Nicole Bracci, Nicholas A. Parenti, Chengjin Ye, Abhishek Asthana, Ebba K. Blomqvist, Li Hui Tan, *et al*. 2024. « SARS-CoV-2 Nsp15 Endoribonuclease Antagonizes dsRNA-Induced Antiviral Signaling ». *Proceedings of the National Academy of Sciences* 121 (15): e2320194121. https://doi.org/10.1073/pnas.2320194121.

Papageorgiou, Nicolas, Afroditi Vaitsopoulou, Awa Diop, Thi Hong Van Nguyen, Bruno Canard, Karine Alvarez, et François Ferron. 2021. « Observation of Arenavirus Nucleoprotein Heptamer Assembly ». *FEBS Open Bio* 11 (4): 1076-83. https://doi.org/10.1002/2211-5463.13106.

Paramasivam, Arumugam. 2020. « RNA 2′-O-Methylation Modification and Its Implication in COVID-19 Immunity ». *Cell Death Discovery* 6 (1): 118. https://doi.org/10.1038/s41420-020-00358-z.

Park, Gina J., Adam Osinski, Genaro Hernandez, Jennifer L. Eitson, Abir Majumdar, Marco Tonelli, Katie Henzler-Wildman, *et al*. 2022. « The Mechanism of RNA Capping by SARS-CoV-2 ». *Nature*, août. https://doi.org/10.1038/s41586-022-05185-z.

Peng, Guiqing, Dawei Sun, Kanagalaghatta R. Rajashankar, Zhaohui Qian, Kathryn V. Holmes, et Fang Li. 2011. « Crystal Structure of Mouse Coronavirus Receptor-Binding Domain Complexed with Its Murine Receptor ». *Proceedings of the National Academy of Sciences* 108 (26): 10696-701. https://doi.org/10.1073/pnas.1104306108.

Philippe, Nadège, Matthieu Legendre, Gabriel Doutre, Yohann Couté, Olivier Poirot, Magali Lescot, Defne Arslan, *et al*. 2013. « Pandoraviruses: Amoeba Viruses with Genomes Up to 2.5 Mb Reaching That of Parasitic Eukaryotes ». *Science* 341 (6143): 281-86. https://doi.org/10.1126/science.1239181.

Prentice, Erik, Josephine McAuliffe, Xiaotao Lu, Kanta Subbarao, et Mark R. Denison. 2004. « Identification and Characterization of Severe Acute Respiratory Syndrome Coronavirus Replicase Proteins ». *Journal of Virology* 78 (18): 9977-86. https://doi.org/10.1128/JVI.78.18.9977-9986.2004.

Pruijssers, Andrea J, et Mark R Denison. 2019. « Nucleoside Analogues for the Treatment of Coronavirus Infections ». *Current Opinion in Virology* 35 (avril):57-62. https://doi.org/10.1016/j.coviro.2019.04.002.

Quiocho, Florante A, Guanghui Hu, et Paul D Gershon. 2000. « Structural Basis of mRNA Cap Recognition by Proteins ». *Current Opinion in Structural Biology* 10 (1): 78-86. https://doi.org/10.1016/S0959-440X(99)00053-6.

Raoult, Didier, Stéphane Audic, Catherine Robert, Chantal Abergel, Patricia Renesto, Hiroyuki Ogata, Bernard La Scola, Marie Suzan, et Jean-Michel Claverie. 2004. « The 1.2-Megabase Genome Sequence of Mimivirus ». *Science* 306 (5700): 1344-50. https://doi.org/10.1126/science.1101485.

Ravi, V., Sonal Saxena, et Pragyan Swagatika Panda. 2022. « Basic Virology of SARS-CoV 2 ». *Indian Journal of Medical Microbiology* 40 (2): 182-86. https://doi.org/10.1016/j.ijmmb.2022.02.005.

Ricciardi, Simona, Andrea Maria Guarino, Laura Giaquinto, Elena V. Polishchuk, Michele Santoro, Giuseppe Di Tullio, Cathal Wilson, *et al*. 2022. « The Role of NSP6 in the Biogenesis of the SARS-CoV-2 Replication Organelle ». *Nature* 606 (7915): 761-68. https://doi.org/10.1038/s41586-022-04835-6.

Rodrigues, Rodrigo Al, Fernanda G De Souza, Bruna L De Azevedo, Lorena Cf Da Silva, et Jônatas S Abrahão. 2021. « The Morphogenesis of Different Giant Viruses as Additional Evidence for a Common Origin of Nucleocytoviricota ». *Current Opinion in Virology* 49 (août):102-10. https://doi.org/10.1016/j.coviro.2021.05.004.

Roe, Molly K., Nathan A. Junod, Audrey R. Young, Dia C. Beachboard, et Christopher C. Stobart. 2021. « Targeting Novel Structural and Functional Features of Coronavirus Protease Nsp5 (3CLpro, Mpro) in the Age of COVID-19 ». *Journal of General Virology* 102 (3). https://doi.org/10.1099/jgv.0.001558.

Rothbauer, Ulrich, Kourosh Zolghadr, Sergei Tillib, Danny Nowak, Lothar Schermelleh, Anja Gahl, Natalija Backmann, *et al*. 2006. « Targeting and Tracing Antigens in Live Cells with Fluorescent Nanobodies ». *Nature Methods* 3 (11): 887-89. https://doi.org/10.1038/nmeth953.

Rubin, Rita. 2022. « Evusheld Treatment Timing Set ». *JAMA* 328 (7): 612-612. https://doi.org/10.1001/jama.2022.12742.

Saez, Natalie J., Hervé Nozach, Marilyne Blemont, et Renaud Vincentelli. 2014. « High Throughput Quantitative Expression criblage and Purification Applied to Recombinant Disulfide-Rich Venom Proteins Produced in E. Coli ». *Journal of Visualized Experiments*, nº 89 (juillet), 51464. https://doi.org/10.3791/51464.

Sanderson, Theo, Ryan Hisner, I'ah Donovan-Banfield, Hassan Hartman, Alessandra Løchen, Thomas P. Peacock, et Christopher Ruis. 2023. « A Molnupiravir-Associated

Mutational Signature in Global SARS-CoV-2 Genomes ». *Nature* 623 (7987): 594-600. https://doi.org/10.1038/s41586-023-06649-6.

Santerre, Maryline, Sterling P. Arjona, Charles Ns Allen, Natalia Shcherbik, et Bassel E. Sawaya. 2021. « Why Do SARS-CoV-2 NSPs Rush to the ER? » *Journal of Neurology* 268 (6): 2013-22. https://doi.org/10.1007/s00415-020-10197-8.

Santos-Mendoza, Teresa. 2023. « The Envelope (E) Protein of SARS-CoV-2 as a Pharmacological Target ». *Viruses* 15 (4): 1000. https://doi.org/10.3390/v15041000.

Saramago, Margarida, Cátia Bárria, Vanessa G. Costa, Caio S. Souza, Sandra C. Viegas, Susana Domingues, Diana Lousa, Cláudio M. Soares, Cecília M. Arraiano, et Rute G. Matos. 2021. « New Targets for Drug Design: Importance of Nsp14/Nsp10 Complex Formation for the 3'-5' Exoribonucleolytic Activity on SARS-CoV-2 ». *The FEBS Journal* 288 (17): 5130-47. https://doi.org/10.1111/febs.15815.

Schmidt, Nora, Sabina Ganskih, Yuanjie Wei, Alexander Gabel, Sebastian Zielinski, Hasmik Keshishian, Caleb A. Lareau, et al. 2023. « SND1 Binds SARS-CoV-2 Negative-Sense RNA and Promotes Viral RNA Synthesis through NSP9 ». *Cell* 186 (22): 4834-4850.e23. https://doi.org/10.1016/j.cell.2023.09.002.

Schulz, Frederik, Natalya Yutin, Natalia N. Ivanova, Davi R. Ortega, Tae Kwon Lee, Julia Vierheilig, Holger Daims, *et al*. 2017. « Giant Viruses with an Expanded Complement of Translation System Components ». *Science* 356 (6333): 82-85. https://doi.org/10.1126/science.aal4657.

Scola, Bernard La, Stéphane Audic, Catherine Robert, Liang Jungang, Xavier De Lamballerie, Michel Drancourt, Richard Birtles, Jean-Michel Claverie, et Didier Raoult. 2003. « A Giant Virus in Amoebae ». *Science* 299 (5615): 2033-2033. https://doi.org/10.1126/science.1081867.

Selisko, Barbara, Nicolas Papageorgiou, François Ferron, et Bruno Canard. 2018. « Structural and Functional Basis of the Fidelity of Nucleotide Selection by Flavivirus RNA-Dependent RNA Polymerases ». *Viruses* 10 (2): 59. https://doi.org/10.3390/v10020059.

Sequeira, Ana Filipa, Jeremy Turchetto, Natalie J. Saez, Fanny Peysson, Laurie Ramond, Yoan Duhoo, Marilyne Blémont, *et al*. 2017. « Gene Design, Fusion Technology and TEV Cleavage Conditions Influence the Purification of Oxidized Disulphide-Rich Venom Peptides in Escherichia Coli ». *Microbial Cell Factories* 16 (1): 4. https://doi.org/10.1186/s12934-016-0618-0.

Shannon, Ashleigh, Aurélie Chazot, Mikael Feracci, Camille Falcou, Véronique Fattorini, Barbara Selisko, Steven Good, *et al*. 2024. « An Exonuclease-Resistant Chain-Terminating Nucleotide Analogue Targeting the SARS-CoV-2 Replicase Complex ». *Nucleic Acids Research* 52 (3): 1325-40. https://doi.org/10.1093/nar/gkad1194.

Shannon, Ashleigh, Véronique Fattorini, Bhawna Sama, Barbara Selisko, Mikael Feracci, Camille Falcou, Pierre Gauffre, *et al*. 2021. « Protein-Primed RNA Synthesis in SARS-CoVs and Structural Basis for Inhibition by AT-527 ». https://doi.org/10.1101/2021.03.23.436564.

Shannon, A. 2022. « A Dual Mechanism of Action of AT-527 against SARS-CoV-2 Polymerase ». *Nature Communications* 13 (1): 621. https://doi.org/10.1038/s41467-022-28113-1.

Shannon, Ashleigh, Nhung Thi-Tuyet Le, Barbara Selisko, Cecilia Eydoux, Karine Alvarez, Jean-Claude Guillemot, Etienne Decroly, Olve Peersen, Francois Ferron, et Bruno Canard. 2020. « Remdesivir and SARS-CoV-2: Structural Requirements at Both Nsp12 RdRp and Nsp14 Exonuclease Active-Sites ». *Antiviral Research* 178 (juin):104793. https://doi.org/10.1016/j.antiviral.2020.104793.

Siddell, Stuart G., Donald B. Smith, Evelien Adriaenssens, Poliane Alfenas-Zerbini, Bas E. Dutilh, Maria Laura Garcia, Sandra Junglen, *et al*. 2023. « Virus Taxonomy and the

Role of the International Committee on Taxonomy of Viruses (ICTV) ». *Journal of General Virology* 104 (5). https://doi.org/10.1099/jgv.0.001840.

Simeoni, Matthieu, Théo Cavinato, Daniel Rodriguez, et David Gatfield. 2021. « I(Nsp1)Ecting SARS-CoV-2–Ribosome Interactions ». *Communications Biology* 4 (1): 715. https://doi.org/10.1038/s42003-021-02265-0.

Slanina, Heiko, Ramakanth Madhugiri, Ganesh Bylapudi, Karin Schultheiß, Nadja Karl, Anastasia Gulyaeva, Alexander E. Gorbalenya, Uwe Linne, et John Ziebuhr. 2021. « Coronavirus Replication–Transcription Complex: Vital and Selective NMPylation of a Conserved Site in Nsp9 by the NiRAN-RdRp Subunit ». *Proceedings of the National Academy of Sciences* 118 (6): e2022310118. https://doi.org/10.1073/pnas.2022310118.

Small, Gabriel I., Olga Fedorova, Paul Dominic B. Olinares, Joshua Chandanani, Anoosha Banerjee, Young Joo Choi, Henrik Molina, Brian Chait, Seth A. Darst, et Elizabeth A. Campbell. 2023. « Structural and Functional Insights into the Enzymatic Plasticity of the SARS-CoV-2 NiRAN Domain ». https://doi.org/10.1101/2023.09.25.558837.

Snijder, E.J., E. Decroly, et J. Ziebuhr. 2016. « The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing ». In *Advances in Virus Research*, 96:59-126. Elsevier. https://doi.org/10.1016/bs.aivir.2016.08.008.

Snijder, Eric J., Peter J. Bredenbeek, Jessika C. Dobbe, Volker Thiel, John Ziebuhr, Leo L.M. Poon, Yi Guan, Mikhail Rozanov, Willy J.M. Spaan, et Alexander E. Gorbalenya. 2003. « Unique and Conserved Features of Genome and Proteome of SARS-Coronavirus, an Early Split-off From the Coronavirus Group 2 Lineage ». *Journal of Molecular Biology* 331 (5): 991-1004. https://doi.org/10.1016/S0022-2836(03)00865-9.

Sreelatha, Anju, Samantha S. Yee, Victor A. Lopez, Brenden C. Park, Lisa N. Kinch, Sylwia Pilch, Kelly A. Servage, *et al*. 2018. « Protein AMPylation by an Evolutionarily Conserved Pseudokinase ». *Cell* 175 (3): 809-821.e19. https://doi.org/10.1016/j.cell.2018.08.046.

Stanley, W. M. 1935. « Isolation of a Crystalline Protein Possessing the Properties of Tobacco-Mosaic Virus ». *Science* 81 (2113): 644-45. https://doi.org/10.1126/science.81.2113.644.

« State-of-the-Art Conference on Azidothymidine Therapy for Early HIV Infection. » 1990. *The American Journal of Medicine* 89 (3): 335-44. https://doi.org/10.1016/0002-9343(90)90347-g.

Steitz, T A, et J A Steitz. 1993. « A General Two-Metal-Ion Mechanism for Catalytic RNA. » *Proceedings of the National Academy of Sciences* 90 (14): 6498-6502. https://doi.org/10.1073/pnas.90.14.6498.

Stevens, Laura J., Andrea J. Pruijssers, Hery W. Lee, Calvin J. Gordon, Egor P. Tchesnokov, Jennifer Gribble, Amelia S. George, *et al*. 2022. « Mutations in the SARS-CoV-2 RNA-Dependent RNA Polymerase Confer Resistance to Remdesivir by Distinct Mechanisms ». *Science Translational Medicine* 14 (656): eabo0718. https://doi.org/10.1126/scitranslmed.abo0718.

Subissi, Lorenzo, Clara C. Posthuma, Axelle Collet, Jessika C. Zevenhoven-Dobbe, Alexander E. Gorbalenya, Etienne Decroly, Eric J. Snijder, Bruno Canard, et Isabelle Imbert. 2014. « One Severe Acute Respiratory Syndrome Coronavirus Protein Complex Integrates Processive RNA Polymerase and Exonuclease Activities ». *Proceedings of the National Academy of Sciences* 111 (37). https://doi.org/10.1073/pnas.1323705111.

Sulimov, Alexey, Danil Kutov, Ivan Ilin, Yibei Xiao, Sheng Jiang, et Vladimir Sulimov. 2022. « Novel Inhibitors of 2′-O-Methyltransferase of the SARS-CoV-2 Coronavirus ». *Molecules* 27 (9): 2721. https://doi.org/10.3390/molecules27092721.

Sun, Haoran, Jasper Fuk-Woo Chan, et Shuofeng Yuan. 2023. « Cellular Sensors and Viral Countermeasures: A Molecular Arms Race between Host and SARS-CoV-2 ». *Viruses* 15 (2): 352. https://doi.org/10.3390/v15020352.

Te Velthuis, Aartjan J. W., Jamie J. Arnold, Craig E. Cameron, Sjoerd H. E. Van Den Worm, et Eric J. Snijder. 2010. « The RNA Polymerase Activity of SARS-Coronavirus Nsp12 Is Primer Dependent ». *Nucleic Acids Research* 38 (1): 203-14. https://doi.org/10.1093/nar/gkp904.

Te Velthuis, Aartjan J.W., Sjoerd H. E. Van Den Worm, et Eric J. Snijder. 2012. « The SARS-Coronavirus Nsp7+nsp8 Complex Is a Unique Multimeric RNA Polymerase Capable of Both de Novo Initiation and Primer Extension ». *Nucleic Acids Research* 40 (4): 1737-47. https://doi.org/10.1093/nar/gkr893.

Thoms, Matthias, Robert Buschauer, Michael Ameismeier, Lennart Koepke, Timo Denk, Maximilian Hirschenberger, Hanna Kratzat, *et al*. 2020. « Structural Basis for Translational Shutdown and Immune Evasion by the Nsp1 Protein of SARS-CoV-2 ». *Science* 369 (6508): 1249-55. https://doi.org/10.1126/science.abc8665.

Trevino, Troy N., Ali A. Almousawi, KaReisha F. Robinson, Avital B. Fogel, Jake Class, Richard D. Minshall, Leon M. Tai, Justin M. Richner, et Sarah E. Lutz. 2024. « Caveolin-1 Mediates Blood-Brain Barrier Permeability, Neuroinflammation, and Cognitive Impairment in SARS-CoV-2 Infection ». *Journal of Neuroimmunology* 388 (mars):578309. https://doi.org/10.1016/j.jneuroim.2024.578309.

Tvarogová, Jana, Ramakanth Madhugiri, Ganesh Bylapudi, Lyndsey J. Ferguson, Nadja Karl, et John Ziebuhr. 2019. « Identification and Characterization of a Human Coronavirus 229E Nonstructural Protein 8-Associated RNA 3′-Terminal Adenylyltransferase Activity ». Édité par Tom Gallagher. *Journal of Virology* 93 (12): e00291-19. https://doi.org/10.1128/JVI.00291-19.

Uchański, Tomasz, Simonas Masiulis, Baptiste Fischer, Valentina Kalichuk, Uriel López-Sánchez, Eleftherios Zarkadas, Miriam Weckener, *et al*. 2021. « Megabodies Expand the Nanobody Toolkit for Protein Structure Determination by Single-Particle Cryo-EM ». *Nature Methods* 18 (1): 60-68. https://doi.org/10.1038/s41592-020-01001-6.

Vabret, A., J. Dina, E. Brison, J. Brouard, et F. Freymuth. 2009. « Coronavirus humains (HCoV) ». *Pathologie Biologie* 57 (2): 149-60. https://doi.org/10.1016/j.patbio.2008.02.018.

Van Der Hoek, Lia, Krzysztof Pyrc, et Ben Berkhout. 2006. « Human Coronavirus NL63, a New Respiratory Virus ». *FEMS Microbiology Reviews* 30 (5): 760-73. https://doi.org/10.1111/j.1574-6976.2006.00032.x.

Vithani, Neha, Michael D. Ward, Maxwell I. Zimmerman, Borna Novak, Jonathan H. Borowsky, Sukrit Singh, et Gregory R. Bowman. 2021. « SARS-CoV-2 Nsp16 Activation Mechanism and a Cryptic Pocket with Pan-Coronavirus Antiviral Potential ». *Biophysical Journal* 120 (14): 2880-89. https://doi.org/10.1016/j.bpj.2021.03.024.

V'kovski, Philip, Hawaa Al-Mulla, Volker Thiel, et Benjamin W. Neuman. 2015. « New Insights on the Role of Paired Membrane Structures in Coronavirus Replication ». *Virus Research* 202 (avril):33-40. https://doi.org/10.1016/j.virusres.2014.12.021.

Walker, Alexander P, Haitian Fan, Jeremy R Keown, Michael L Knight, Jonathan M Grimes, et Ervin Fodor. 2021. « The SARS-CoV-2 RNA Polymerase Is a Viral RNA Capping Enzyme ». *Nucleic Acids Research* 49 (22): 13019-30. https://doi.org/10.1093/nar/gkab1160.

Walpita, Pramila, et Ramon Flick. 2005. « Reverse Genetics of Negative-Stranded RNA Viruses: A Global Perspective ». *FEMS Microbiology Letters* 244 (1): 9-18. https://doi.org/10.1016/j.femsle.2005.01.046.

Wang, Dehe, Ao Jiang, Jiangpeng Feng, Guangnan Li, Dong Guo, Muhammad Sajid, Kai Wu, *et al*. 2021. « The SARS-CoV-2 Subgenome Landscape and Its Novel Regulatory Features ». *Molecular Cell* 81 (10): 2135-2147.e5. https://doi.org/10.1016/j.molcel.2021.02.036.

Wang, Quan, Jiqin Wu, Haofeng Wang, Yan Gao, Qiaojie Liu, An Mu, Wenxin Ji, *et al*. 2020. « Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase ». *Cell* 182 (2): 417-428.e13. https://doi.org/10.1016/j.cell.2020.05.034.

Watson, J. D., et F. H. C. Crick. 1953a. « Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid ». *Nature* 171 (4356): 737-38. https://doi.org/10.1038/171737a0.

Watson et Crick 1953b. « THE STRUCTURE OF DNA ». *Cold Spring Harbor Symposia on Quantitative Biology* 18 (0): 123-31. https://doi.org/10.1101/SQB.1953.018.01.020.

Wilkins, M. H. F., A. R. Stokes, et H. R. Wilson. 1953. « Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids ». *Nature* 171 (4356): 738-40. https://doi.org/10.1038/171738a0.

Witz, Jean. 1969. « La détermination de la structure des virus simples par diffraction des rayons X et microscopie électronique ». *Bulletin de la Société française de Minéralogie et de Cristallographie* 92 (6): 621-31. https://doi.org/10.3406/bulmi.1969.6296.

Yan, Liming, Ji Ge, Litao Zheng, Ying Zhang, Yan Gao, Tao Wang, Yucen Huang, *et al*. 2021. « Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis ». *Cell* 184 (1): 184-193.e10. https://doi.org/10.1016/j.cell.2020.11.016.

Yan, Liming, Yucen Huang, Ji Ge, Zhenyu Liu, Pengchi Lu, Bo Huang, Shan Gao, *et al*. 2022. « A Mechanism for SARS-CoV-2 RNA Capping and Its Inhibition by Nucleotide Analog Inhibitors ». *Cell* 185 (23): 4347-4360.e17. https://doi.org/10.1016/j.cell.2022.09.037.

Yan, Liming, Yunxiang Yang, Mingyu Li, Ying Zhang, Litao Zheng, Ji Ge, Yucen C. Huang, *et al*. 2021. « Coupling of N7-Methyltransferase and 3′-5′ Exoribonuclease with SARS-CoV-2 Polymerase Reveals Mechanisms for Capping and Proofreading ». *Cell* 184 (13): 3474-3485.e11. https://doi.org/10.1016/j.cell.2021.05.033.

Yang, Mei, Jennifer Mariano, Rebecca Su, Christopher E. Smith, Sudipto Das, Catherine Gill, Thorkell Andresson, Jadranka Loncarek, Yien Che Tsai, et Allan M. Weissman. 2023. « SARS-CoV-2 Papain-like Protease Plays Multiple Roles in Regulating Cellular Proteins in the Endoplasmic Reticulum ». *Journal of Biological Chemistry* 299 (12): 105346. https://doi.org/10.1016/j.jbc.2023.105346.

Yang, Xiaorong, Xinran Liu, Derek M. Musser, Ibrahim M. Moustafa, Jamie J. Arnold, Craig E. Cameron, et David D. Boehr. 2017. « Triphosphate Reorientation of the Incoming Nucleotide as a Fidelity Checkpoint in Viral RNA-Dependent RNA Polymerases ». *Journal of Biological Chemistry* 292 (9): 3810-26. https://doi.org/10.1074/jbc.m116.750638.

Yates, Mary K., et Katherine L. Seley-Radtke. 2019. « The Evolution of Antiviral Nucleoside Analogues: A Review for Chemists and Non-Chemists. Part II: Complex Modifications to the Nucleoside Scaffold ». *Antiviral Research* 162 (février):5-21. https://doi.org/10.1016/j.antiviral.2018.11.016.

Yazdi, Aliakbar Khalili, Paknoosh Pakarian, Sumera Perveen, Taraneh Hajian, Vijayaratnam Santhakumar, Albina Bolotokova, Fengling Li, et Masoud Vedadi. 2022. « Kinetic Characterization of SARS-CoV-2 Nsp13 ATPase Activity and Discovery of Small-Molecule Inhibitors ». *ACS Infectious Diseases* 8 (8): 1533-42. https://doi.org/10.1021/acsinfecdis.2c00165.

Yue, Kun, Bin Yao, Yingchao Shi, Yang Yang, Zhaohui Qian, Yali Ci, et Lei Shi. 2022. « The Stalk Domain of SARS-CoV-2 NSP13 Is Essential for Its Helicase Activity ». *Biochemical and Biophysical Research Communications* 601 (avril):129-36. https://doi.org/10.1016/j.bbrc.2022.02.068.

Zerbini, Francisco Murilo, Stuart G. Siddell, Elliot J. Lefkowitz, Arcady R. Mushegian, Evelien M. Adriaenssens, Poliane Alfenas-Zerbini, Donald M. Dempsey, *et al*. 2023. « Changes to Virus Taxonomy and the ICTV Statutes Ratified by the International Committee on Taxonomy of Viruses (2023) ». *Archives of Virology* 168 (7): 175. https://doi.org/10.1007/s00705-023-05797-4.

Zhang, Linlin, Daizong Lin, Xinyuanyuan Sun, Ute Curth, Christian Drosten, Lucie Sauerhering, Stephan Becker, Katharina Rox, et Rolf Hilgenfeld. 2020. « Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved α-Ketoamide Inhibitors ». *Science* 368 (6489): 409-12. https://doi.org/10.1126/science.abb3405.

Zimmermann, Liv, Xiaohan Zhao, Jana Makroczyova, Moritz Wachsmuth-Melm, Vibhu Prasad, Zach Hensel, Ralf Bartenschlager, et Petr Chlanda. 2023. « SARS-CoV-2 Nsp3 and Nsp4 Are Minimal Constituents of a Pore Spanning Replication Organelle ». *Nature Communications* 14 (1): 7894. https://doi.org/10.1038/s41467-023-43666-5.

Zuo, Y. 2001. « Exoribonuclease superfamilies: structural analysis and phylogenetic distribution ». *Nucleic Acids Research* 29 (5): 1017-26. https://doi.org/10.1093/nar/29.5.1017.

Züst, Roland, Luisa Cervantes-Barragan, Matthias Habjan, Reinhard Maier, Benjamin W Neuman, John Ziebuhr, Kristy J Szretter, *et al*. 2011. « Ribose 2′-O-Methylation Provides a Molecular Signature for the Distinction of Self and Non-Self mRNA Dependent on the RNA Sensor Mda5 ». *Nature Immunology* 12 (2): 137-43. https://doi.org/10.1038/ni.1979.

Sommaires des figures

Figure 1 : Cas d'infection d'une capside de Mimivirus par le virus Sputnik.....29
Figure 2 : Clichés de TMV.....32
Figure 3 : Classification de Baltimore.....35
Figure 4 : Schéma du Nucléotide.....36
Figure 5 : Schéma des bases azotées et de leurs liaisons hydrogènes.....37
Figure 6 : Schéma du virus de la grippe (Influenza).....39
Figure 7 : Schéma du génome ambisens d'Arenavirus.....42
Figure 8 : Diversité génomique et structurel de familles et sous familles de l'ordre des *Nidovirales* 47
Figure 9 : Schéma de la transcription des ARNm sg.....49
Figure 10 : Cycle viral type d'un Coronavirus.....50
Figure 11 : Arbre phylogénétique de la famille des Coronavirus.....54
Figure 12 : Structure schéma du virions de SARS-CoV-2.....55
Figure 13 : Description du *frameshifting* chez les Coronavirus.....58
Figure 14 : Description des ORF codant pour les Protéines Non Structurales (NSP). 59
Figure 15 : Schéma de la coiffe canonique de l'ARN.....65
Figure 16 : Schéma d'exemples de voies non canoniques de coiffage de l'ARN66
Figure 17 : Schéma de la coiffe chez les Coronavirus 67
Figure 18 : Schéma de l'étape de RNAylation de nsp9 lors de la seconde étape (GMPlysation) de la formation de la coiffe pour former l'intermédiaire GpppA-ARN68
Figure 19 : Schéma des différents sous domaines de nsp12.....71
Figure 20 : Schéma des motifs du site catalytique de la RdRp de nsp12.....73
Figure 21 : Structure du Complexe de réplication et transcription (RTC) nsp12/7/8 76
Figure 22 : Structure du Complexe de réplication et transcription (RTC) incluant nsp13 et nsp9.....78
Figure 23 : Structure hypothètique du Complexe de réplication et transcription (RTC) complet, formant un mono ou di-mère.....80
Figure 24 : Structure Cryo-EM du complexe nsp14/nsp10 à 3.90 Å 84
Figure 25 : Schéma des modifications des nucléosides analogues (NA) de l'exemple du Remdesivir.....91
Figure 26 : Schéma des molécules d'exemples de NAs (Prodrugs).....93
Figure 27 : Incorporation de l'AT-9010 et du Sofosbuvir triphosphate (STP) par le SARS-CoV-2 RTC, et l'excision par le complexe d'exonucléase de SARS-CoV-2 nsp14/10 (ExoN). 119
Figure 28 : Structure cryo-EM du RTC avec des molécules liées et des molécules AT-9010 et carte cryo-EM correspondante. 121
Figure 29 : Base structurale de l'inhibition NiRAN du SARS-CoV-2 par AT-9010.122

Figure 30 : Comparaison des structures des nsp12 de SARS-CoV, MERS et SARS-CoV-2 sous forme de représentation ESPript 126
Figure 31 : Gel SDS-Page de purification de nsp12 de MERS avec ou sans étape de sonication. 128
Figure 32 : Gel SDS-PAGE 4-20% des échantillons élution des 12 premières purifications IMAC-Ni effectuées en haut débit. 131
Figure 33 : Gel SDS-Page 12% d'une purification de nsp12 de MERS sur IMAC-Ni de 5ml 133
Figure 34 : Base fonctionnelle et structurale de l'inhibition de la MTase du DENV par l'AT-9010. 139
Figure 35 : Incorporation et inhibition de l'AT-9010 lors de la réaction d'initiation de novo par le DENV2 NS5. 141
Figure 36 : Panel de figures sur la méthyltransférase cap-dépendante du domaine N7-MT-1a du WBV. 144
Figure 37 : Séquence et structure de la N7-MT-1a de WBV. 146
Figure 38 : Graphique de fitting des phases d'associations et de dissociation pour le vhh1 151
Figure 39 : Graphique de fitting des phases d'associations et de dissociation pour le vhh1. 152
Figure 40 : Graphique de l'essai de l'association de nsp14 sur le vhh, suivi de l'ajout de nsp10 153
Figure 41 : Graphique de l'essai de l'association du complexe nsp14/nsp10 sur le vhh, suivi d'une phase de dissociation 154
Figure 42 : Superposition des essais de SEC-MALS pour le complexe nsp14/vhh1 et les protéines nsp14 ou vhh1 155
Figure 43 : Essais de SEC-MALS pour le complexe nsp14/nsp10/vhh1 156
Figure 44 : Panel de gels urée 20% pour la caractérisations de l'activité ExoN158
Figure 45 : Courbes normalisées des caractérisations de l'activité MTase 161
Figure 46 : Superposition des Fonction de distribution des distances de paires (P(r)) de différents complexes avec nsp14 164
Figure 47 : Kratky Plot 165
Figure 48 : Analyse par DAMMIF et CRYSOL des profils de diffusion en SAXS des complexes nsp14/vhh et nsp14/nsp10/vhh 167
Figure 49 : Structure de nsp14 stabilisée par un vhh (PDB : 9H01) 171
Figure 50 : Représentation ENDscript de la stucture nsp14/vhh 173
Figure 51 : Comparaison des domaines MTase de nsp14 avec la "poche à SAM" dans ses deux configurations 174
Figure 52 : Représentation de la symétrie cristallographique, ainsi que de l'emplacement théorique de nsp10 175

Sommaires des tableaux

Tableau 1 : Diversité des tailles de génomes viraux.....27
Tableau 2 : Classification du Comité international de taxonomie des virus (ICTV) de 2023, ainsi qu'un exemple de nomenclature.....44
Tableau 3 : Conditions et programme type des PCR.....101
Tableau 4 : Activité *in vitro* de l'AT-511 et d'autres médicaments antiviraux oraux contre divers Coronavirus humains. Extrait de (Good *et al*. 2021).....124
Tableau 5 : Conditions testées pour la production de la nsp12 de MERS en système bactérien, incluant différentes combinaisons de cellules, plasmides et milieux.....130
Tableau 6 : Données cristallographiques et statistiques pour la nsp14 du SARS-CoV-2 en complexe avec un nanocorps de camélidé.....170

ANNEXES

A. Publications 201
A1 : A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase 202
A2 : AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5 212
A3 : A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome 249
B. Financement du projet de thèse 264
C. Figures supplémentaires 265
C1 : Tableaux des Essais d'expression à haut débit 266
C2 : Gel SDS-PAGE 4-20% des échantillons élution effectuées en haut débit, après purifications IMAC-Ni 272
C3 : Compléments des données de SEC-SAXS 275
282

a. Publications

A1 : A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-022-28113-1 OPEN

A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase

Ashleigh Shannon[1], Véronique Fattorini[1], Bhawna Sama[1], Barbara Selisko[1], Mikael Feracci[1], Camille Falcou[1], Pierre Gauffre[1], Priscila El Kazzi[1], Adrien Delpal[1], Etienne Decroly[1], Karine Alvarez[1], Cécilia Eydoux[1], Jean-Claude Guillemot[1], Adel Moussa[2], Steven S. Good[2], Paolo La Colla[3], Kai Lin[2], Jean-Pierre Sommadossi[2], Yingxiao Zhu[4], Xiaodong Yan[4], Hui Shi[4], François Ferron[1,5] & Bruno Canard[1,5]✉

The guanosine analog AT-527 represents a promising candidate against Severe Acute Respiratory Syndrome coronavirus type 2 (SARS-CoV-2). AT-527 recently entered phase III clinical trials for the treatment of COVID-19. Once in cells, AT-527 is converted into its triphosphate form, AT-9010, that presumably targets the viral RNA-dependent RNA polymerase (RdRp, nsp12), for incorporation into viral RNA. Here we report a 2.98 Å cryo-EM structure of the SARS-CoV-2 nsp12-nsp7-nsp8$_2$-RNA complex, showing AT-9010 bound at three sites of nsp12. In the RdRp active-site, one AT-9010 is incorporated at the 3′ end of the RNA product strand. Its modified ribose group (2′-fluoro, 2′-methyl) prevents correct alignment of the incoming NTP, in this case a second AT-9010, causing immediate termination of RNA synthesis. The third AT-9010 is bound to the N-terminal domain of nsp12 - known as the NiRAN. In contrast to native NTPs, AT-9010 is in a flipped orientation in the active-site, with its guanine base unexpectedly occupying a previously unnoticed cavity. AT-9010 outcompetes all native nucleotides for NiRAN binding, inhibiting its nucleotidyltransferase activity. The dual mechanism of action of AT-527 at both RdRp and NiRAN active sites represents a promising research avenue against COVID-19.

[1] Architecture et Fonction des Macromolécules Biologiques, CNRS and Aix Marseille Université, UMR 7257, Polytech Case 925, 13009 Marseille, France. [2] Atea Pharmaceuticals, Inc., 125 Summer Street, Suite 1675, Boston, MA 02110, USA. [3] Università degli Studi di Cagliari, Monserrato, Italy. [4] Wuxi Biortus Biosciences Co. Ltd, 214437 Jiangyin, Jiangsu, China. [5] European Virus Bioinformatics Center, Leutragraben 1, 07743 Jena, Germany.
✉email: bruno.canard@univ-amu.fr

The highly conserved replicative enzymes of positive-sense RNA viruses, including the viral RdRp, remain at the forefront of drug-design strategies. Nucleoside analogs (NA) represent a promising class of RdRp inhibitors, and are currently used for the treatment of several other viral infections. NA prodrugs are metabolized in the host cell into active 5′-triphosphate forms that compete with natural nucleoside triphosphates (NTP) for incorporation into the viral RNA. This results in either chain-termination of viral RNA synthesis, or increases the viral mutation load to levels that can lethally alter the genetic make-up of the virus. Many details of NTP incorporation into viral RNA have been determined at the structural level following the pioneering work on the Reovirus RdRp[1] (reviewed in ref. [2]). However, CoVs stand out among RNA viruses for possessing an RNA-repair 3′-to-5′ exonuclease (ExoN, nsp14 stimulated by nsp10) able to excise mismatched bases as well as NAs incorporated into viral RNA, generally compromising the efficacy of these drugs[3–6]. Recently, the guanosine analog phosphoramidate prodrug AT-527 (Fig. 1a, left), was shown to act as a potent broad-spectrum anti-CoV inhibitor in a variety of cell lines[7]. It is now in phase III and II clinical trials for the treatment of COVID-19 and hepatitis C virus (HCV) infections[8], respectively[9] (www.clinicaltrials.gov/ct2/). AT-527 and its active 5′-triphosphate AT-9010 (Fig. 1a, right) carries a 2′-fluoro-2′-C-methyl modified ribose, identical to that of the clinically relevant anti-HCV uracil prodrug Sofosbuvir[10].

The CoV genome contains ~30,000 nucleotides, about three times more than that of other significant human pathogenic +RNA viruses (see ref. [11] for review). Genome size reflects the presence of novel domains, many of which remain poorly characterized. Among these is the Nidovirus RdRp-Associated Nucleotidyltransferase (NiRAN) N-terminal domain of nsp12, just upstream of the RdRp[12]. The NiRAN is structurally related to the pseudo-kinase family of enzymes[13], and has been shown to mediate the covalent transfer of nucleoside monophosphates (NMP) to viral cofactor proteins nsp7, nsp8 and nsp9, in a process known as NMPylation[14–16]. However, the role of this activity in the viral life-cycle remains unknown. The NiRAN has

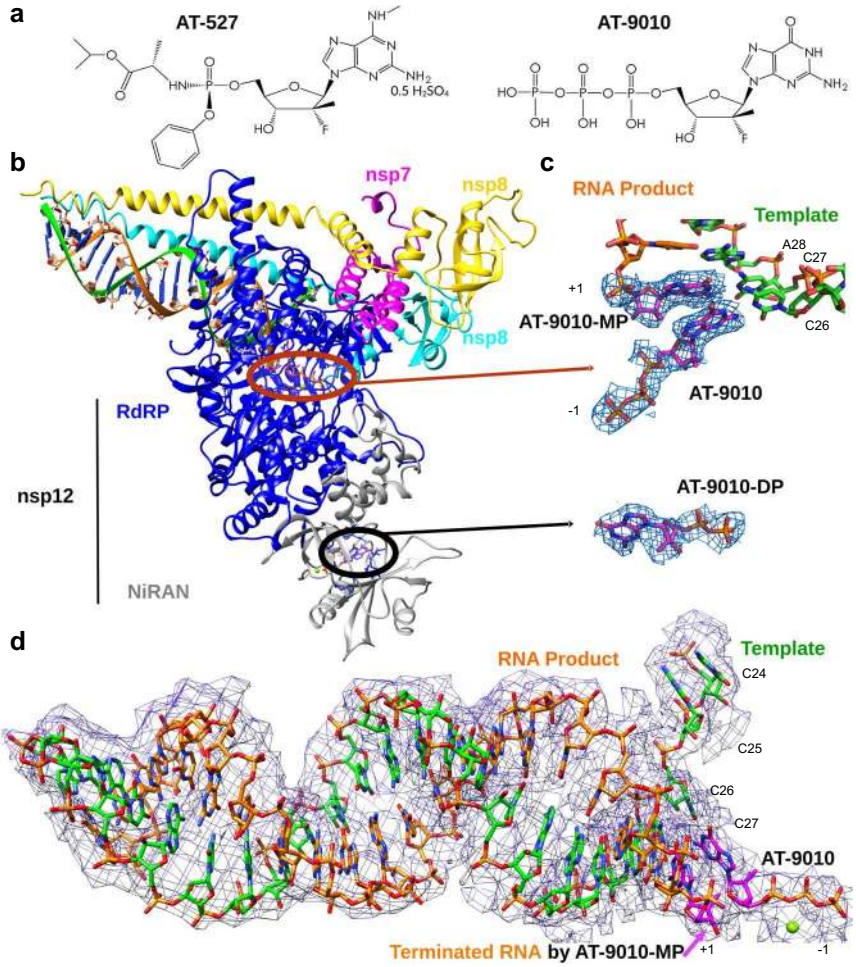


Fig. 1 Cryo-EM structure of the RTC with bound RNA and AT-9010 molecules and corresponding cryo-EM map. a Structure of the guanine analog phosphoramidate prodrug AT-527 (left) and its active triphosphate form AT-9010 (right) following activation by cellular kinases. **b** Ribbon and stick representation of the cryo-EM structure of nsp7-$(nsp8)_2$-nsp12:AT-9010-terminated-RNA:$(AT\text{-}9010)_2$ complex. RNA, AT-9010 and protein shown with the following colors: template RNA, green; RNA product, orange; AT-9010, magenta; nsp7, pink; $nsp8_1$ and $nsp8_2$, yellow and cyan respectively; and nsp12 in gray and blue for NiRAN and RdRp domains respectively. **c** Brown and black ovals are enlarged for RdRp domain and NiRAN domains, respectively, showing experimental cryo-EM map around AT-9010, In the RdRp, one AT-9010 monophosphate (AT-9010-MP) is incorporated at (+1) position, with an incoming AT-9010 occupying the (−1) position. In the NiRAN domain, AT-9010 is bound it its diphosphate form (AT-9010-DP). **d** Experimental cryo-EM map for RNA (stick representation), showing incorporated AT-9010-MP (+1) and incoming AT-9010 (−1). Bases of the template RNA involved in the interaction with AT-9010 are numbered as C24–C27. All cryo-EM maps are represented at 3.5σ representative of the general map of the entire complex at 2.98 Å resolution.

additionally been proposed to participate in the guanylyltransferase step of viral RNA capping[17]. Mutations in the NiRAN are lethal to the virus[12], making it an attractive candidate for CoV-specific drug targeting.

Here, we show that AT-9010 inhibits two separate enzyme activities of the SARS-CoV-2 replicase/transcriptase complex. First, it acts as an immediate RNA chain terminator at the RdRp active site and provides partial resistance to excision by the nsp14/nsp10 ExoN. Second, it is a NMPylation inhibitor of nsp8 and nsp9 when bound into a deep pocket located at the NiRAN active site.

Results

Structural studies of AT-9010 bound to the SARS-CoV-2 replicase:RNA complex. To investigate the AT-527/AT-9010 mechanism of action we performed cryo-EM studies with the SARS-CoV-2 minimal replication-transcription complex (RTC), comprised of nsp12 and essential cofactors nsp7 and nsp8[18,19]. The complex was formed in the presence of both AT-9010 and an annealed primer-template RNA. Image processing of single particles allowed the reconstruction of RTC:RNA:AT-9010 complexes at 2.98 Å resolution (Supplementary Table 1 and Supplementary Fig. 1).

The overall RTC conformation resembles previous structures, with one nsp12, one nsp7 and two nsp8 proteins (Fig. 1b)[20–22]. Both nsp12 (residues 4-929) and nsp7 (residues 2–73) are almost fully resolved, while the two nsp8s are mostly resolved (residues 38–191 and 43–192). As observed in ref. [23], the dsRNA stabilizes both nsp8 alpha-helical extensions in their N-terminus regions. The RdRp domain adopts a canonical right-hand fold, with fingers, palm and thumb subdomains (for review[2]). As for all viral RdRps, the active site is formed by five conserved motifs, A-E, located on the palm subdomain, with two additional motifs, F and G, located in the finger subdomains. The fully resolved N-terminus NiRAN is located alongside the palm domain of the RdRp[23,24] (Fig. 1b).

AT-9010 terminates the growing RNA chain. The 5′-end of the RNA template contains four consecutive cytidine bases (C24–C27), directing incorporation of AT-9010 into RNA (Figs. 1c, d and 2a). The 3′ end of the RNA primer, is stabilized by Ser814 bonding through a non-bridging oxygen of the last phosphodiester bond. AT-9010 is incorporated at the first position (+1), base pairing with C27 of the template strand (Fig. 1b, c and Supplementary Movie 1). A second, free AT-9010 occupies the NTP-binding site, (−1) position, but is not correctly poised for incorporation (Fig. 1c). Palm domain residues Asp618 and Asp760 in motifs A and C, respectively, coordinate a single magnesium ion, which interacts with oxygens of the α and β phosphates of the second AT-9010 (Fig. 2). Motif A and C acidic residues usually coordinate two metal ions[22]. The missing catalytic Mg^{2+} in the structure normally plays a critical role in the positioning of the 3′ end of the RNA primer and the incoming NTP.

Chain-terminated RNA and next AT-9010 disrupts the catalytic site. The second, incoming (−1) AT-9010 guanine base is partially base-paired with cytosine (C26) of the template strand, and further stabilized by motif F Lys545 and motif B Ser682, two residues important during the fidelity check prior to incorporation (Figs. 1c and 2b–d)[25,26]. The α- and β- phosphates are coordinated by the single Mg^{2+}, while the γ-phosphate is stabilized by motif A Lys621 and motif D Lys795 (Fig. 2d). The 3′-OH of the elongated primer is too far from the α-phosphate for incorporation (6.3 Å) and furthermore the metal is not correctly aligned for catalysis. Rather, we observe that AT-9010 has its α- and β-phosphates spatially overlapping the expected position of the leaving pyrophosphate (β- and γ-phosphates) (Fig. 2c), as judged by comparing with either superimposed enterovirus RdRp structures or incorporated Remdesivir by SARS-CoV-2 nsp12[22,27].

The effect of the 2′-fluoro and 2′-C-methyl ribose modifications of the incorporated AT-9010 are as follows: The replacement of the 2′ hydroxyl by a fluoro group eliminates a stabilizing interaction with Ser759 of the SDD palm domain motif C. Comparison of the SARS-CoV-2 nsp12:RNA:Remdesivir and Favipiravir-ribose 5′-triphosphate cryo-EM structures with HCV NS5:RNA:Nucleotide ternary complexes[22,28,29] shows a tight superimposition of catalytic residues, RNA backbone, and nucleotides/NAs except in the case of AT-9010 (Supplementary Fig. 2). The incoming AT-9010 ribose group is shifted in comparison to its expected position[22,28,30] (Fig. 2c, d and Supplementary Fig. 2a, b, c). Its orientation is also different to that of pre-incorporated, (−1) translocated Sofosbuvir 5′-diphosphate and GS-9813 5′-diphosphate in HCV-RdRp structures[29]. As these two NAs carry the same 2′-fluoro-2′-C-methyl ribose as AT-9010, this demonstrates that this shift is neither dependent on the nature of the base nor on the ribose modification of the incoming NA 5′-triphophate (Supplementary Fig. 2). Rather, the hydrophobic 2′-methyl group of the incorporated AT-9010 ribose creates a repulsive hydrophobic-polar contact with the ribose ring oxygen of the incoming NA (or NTP) (Fig. 2b), causing chain-termination. Following 2′-fluoro-2′-C-methyl NA incorporation into RNA by any given viral RdRp, this ribose modification may well promote a universal mechanism of chain-termination.

AT-9010 is a potent RNA chain terminator substrate. We compared AT-9010 incorporation and elongation by the purified, recombinant SARS-CoV and SARS-CoV-2 RTCs using a heteropolymeric RNA primer:template pair, corresponding to the 3′ end of the SARS-CoV-2 genome[26]. In the absence of GTP, AT-9010 is rapidly incorporated into viral RNA causing immediate termination of RNA synthesis (Fig. 3a). Even at concentrations of 500 μM of ATP (the next templated nucleotide), elongation past the incorporated AT-9010 is not observed (Supplementary Fig. 3a). Consistent with structural analysis, this indicates that chain-termination is independent of the identity of the incoming NTP. Rather, the AT-9010 ribose modification causes a misalignment that prevents further elongation (Fig. 2). Importantly, in the presence of equimolar concentrations of GTP, AT-9010 acts as a competitive guanosine substrate, discriminated against only ~5-fold (Fig. 3b). We additionally compared AT-9010 incorporation with the structurally related, uracil analog Sofosbuvir triphosphate (STP). Although STP is able to be incorporated by nsp12 as a substitute for UTP, it is not competitive with UTP, even at 5-fold higher concentrations (Supplementary Fig. 3b).

In both human bronchial and nasal epithelial cells incubated with 10 μM of the AT-527 prodrug, intracellular concentrations of AT-9010 peak at ~700 μM and 240 μM, respectively[7]. Considering an average intracellular concentrations of GTP of ~500 μM[31], it is expected that AT-9010 would be highly competitive for insertion into viral RNA, frequently terminating synthesis.

RNA chain-terminated by AT-9010 shows resistance to ExoN removal. The stalling of the polymerase following insertion of a chain-terminating NA may allow excision by the proofreading 3′-to-5′ exonuclease nsp14/nsp10[6] potentially dampening AT-9010 efficacy. Following incorporation into RNA, both AT-9010 and STP are excised by the SARS-CoV-2 ExoN (Fig. 3b, c). Interestingly, however, AT-9010 is ~4.3-fold more resistant to excision

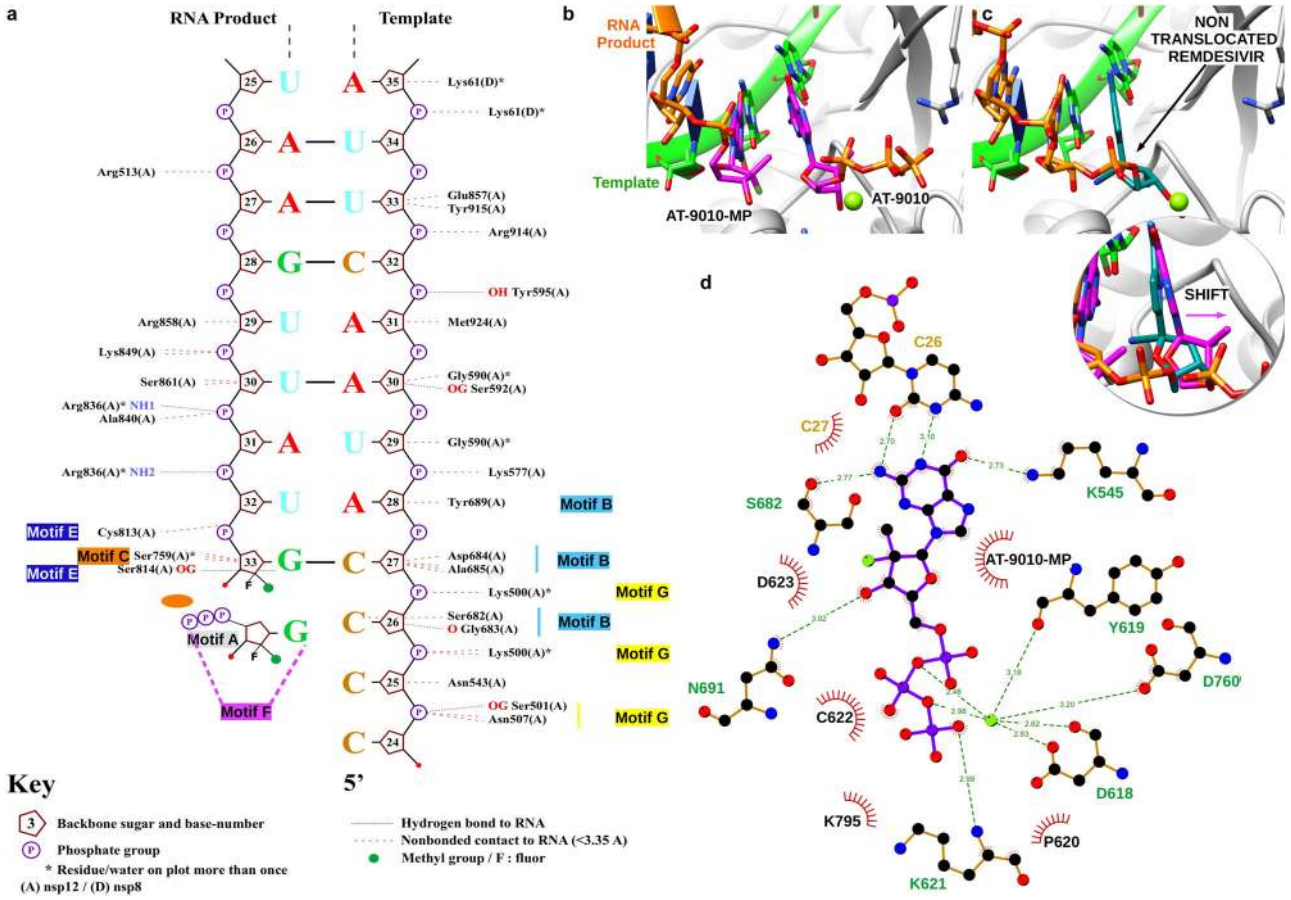


Fig. 2 Structural basis for inhibition of the SARS-CoV-2 polymerase complex by AT-9010. a Nucplot molecular analysis with RNA, AT-9010 and nsp12. **b** Close up of RdRp catalytic site following AT-9010 incorporation. **c** Close up of RdRp catalytic site following remdesivir incorporation (PDB 7BV2). Both structures in panel **b** and **c** are in the same orientation, superimposed by least squares fit method (shown in circle). One AT-9010 5′-monophosphate (AT-9010-MP) is incorporated into the primer RNA strand, and terminates RNA elongation. The second AT-9010 molecule, coordinated by one ion, occupies the NTP-binding site in comparison, remdesivir is terminally incorporated, and untranslocated. Superimposition shows the incoming (−1) AT-9010 ribose group is shifted with the phosphates in a post-incorporation position. Ribbons are depicted as follows: nsp12, blue; nsp7, pink; $nsp8_1$ and $nsp8_2$, yellow and cyan, respectively; RNA, green sticks; AT-9010, magenta sticks. **d** Ligplot 2D analysis of the contacts of AT-9010 molecule with nsp12, and incorporated AT-9010-MP.

relative to an unmodified or STP-terminated RNA 3′-end (Fig. 3b, c). Given the highly competitive incorporation of AT-9010, this reduced ExoN-rate likely slows viral replication, even if complete excision may eventually be achieved.

AT-9010 binds into the active site of the NiRAN domain. A third AT-9010 is found in the N-terminal NiRAN domain of nsp12 (Figs. 1b and 4 and Supplementary Movie 2). Initial attempts to place the AT-9010 triphosphate group lead to a steric clash between a non-bridging oxygen of the γ-phosphate and the carboxyl group of Asp218. However, the density suggests the presence of an additional metal ion, as seen in the SelO pseudo-kinase structure[13]. Density and coordination distances between this ion, the AT-9010-β phosphate and the Asp218 carboxyl group are coherent, showing AT-9010 is in its diphosphate (DP) form—herein referred to as AT-9010-DP (Fig. 4b and Supplementary Fig. 4a). In support of the DP form, HPLC analysis of AT-9010 before and after incubation with either nsp12 wild-type (WT), or NiRAN mutant, or RdRp active-site mutant shows that the NiRAN domain mediates hydrolysis of the γ-phosphate. The presence of the DP form increases 2.3–2.8-fold with WT and RdRp active-site mutants, but not following incubation with the NiRAN active-site mutant (Supplementary Fig. 4b).

The NiRAN catalytic center contains a groove harboring two catalytic ions, coordinated by conserved residues Asn209 and Asp218 (Fig. 4c, e), leading to a narrow cavity. The α and β phosphates of AT-9010 are stabilized in the groove by the two ions and residues Lys50, Lys73 and Arg 116 (Fig. 4a). Intriguingly, the orientation of AT-9010-DP is strikingly different to that of ADP and GDP in existing NiRAN structures[17,32], as well as other nucleotide-bound pseudo-kinase structures (Fig. S4a–c)[20,21]. Rather, the binding mode is reminiscent of the orientation of ATP bound to casein kinase[33]. The AT-9010-DP guanine base and modified ribose are accommodated in an inner narrow cavity (Fig. 4), which is unoccupied in existing NDP-NiRAN structures[17,32]. The guanine base is extensively stabilized by hydrophobic interactions and through hydrogen bonding with Arg55, Thr120 and Tyr217, residues which are conserved in CoV nsp12 sequences (Fig. 4a, b), but absent in other pseudo-kinases (Supplementary Fig. 5). The AT-9010-DP α- and β-phosphates are coordinated in the same groove as the phosphates of GDP and ADP in other NiRAN structures, but are in a flipped orientation, occupying the usual positions of the β- and α-phosphates, respectively[17,32]. AT-9010-DP has thus a unique binding mode, driven by both the hydrophobic nature of the cavity and modified ribose.

AT-9010 inhibits nsp9 and nsp8 UMPylation. The NiRAN has previously been shown to mediate the covalent transfer of NMPs[13] to various cofactor proteins[14–16]. In these studies, UTP

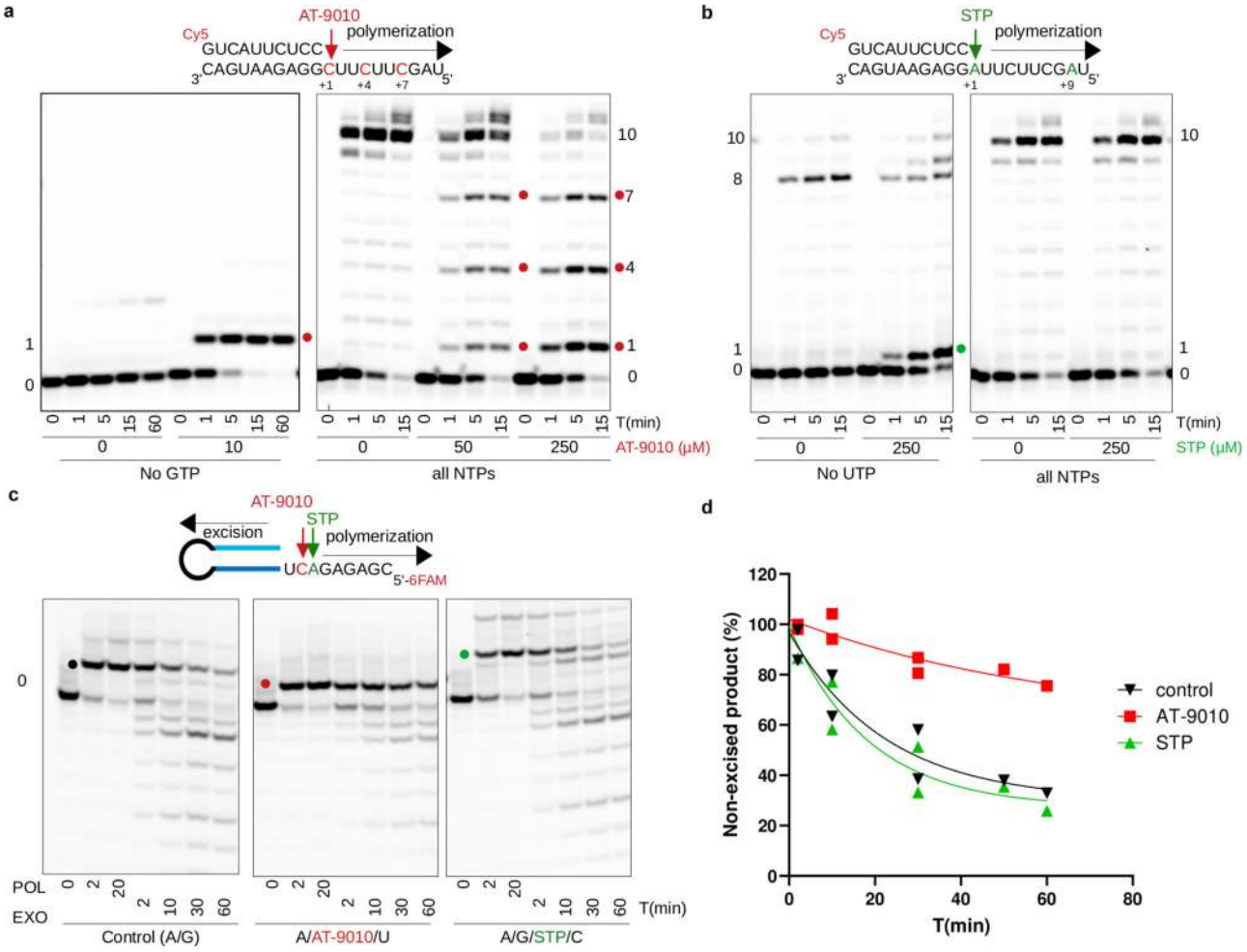


Fig. 3 Incorporation of AT-9010 and Sofosbuvir triphosphate (STP) by the SARS-CoV-2 RTC, and excision by the SARS-CoV-2 nsp14/10 exonuclease complex (ExoN). a Timecourse of AT-9010 incorporation (red dots) as a substitute for GTP (left panel, 50 μM each of ATP, UTP, and CTP), or in competition with GTP (right panel, 50 μM each NTP) at indicated concentrations of AT-9010 (0–250 μM). Fold-preference for GTP over AT-9010 incorporation is calculated by comparing the amount of AT-9010 insertion (red dots) relative to full-length product at two concentrations and at three time-points. **b** Timecourse of Sofosbuvir triphosphate (STP) incorporation (green dots) as a substitute for UTP (left panel, 50 μM each of ATP, GTP and CTP), or in competition with UTP (50 μM each NTP) at indicated concentrations (0 or 250 μM). For both **a** and **b**, reactions were run at least in duplicate at multiple nucleotide/analog concentrations for two different RNA substrates, with consistent results. **c** Incorporation of ATP+GTP control (black, left panel), ATP + AT-9010 (red, middle panel) and ATP + GTP + STP (green, right panel) by the polymerase complex (POL), followed by excision time-course with the nsp14 ExoN (EXO). For both AT-9010 and STP experiments, the next templated nucleotide for incorporation was additionally added. **d** Quantitation of remaining product after ExoN excision shown in **c**, a representative gel of experiments done in duplicate for two separate RNAs. The RNA marked at position 0 on each gel corresponds to the size of the fluorescently labeled primer shown above each gel, prior to elongation. Source data are provided as a Source Data file.

was demonstrated to be the preferred substrate for cofactor labeling. However, our structural analysis shows that the AT-9010 guanosine base is well accommodated into the NiRAN. To determine whether AT-9010 is able to inhibit NiRAN-mediated transferase activity, we performed competition experiments measuring the efficiency of nsp9-UMPylation by both SARS-CoV and SARS-CoV-2 RTCs in the presence of increasing concentrations of AT-9010, or its uracil equivalent STP (Supplementary Fig. 6a). Both drugs inhibit nsp9 labeling at comparable levels for SARS-CoV and SARS-CoV-2 RTCs. When provided at equimolar concentrations to UTP, nsp9-UMPylation is inhibited ~85–90% by both STP and AT-9010 for the SARS-CoV-2 complex, showing both drugs outcompete UTP for NiRAN-binding. Nsp8 has additionally been shown to act as a target substrate for NiRAN-mediated UMPylation by the CoV RTC[14,16]. Interestingly, AT-9010 is ~4-5-fold more efficient at blocking nsp8 labeling than the uracil equivalent STP (Supplementary Fig. 6b). Furthermore, given the excess of nsp8 provided in the reaction, these inhibition data suggest that AT-9010 remains stably bound into the NiRAN active site, rather than being transferred to nsp8.

AT-9010 is stably bound in the NiRAN active site. Thermal shift assays with nsp12 confirms that AT-9010 provides more thermodynamic stability than any other native nucleotide (Supplementary Fig. 6c). Comparison of NiRAN and RdRp active-site mutants (K73A and SAA, respectively) shows that this stability increase is provided by AT-9010 binding preferentially to the NiRAN, rather than the RdRp active site. Both GTP: and AT-9010:nsp12 complexes show an increased thermal stability compared with UTP or STP-bound complexes. Overall, and consistent with structural analysis, these results indicate that guanosine is a stable ligand of the NiRAN active site, with the 2′-fluoro-2′-C-methyl ribose modification of AT-9010 likely allowing insertion into the pocket and/or providing additional stability.

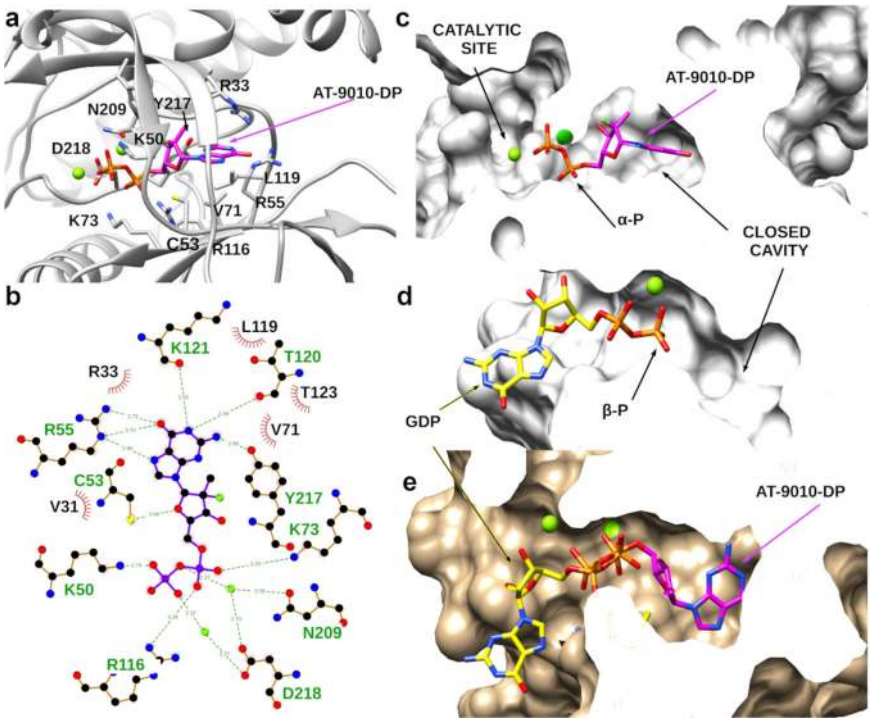


Fig. 4 Structural basis for NiRAN inhibition of the SARS-CoV-2 by AT-9010. a Binding of AT-9010 5′-diphosphate (AT-9010-DP) in the NiRAN domain. Interacting residues are labeled and shown in stick representation. **b** Ligplot 2D representation of the detailed interactions of AT-9010-DP binding in the NiRAN cavity. **c** Surface representation of the NiRAN sliced orthogonally in Z showing AT-9010-DP engulfed in the cavity. **d** Surface representation of the NiRAN sliced orthogonally in Z showing GDP binding (PDB 7CYQ); Orientation is rotated by ~+90° compared to **c**. **e** Superimposition of GDP and AT-9010-DP showing the overlap of the α-phosphate of AT-9010 with the β-phosphate of GDP and Mg^{2+} in the two structures, same orientation as **d**. Color code is as follows: AT-9010, magenta sticks; GDP, yellow sticks; and Mg^{2+}, light green.

In contrast, the binding of UTP is weak and putatively transient, likely to facilitate transfer to CoV nsps and possibly other targets.

Discussion

AT-9010 is efficiently incorporated into viral RNA, causes immediate chain-termination, and shows resistance to excision. Additionally, the superior affinity of guanosine analogs for the NiRAN as well as its deep AT-9010-binding pocket uncovers a specific druggable site of remarkable interest for anti-coronavirus research. The respective inhibition impact of AT-9010 on each of these activities is unknown. We note that the precise role of the NiRAN in the viral life-cycle is still somewhat speculative, but mutagenesis has shown that its NMPylation activity is essential[12]. Nucleotide transferase enzymes are required for a range of reactions including, e.g., RNA capping, DNA/RNA ligation, and priming of RNA synthesis, making it possible that the NiRAN is involved in various viral processes. Currently, there is some evidence for both guanylyltransferase and RNA-priming activity for the NiRAN domain of SARS-CoV-2[16,17]. Here, we show through both structural and biochemical analysis that AT-9010 binds preferentially, and stably in the NiRAN active-site pocket. Both of the reported NMPylation events (nsp8 and nsp9) are blocked efficiently by AT-9010. Therefore, any downstream activities dependent on this nucleotidylation event, or dependent on NiRAN-NTP binding in general (including capping, DNA/RNA ligation, or priming of RNA synthesis) would be blocked.

We note that the serendipitous simultaneous targeting of RdRp and NiRAN essential activities by the same drug adds momentum to the use of inhibitors against COVID-19 and the SARS-CoV-2 RTC, which remarkably has remained genetically stable since the beginning of the pandemic. Such AT-9010 pleiotropic action should attenuate the chance of simultaneous resistance mutations, and be an important asset in the control of the expanding genetic pool of SARS-CoV-2 variants observed in the current pandemic.

Methods

5′-triphosphate nucleosides. AT-9010, and 5′-triphosphate 2′-fluoro-2′-C-methyl uridine (Sofosbuvir 5′-TP, STP) were provided by NuBlocks LLC, Oceanside, CA, USA. Other NTPs were HPLC grade, purchased from GE Healthcare.

Expression and purification of SARS-CoV proteins. The SARS-CoV and SARS-CoV-2 sequences used in this study are from the Tor2 (NCBI ref. seq.: NC_004718.3) and Wuhan-Hu-1 (NCBI ref. seq.: YP_009724389.1) isolates, respectively. Unless specified ortherwise, a N-terminus 6His tag was fused to a TEV cleavage site ENLYFQG on which the last glycine was replaced by the first amino acid of the coding sequence of the gene of interest.

SARS-CoV proteins 6His(TEV)nsp7 and 6His(TEV)nsp8 were expressed under the control of a T5-promoter in pQE30 vectors in Escherichia coli (*E. coli*) NEB Express C2523 cells (New England Biolabs) carrying the pRare2LacI (Novagen) plasmid. Protein was expressed overnight at 17 °C (with 100 µg/mL ampicillin and 17 µg/mL Chloramphenicol), following induction with 100 µM IPTG at an $OD_{600} = 0.5–0.6$. Cells were lysed by sonication in lysis buffer (50 mM Tris-HCl pH 8, 300 mM NaCl, 10 mM imidazole, supplemented with 20 mM $MgSO_4$, 0.25 mg/mL Lysozyme, 10 µg/mL DNase and 1 mM PMSF) and protein was purified through affinity chromatography with TALON® Superflow™ cobalt-based IMAC resin (Cytiva). A wash step with buffer supplemented with 500 mM NaCl was performed prior to elution with 200 mM imidazole. The affinity tag was removed via overnight cleavage with TEV protease (1:10 w/w ratio to TEV:protein) in a dialysis buffer containing no imidazole and supplemented with 1 mM DTT. Cleaved protein was re-purified through a second cobalt column to remove the histidine-labeled TEV protease, and further purified with size exclusion chromatography (Cytiva Superdex S200) in a final buffer of 25 mM HEPES pH 8, 150 mM NaCl, 5 mM $MgCl_2$ and 5 mM TCEP.

SARS-CoV nsp12-8His was expressed from a pJ404 vector in *E. coli* strain BL21/pG-Tf2 (Takara 9124), in the presence of ampicillin (100 µM/mL) and Chloramphenicol (17 µg/mL). Expression was induced at $OD_{600} = 0.5–0.6$ with 250 µM IPTG and 5 ng/mL of tetracycline for induction of chaperones proteins (groES-groEL-tig), and left overnight at 23 °C at 220 rpm. Cells were lysed over 45–60 mins at 4 °C, in a buffer containing 50 mM Tris pH 8, 300 mM NaCl, 5 mM

$MgSO_4$, 10% glycerol, 1% CHAPS, supplemented with 5 mM 2-mercaptoethanol, 0.5 mg/mL Lysozyme, 10 μg/mL DNase, 1 mM PMSF, 0.2 mM benzamidine. NaCl was gradually added to a final concentration of 1 M, to aid in the removal of contaminating nucleic acids. Following centrifugation (30,000 × *g* for 30 min), the supernatant was diluted to reduce NaCl concentration to 300 mM. Protein was purified using cobalt-based IMAC resin TALON® Superflow™ (Cytiva), washing three times with wash buffer (50 mM Tris pH 8, 10% glycerol) with alternating NaCl concentration (300 mM, 1 M, 300 mM) before elution with 200 mM imidazole. Protein was further purified through size exclusion chromatography (Cytiva Superdex S200) in the same final buffer as nsp7 and nsp8 proteins, supplemented with 10% glycerol. Concentrated aliquots of nsp12, 7, and 8 were flash-frozen in liquid nitrogen and stored at −80 °C.

SARS-CoV nsp9 was expressed with self-cleavable ubiquitin fused at its N-terminus and 6His at its C-terminus under the control of a Tet-promoter in a pASK vector in *E. coli* NEB Express C2523 cells carrying the pCG1 plasmid (New England Biolabs). Protein was expressed overnight at 20 °C (with 50 μM/mL kanamycin, 17 μg/mL chloramphenicol), following induction with 200 μg/L anhydrotetracycline at an $OD_{600} = 0.6–0.7$. Cells were incubated in lysis buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 10 mM midazole, 5 mM $MgSO_4$, 1 mM of β-mercaptoethanol, 0.25 mg/mL Lysozyme, 10 μg/mL DNase, 0.1% triton and 1 mM PMSF) and lysed by sonication. Protein was purified first through affinity chromatography with HisPur Cobalt resin (Thermo Scientific), eluted in 100 mM imidazole, then through size exclusion chomatography (GE Superdex S200) in a final buffer of 50 mM HEPES pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$ and 1 mM of β-mercaptoethanol.

SARS-CoV protein nsp10 carrying a non-cleavable 6His tag in N-terminus was expressed under the control of a Tet-promoter in a pASK vector in *E. coli* NEB Express C2523 cells (New England Biolabs) carrying the pRare2LacI (Novagen) plasmid. Protein was expressed overnight at 17 °C (with 50 μM/mL kanamycin, 17 μg/mL chloramphenicol), following induction with 200 μg/L tetracycline at an OD600 = 0.6–0.7. Cells were incubated in lysis buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 10 mM imidazole, 5 mM $MgSO_4$, 1 mM of β-mercaptoethanol, 0.25 mg/mL Lysozyme, 10 μg/mL DNase, 0.1% Triton and 1 mM PMSF) and lysed by sonication. Protein was purified first through affinity chromatography with HisPur Cobalt resin (Thermo Scientific), eluted in 100 mM imidazole, then through size exclusion chomatography (GE Superdex S200) in a final buffer of 50 mM HEPES pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$ and 1 mM of β-mercaptoethanol.

SARS-CoV nsp14 carrying a non-cleavable 6His tag in N-terminus was expressed from a pDEST14 vector in *E. coli* strain NEB Express C2566 cells (New England Biolabs) carrying the pRare2 plasmid, in the presence of ampicillin (100 μM/mL) and chloramphenicol (17 μg/mL). Protein expression was induced at an $OD_{600} = 0.8$ with 2 μM IPTG, and left overnight at 17 °C with shaking. Cells were lysed by sonication in a buffer containing 50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole, supplemented with 0.25 mg/mL Lysozyme, 10 μg/mL DNase and 1 mM PMSF. The protein was purified through affinity chromatography with HisPur Cobalt resin (Thermo Scientific), washing with an increased concentration of salt (1 M NaCl), prior to elution in buffer supplemented with 250 mM imidazole. The protein was further purified by a size exclusion chomatography (GE Superdex S200) in a final buffer of 10 mM HEPES pH 7.5, 150 mM NaCl.

SARS-CoV-2 proteins were either purchased from Biortus (en.wuxibiortus.com) or purified using the following protocols.

For Cryo-EM, the gene of the full-length SARS-CoV-2 nsp12 (residues 1-932) was synthesized with codon optimization (General Biosystems) and cloned into pFastBac1 baculovirus expression vector. An additional peptide (MHHHHHHHHHWSHPQFEKENLYFQG) was added to the N-Terminus of nsp12. *Spodoptera frugiperda* (*Sf21*) cells expressing the target protein were collected 48 h after infection at 27 °C and were centrifuged at 4300 × *g* for 10 min. Pellets were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 2 mM $MgCl_2$, Complete Protease Inhibitor Tablet) and homogenized with High-Pressure Homogenizer at 4 °C. Cell lysate was centrifuged at 27,000 × *g* for 60 min at 4 °C. The fusion protein was first purified by Strep-Tactin (Strep-actin®XT) affinity chromatography and the tag was removed by incubation of TEV protease overnight at 4 °C after elution. The protein was reloaded onto a Heparin HP column after buffer exchanged by ultrafiltration tubes to buffer A (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5% glycerol, 2 mM $MgCl_2$). Flow through was collected and loaded on to a HiLoad 16/600 Superdex 200 pg column (GE healthcare) equilibrated in 10 mM Tris-HCl, pH 8.0, 500 mM NaCl, 2 mM $MgCl_2$. Purified nsp12 was concentrated to 6.86 mg/mL with ultrafiltration tubes and stored at −80 °C.

The gene of SARS-CoV-2 nsp7 (residues 1-83) possessing a C-terminal Avi-6His tag (GLNDIFEAQKIEWHEHHHHHH) was cloned into a modified pET-32a vector. BL21 (*E. coli*, T7 Express) containing the plasmid were grown to an OD600 of 0.6 at 37 °C, and protein was expressed at 15 °C for 16 h after the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. The cells were harvested then resuspended in buffer B (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 10 mM imidazole). Cells were disrupted by a High-Pressure Homogenizer at 4 °C. The insoluble material was removed by centrifugation at 27,000 × *g*, 60 min at 4 °C. The fusion protein was purified by Ni-NTA (Novagen, USA) affinity chromatography followed by a Superdex HiLoad 16/600 Superdex 75 pg column (GE Healthcare, USA) in buffer C (10 mM Tris-HCl, pH 8.0, 150 mM NaCl). Purified nsp7 was concentrated to 7.27 mg/mL with ultrafiltration tubes and stored at −80 °C.

The gene of SARS-CoV-2 nsp8 (residues 1–198) was cloned into a modified pET-28a vector containing an N-terminal His6-flag-tag with a TEV cleavage site (MHHHHHHDYK DDDDKENLYFQG) for expression in *E. coli*. Nsp8 was expressed in the same way as that for nsp7. Cells were harvested then resuspended in buffer D (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol). Cells were lysed using a High-Pressure Homogenizer at 4 °C. Cell lysate was clarified by centrifugation at 27,000 × *g*, 60 min at 4 °C. Supernatant was applied onto a Talon affinity chromatography column and the tag was removed by on-column cleavage overnight using TEV protease. The mixture was buffer exchanged to buffer D with ultrafiltration tubes and reloaded onto a His-chelating NTA column again to remove the His tag and TEV protease. Target protein was further purified by passage through a HiLoad 16/600 Superdex 75 pg (GE Healthcare, USA) in buffer E (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol). The fractions near the maximum height of the peak were combined and further purified by a Mono Q 10/100 GL column (GE Healthcare, USA). Purified nsp8 was buffer exchanged to buffer E with ultrafiltration tubes, then concentrated to 11.63 mg/mL and stored at −80 °C.

For enzyme studies, SARS-CoV-2 proteins 6His(TEV)nsp7, 6His(TEV)nsp8, and nsp9-6His were obtained from *E.coli* expression clones. The SARS-CoV-2 sequences replaced those of SARS-CoV in cognate vectors. For SARS-CoV-2 nsp10 and nsp14, genes were cloned into the pET28a(+) vector with a 6His tag fused to the TEV cleavage sequence as described above; The proteins were purified following the same protocol as SARS-CoV 6His(TEV)nsp7 and 6His(TEV)nsp8, omitting hloramphenicol in the culture medium. Unlike nsp7 and nsp8, which always had their tags removed, nsp12 was used with or without 6His tag without noticeable different properties. Likewise, no noticeable differences were noted between SARS-CoV and SARS-CoV2 for their RTC or ExoN activities.

Assembly of the extended nsp12-nsp7-nsp8-RNA complex for Cryo-EM. To assemble the Extended RdRp complex, nsp12 was incubated with nsp7 and nsp8 at 4 °C for 3 h with a molar ratio of 1:3:6 in a buffer containing 25 mM Tris-HCl, pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 4 mM DTT. Then the mixture was purified by Mono Q 5/50 GL ion-exchange chromatography (GE Healthcare, USA), resulting in a stable nsp7-nsp8-nsp12 complex. The protein complex was desalted to buffer F (25 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 4 mM DTT). Purified RdRp complex was buffer exchanged to 25 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5 mM $MgCl_2$, 4 mM DTT and concentrated to 10 mg/mL for Cryo-EM experiments. A 30-mer oligoribonucleotide template (5′-CCCCCCCCCCAUAACUUA AUCUCACAUAGC-3′) and a 20-mer oligoribonucleotide primer (5′-GCUAU GUGAGAUUAAGUUAU-3′) were chemically synthesized by GenScript. They were annealed by heating an equimolar solution in RdRp buffer to 95 °C and gradually cooling to 4 °C. The annealed RNA scaffold was incubated with the nsp12-nsp7-nsp8 complex for 30 min at 4 °C with a molar ratio of 2:1 to form the nsp12-nsp7-nsp8-RNA complex. AT-9010 with a final concentration of 1 mM was added subsequently for compound incorporation.

Cryo-EM sample preparation and data collection. In total, 3 μL of protein solution at 5 mg/mL (with 0.025% DDM) was applied onto a glow-discharged holey carbon grid (Quantifoil, 300 mesh, R1.2/1.3). Excess samples were blotted for 5.0 s with a blotting force of 3, then the remaining solution was vitrified by plunging into liquid ethane using a Vitrobot Mark IV (Thermo Fischer Scientific) at 4 °C and 100% humidity. Cryo-EM data were collected with a 300 keV Titan Krios electron microscope (Thermo Fisher Scientific, USA) equipped with a K3 direct electron detector (Gatan, USA) operating in a super-resolution counting mode. All movies were automatically recorded using SerialEM[34] at a magnification of 105 K, with a physical pixel size of 0.83 Å. A total dose of 80.5 e-/Å^2 was fractionated into 50 frames. 7,459 movie micrographs were collected with a defocus range from −1.5 to −2.5 μm, and the slit width of Gatan Quantum GIF energy filter (Gatan, USA) was set to be 20 eV. Statistics for data collection and refinement are shown in Supplementary Table 1.

Cryo-EM image processing. All dose-fractioned movies were motion-corrected with Relion′s own implementation. CTF estimation, 2D classification, 3D classification and refinements were all performed in cryoSPARC. A total of 2,410,466 particles were auto-picked using blob picker and extracted with a box size of 320 pixels. 248,401 particles were selected after three rounds of 2D classification based on the complex integrity. This particle set was used for Ab-Initio reconstruction with three classes, which were then used as 3D volume templates for heterogeneous refinement. The 3D volume corresponding to the intact nsp12-nsp7-nsp8-RNA complex was used for creating 100 2D projections, which were then used as templates for template-based particle picking. Approximately 3,640,595 particles were picked from a set of 3622 micrographs filtered based on fitted resolution better than 5 Å as estimated by CTFFIND4, using template picker. Particles were extracted with a box size of 360 pixels. A total of 234,421 particles were selected after four rounds of 2D classification based on the complex integrity. This particle set was used for Ab-Initio reconstruction with three classes, followed by

heterogeneous refinement. A subset of 181,669 particles from the class with good features was subjected to Homogeneous Refinement, Local Refinement and Non-uniform Refinement, resulting in a 2.98 Å map.

Model building and refinement. To build the model of nsp12-nsp7-nsp8-RNA complex, the structure of SARS-CoV-2 nsp12-nsp7-nsp8-RNA complex (from PDB 7CYQ), with one nsp9 and nsp13 removed) was placed and rigid-body fitted into the Cryo-EM map using UCSF Chimera. The model was manually built in Coot[35] with the guidance of the Cryo-EM map, and in combination with real space refinement using Phenix[36]. The model validation statistics are shown in Supplementary Table 1.

HPLC analysis of nsp12-AT-9010 products. Wild-type nsp12 of SARS-CoV and SARS-CoV-2, as well as SARS-CoV NiRAN (K73A) and RdRp active-site mutants (SDD → SAA) were incubated with equimolar concentrations of AT-9010 triphosphate for one hour at 37 °C in a reaction buffer containing 50 mM HEPES, pH 7.5, 1 mM DTT and 1 mM MnCl. Reactions were heat inactivated at 70 °C for 10 min to promote dissociation of bound AT-9010 and products were purified through micocon spin columns (molecular weight cutoff 10,000). In parallel, a mock reaction without enzyme was performed under analogous conditions. Products were separated on HPLC C18 columns under a water: acetonitrile gradient to separate tri-, di- and mono-phosphorylated versions of AT-9010, UV detected and quantitated by determining peak surface areas.

Primer-dependent polymerization and excision assays. Primer (5′cy5-GUCAUUCUCC-3′) and template (5′-UAGCUUCUU(A/C)GGAGAAUGAC-3′) RNA pairs, corresponding to the 3′ end of the SARS-CoV genome, were annealed at a molar ratio of 1:1.5 in 110 mM KCl at 70 °C for 10 min, then cooled slowly to room temperature over several hours. Hairpin RNAs were synthesized by Integrated DNA Technologies (Coralville, IA). The active RTC was formed by first incubating nsp7 and 8 together at equimolar concentrations (100 µM) for 30 min at room temperature. Nsp12, extra nsp8 and protein gel filtration buffer were added to form a final complex consisting of nsp12:7:8 at a 1:3:6 ratio, with 10 µM nsp12. The complex was further incubated for 10 mins at room temperature, then pre-incubated with RNA in a pre-mix containing 20 mM HEPES pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$. For single nucleotide incorporation assays, reactions were initiated with 50 µM (final concentration) of all AT-9010 or STP, with or without the following nucleotide (ATP). Final reaction concentrations were 0.5 µM nsp12, 0.4 µM RNA. Reactions were quenched after indicated time-points with 5X volume of FBD stop solution (formamide, 10 mM EDTA). For AT-9010–GTP competition experiments, protein-RNA complexes were preincubated as described above, and initiated with either all four NTPs, or with only CTP, UTP and ATP (50 µM each NTP), supplemented with various concentrations of AT-9010 (10–250 µM). To calculate the discrimination between AT-9010 and GTP, the AT-9010 product band (from 50 or 250 µM concentrations) was compared with the sum of fractions of product bands derived from GTP incorporation at three time-points. Discrimination was corrected to account for concentration difference between AT-9010 and GTP. For analog excision, polymerization reactions were performed on hairpin RNAs (5′-6-FAM-GCUUACGAGAAUGACAAAAGUCAUUCU-3′ and 5′-6FAM-GAAAGGUCUCGCGCCCGGGUAAUCGGGCGCGAGA-3′) in the same conditions as described above (labeled RTC, 2′ and 20′), then stopped by heating at 70 °C for 10 min. The hairpin was re-annealed at 30° for >30 min, and 50 nM nsp14/nsp10 (1:5) were added for time-course reactions (labeled Exo, 2′, 10′, and 60‘). Reaction products were separated using denaturing polyacrylamide gel electrophoresis (20% acrylamide, 7 M urea), visualized using a Typhoon FluorImager, and analyzed with ImageQuant software. Percent excision was calculated by dividing the intensity/volume of the band of interest (unexcised product) by the sum of all bands in the lane, to account for loading error. Excision percentage was plotted over time using GraphPad Prism, and fitted using a one-phase decay to obtain the reaction rate. The reaction was run in triplicate and in duplicate using two different hairpin RNAs.

Nucleotidyl transferase inhibition reactions. Standard nucleotide-transfer reactions were performed at 37 °C in a buffer containing 20 mM HEPES pH 7.5, 1 mM DTT, 2 mM MnCl2, 1 µCi α^{32}P-UTP, 30 mM NaCl and a 5 µM constant concentration of cold UTP, in competition with increasing concentrations of AT-9010 and STP (0.5–300 µM). Final protein concentrations of 1 µM nsp12 with 2 µM nsp9 or 5 µM nsp8 were used. Reactions were stopped after 15 min incubation, and analyzed on 15% sodium dodecyl sulphate–polyacrylamide gel electrophoresis gels, stained with InstantBlue for total protein and exposed for 2 h—overnight to reveal radiolabelled proteins. Nsp9 and Nsp8-UMPylation was calculated as a percentage relative to the no-inhibitor controls. Reactions were performed in triplicate with 6–7 concentrations of each NA.

Thermal-shift assays. The influence of compound binding on protein stability was measured by Themofluor assay, using a CFX Connect BioRad real-time PCR machine. Reactions were run in thin-walled, 96-well plates, in a buffer composed of 10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$ and 0.5 mM $MnCl_2$, in a final reaction volume of 20 µL. Compounds and NTP concentrations were fixed at 100 µM, with a final protein concentration of 2 µM (nsp12 WT and mutants) SYPRO Orange dye was used at a five-fold final concentration. Melting-temperature (T_m) values given are the average and standard deviation of three independent experiments.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The coordinates and structure factors for SARS-CoV-2 nsp7:nsp8:nsp12:RNA:AT-9010 quaternary structure have been deposited in the PDB with the following accession codes: 7ED5 and the EMDB: EMD-31061, respectively. Source data are provided with this paper.

Received: 27 November 2021; Accepted: 9 January 2022;
Published online: 02 February 2022

References

1. Tao, Y., Farsetta, D. L., Nibert, M. L. & Harrison, S. C. RNA synthesis in a cage–structural studies of reovirus polymerase lambda3. *Cell* **111**, 733–745 (2002).
2. Peersen, O. B. A comprehensive superposition of viral polymerase structures. *Viruses* **11**, https://doi.org/10.3390/v11080745 (2019).
3. Minskaia, E. et al. Discovery of an RNA virus 3′->5′ exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl Acad. Sci. USA* **103**, 5108–5113 (2006).
4. Eckerle, L. D., Lu, X., Sperry, S. M., Choi, L. & Denison, M. R. High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J. Virol.* **81**, 12135–12144 (2007).
5. Bouvet, M. et al. RNA 3′-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl Acad. Sci. USA* **109**, 9372–9377 (2012).
6. Ferron, F. et al. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl Acad. Sci. USA* **115**, E162–E171 (2018).
7. Good, S. S. et al. AT-527, a double prodrug of a guanosine nucleotide analog, is a potent inhibitor of SARS-CoV-2 in vitro and a promising oral antiviral for treatment of COVID-19. *Antimicrob. Agents Chemother.* https://doi.org/10.1128/AAC.02479-20 (2021).
8. Good, S. S., Moussa, A., Zhou, X.-J., Pietropaolo, K. & Sommadossi, J.-P. Preclinical evaluation of AT-527, a novel guanosine nucleotide prodrug with potent, pan-genotypic activity against hepatitis C virus. *PLoS ONE* **15**, e0227104 (2020).
9. Han, Y. J. et al. Treatment of severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and coronavirus disease 2019 (COVID-19): a systematic review of in vitro, in vivo, and clinical trials. *Theranostics* **11**, 1207–1231 (2021).
10. Dousson, C. B. Current and future use of nucleo(s)tide prodrugs in the treatment of hepatitis C virus infection. *Antivir. Chem. Chemother.* **26**, 2040206618756430 (2018).
11. Hartenian, E. et al. The molecular virology of coronaviruses. *J. Biol. Chem.* **295**, 12910–12934 (2020).
12. Lehmann, K. C. et al. Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. *Nucleic Acids Res.* **43**, 8416–8434 (2015).
13. Sreelatha, A. et al. Protein AMPylation by an evolutionarily conserved pseudokinase. *Cell* **175**, 809–821.e19 (2018).
14. Conti, B. J., Leicht, A. S., Kirchdoerfer, R. N. & Sussman, M. R. Mass spectrometric based detection of protein nucleotidylation in the RNA polymerase of SARS-CoV-2. *Commun. Chem.* **4**, 41 (2021).
15. Slanina, H. et al. Coronavirus replication-transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit. *Proc Natl Acad Sci USA* **118**, https://doi.org/10.1073/pnas.2022310118 (2021).
16. Shannon, A. et al. Protein-primed RNA synthesis in SARS-CoVs and structural basis for inhibition by AT-527. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.23.436564 (2021).
17. Yan, L. et al. Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell* **184**, 184–193.e10 (2021).
18. Subissi, L. et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl Acad. Sci. USA* **111**, E3900–E3909 (2014).
19. Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **10**, 2342 (2019).

20. Gao, Y. et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779–782 (2020).
21. Wang, Q. et al. Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell* **182**, 417–428.e13 (2020).
22. Yin, W. et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* **368**, 1499–1504 (2020).
23. Hillen, H. S. et al. Structure of replicating SARS-CoV-2 polymerase. *Nature* **584**, 154–156 (2020).
24. Malone, B. et al. Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *Proc Natl Acad Sci USA*. **118**, https://doi.org/10.1073/pnas.2102516118 (2021).
25. Peersen, O. B. Picornaviral polymerase structure, function, and fidelity modulation. *Virus Res.* **234**, 4–20 (2017).
26. Shannon, A. et al. B. Canard, rapid incorporation of Favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. *Nat. Commun.* **11**, 4682 (2020).
27. Gong, P. & Peersen, O. B. Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *Proc. Natl Acad. Sci. USA* **107**, 22505–22510 (2010).
28. Naydenova, K. et al. Structure of the SARS-CoV-2 RNA-dependent RNA polymerase in the presence of favipiravir-RTP. *Proc Natl Acad Sci USA*. **118**, https://doi.org/10.1073/pnas.2021946118 (2021).
29. Appleby, T. C. et al. Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* **347**, 771–775 (2015).
30. Kokic, G. et al. Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **12**, 279 (2021).
31. Traut, T. W. Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* **140**, 1–22 (1994).
32. Chen, J. et al. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* **182**, 1560–1573.e13 (2020).
33. Xu, R. M., Carmel, G., Sweet, R. M., Kuret, J. & Cheng, X. Crystal structure of casein kinase-1, a phosphate-directed protein kinase. *EMBO J.* **14**, 1015–1023 (1995).
34. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
36. Afonine, P. V. et al. Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D. Struct. Biol.* **74**, 531–544 (2018).

Acknowledgements

We thank Dr. Manfu Wang, and Dr. Zenglin Yuan from Wuxi Biortus Biosciences Co. Ltd. for their help in data processing and refinement. We thank Minqi Gao from Wuxi Biortus Biosciences Co. Ltd. for his help in enzyme activity assays, and Dr. Zheng Liu from the Centre for Electron Microscopy, CUHK, for his kind help in data collection. We thank Léa Lo Bello for technical assistance in protein purification. We are in debt to Pr. Olve Peersen for his constant support, data sharing, and valuable biochemical insight during the course of the project. We thank Stephen Harrison, Sylvie Doublié, Julien Lescar, Olve Peersen, and Juan Reguera for helpful suggestions and critical reading of the manuscript. This project has received funding from the Fondation pour la Recherche Médicale (Aide aux Équipes), SCORE project H2020 SC1-PHE-Coronavirus-2020, grant#101003627, ATEA Pharmaceuticals, the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No. 101005077 (IMI-CARE).

Author contributions

Conceptualization: A.S., B.C., Methodology: A.S., V.F., A.M., S.G., B. Sama, B. Selisko, A.D., M.F., C.F., P.G., P.E.K., C.E., Y.Z., X.Y., H.S., B.C., Analysis: A.S., E.D., K.A., K.A., J.C.G., S.G., P.L.C., K.L., Y.Z., X.Y., H.S., B.C., Visualization: A.S., F.F., B.C., Funding acquisition: E.D., F.F., J.J.V., J.P.S., B.C., Writing: A.S., F.F., B.C.

Competing interests

S.G., A.M., K.L., and J.P.S. are employees of ATEA Pharmaceuticals, Inc. The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-28113-1.

Correspondence and requests for materials should be addressed to Bruno Canard.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022

A2 : AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5

Version of Record: https://www.sciencedirect.com/science/article/pii/S0166354223000529
Manuscript_27af1d3134d44fa83517ed3342d71f60

AT-752 targets multiple sites and activities on the Dengue virus replication enzyme NS5

Mikael Feracci[1], Cécilia Eydoux[1], Véronique Fattorini[1], Lea Lo Bello[1], Pierre Gauffre[1], Barbara Selisko[1], Priscila Sutto-Ortiz[1], Ashleigh Shannon[1], Hongjie Xia[2], Pei-Yong Shi[2], Mathieu Noel[3], Françoise Debart[3], Jean-Jacques Vasseur[3], Steve Good[4], Kai Lin[4], Adel Moussa[4], Jean-Pierre Sommadossi[4], Aurélie Chazot[1], Karine Alvarez[1], Jean-Claude Guillemot[1], Etienne Decroly[1], François Ferron[1], & Bruno Canard[1]*

[1]AFMB, CNRS, Aix-Marseille University, UMR 7257, Case 925, 163 Avenue de Luminy, 13288 Marseille Cedex 09, France
[2] Department of Biochemistry and Molecular Biology, Sealy Institute for Drug Discovery, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX, USA. peshi@utmb.edu.
[3]IBMM, UMR 5247 CNRS-UM1-UM2, Department of Nucleic Acids, Montpellier University, Place E. Bataillon, 34095 Montpellier Cedex 05, France
[4]Atea Pharmaceuticals, Inc.; 225 Franklin St., Suite 2100, Boston, MA 02110 USA.

*Corresponding author: bruno.canard@univ-amu.fr

© 2023 published by Elsevier. This manuscript is made available under the Elsevier user license
https://www.elsevier.com/open-access/userlicense/1.0/

Abstract

AT-752 is a guanosine analogue prodrug active against dengue virus (DENV). In infected cells, it is metabolized into 2'-methyl-2'-fluoro guanosine 5'-triphosphate (AT-9010) which inhibits RNA synthesis in acting as a RNA chain terminator. Here we show that AT-9010 has several modes of action on DENV full-length NS5. AT-9010 does not inhibit the primer pppApG synthesis step significantly. However, AT-9010 targets two NS5-associated enzyme activities, the RNA 2'-O-MTase and the RNA-dependent RNA polymerase (RdRp) at its RNA elongation step. Crystal structure and RNA methyltransferase (MTase) activities of the DENV 2 MTase domain in complex with AT-9010 at 1.97 Å resolution shows the latter bound to the GTP/RNA-cap binding site, accounting for the observed inhibition of 2'-O but not N7-methylation activity. AT-9010 is discriminated ~10 to 14-fold against GTP at the NS5 active site of all four DENV1-4 NS5 RdRps, arguing for significant inhibition through viral RNA synthesis termination. In Huh-7 cells, DENV1-4 are equally sensitive to AT-281, the free base of AT-752 ($EC_{50} \approx 0.50$ µM), suggesting broad spectrum antiviral properties of AT-752 against flaviviruses.

Introduction

The *Flavivirus* genus includes a number of significant human pathogens, such as Yellow Fever (YFV), Zika (ZIKV), West-Nile (WNV), Tick-Borne Encephalitis (TBEV), Japanese Encephalitis (JEV), and Dengue (DENV) viruses (Pierson and Diamond, 2020). They productively infect the salivary glands of their arthropod vectors, which transmit the virus to humans through subsequent bites during blood-meals. Flaviviruses are enveloped viruses infecting a wide variety of cells through binding to diverse receptors at the host cell surface. Dengue disease outbreaks and associated conditions have devastating public health and economic consequences in afflicted countries. With ~20,000 deaths and more than 100 million cases annually, Dengue morbidity and burden are considerable. DENV designs a group of at least four antigenically distinct members classified as Dengue serotype 1-4, homologous about 70% to each other at the amino acid level in their most conserved NS5 gene (van den Elsen et al., 2021).

The emergence of Zika virus in 2016 and its spreading from the pacific zone to South America and later worldwide illustrates the pandemic potential of either known or yet-to-be discovered flaviviruses (Masmejan et al., 2020). The recent emergence in the USA of Powassan virus (POWV), a BSL-3 deadly tick-borne flavivirus (Doughty et al., 2017), indicates that this viral genus deserves to be surveilled closely as it contains several BSL-4 agents such as, *eg.*, Omsk hemorrhagic fever virus (OHFV), and Kyasanur Forest disease virus (KFDV).

Besides the successful YFV vaccine, JEV and TBE are amongst the few protective vaccines available to contain flavivirus outbreaks. Ideally, DENV vaccine design requires a yet unreached full-protection against all serotypes at once. DengVaxia™, a vaccine approved in several countries, is recommended to persons aged >9 years old, without previous exposure to DENV (www.who.int/publications/i/item/who-wer9335-457-476). More recently, QDENGA™ was approved in Europe and in some endemic countries for individuals 4-years of age and older by preventing 84% of hospitalized Dengue cases but immunity is fading after 4 years, with only 65% protection (Rivera et al., 2022).

The genome size of human pathogenic flaviviruses is in the vicinity of 11 kb. It is made of a single-stranded RNA molecule of positive polarity ((+)RNA), which is 5'-capped but not 3'-poly-adenylated. The absence of a polyA tail is consistent with the functionalities provided by the 5' and 3' end sequences: the latter RNA ends partially anneal and fold into a defined structure promoting cyclization of the genome and a binding platform for the RNA replication/transcription complex (Filomatori et al., 2011). The RTC comprises seven non-structural protein NS1, NS2a & b, NS3, NS4a & b, and NS5. It associates to membranes within vesile-like structures into a viral factory able to produce proviral RNA (van den Elsen et al., 2021). The latter and protein components assemble to form new virions.

NS5 is an RNA-dependent RNA polymerase (RdRp) of ~100 kDa, responsible for viral RNA synthesis, and the most structurally conserved protein across flavivirus members. Along with RdRp activity, it carries both N7-guanine and 2'-O-methyltransferase (MTase) activities involved in the synthesis of a canonical eukaryotic RNA-cap (Ray et al., 2006; van den Elsen et al., 2021). In the eukaryotic cell, RNA capping proceeds through three reactions occurring onto the neo-transcribed RNA (Ramanathan et al., 2016). The nascent 5'-triphosphate transcript is first converted to the 5'-diphosphate by an RNA triphosphatase. The flavivirus NS3 helicase possesses this activity, although this has never been demonstrated as being involved in the RNA capping process. The 5'-diphosphate RNA transcript is then capped by the addition of GMP

originating from GTP by a guanylyltransferase (GTase). The flavivirus NS5 is suspected to carry this activity, yet the active site, catalytic residues, and mechanism of the reaction remain elusive (El Sahili and Lescar, 2017). The capped RNA is then sequentially methylated both at the N7-guanine of the guanine cap and then the 2'-hydroxyl of the first transcribed nucleotide, a flavivirus-conserved adenosine at position 1 (Ray et al., 2006).

Although Dengue outbreaks have been plaguing the tropical/subtropical belt for decades, the COVID19 crisis brought increased awareness and progress in the potential control and management of pandemic viruses. In particular, prophylactic and therapeutic antiviral drugs have been deemed useful together with the requirement of both rapid diagnostics and auxiliary treatment as early as possible after infection.

Fortunately, such diagnostics exist for Dengue, in particular in the early detection of minute amounts of NS1 in the bloodstream of infected individuals. The amount of NS1 correlates with disease severity (reviewed in (Ghetia et al., 2022)): the severity of the disease has been shown to be positively co-related to the height of the post-infection viremia peak. Therefore, as for many other serious infectious diseases caused by an RNA virus, there is a renewed interest in the design of potent antiviral drugs. The recent discovery of potent pan-flavivirus inhibitors targeting NS4b confirms the renewed interest in drug discovery and design against this important virus family (Kaptein et al., 2021; Moquin et al., 2021; Wang et al., 2022), which would likely benefit from combination therapies, as well as with concomitant addressing of the related cytokine storm occurring when virus titers have significantly declined (Kuczera et al., 2018).

NS5 represents a target of choice for antiviral drugs, and in particular nucleoside/tide analogues (NAs). At least four decades of research and many successful clinical developments have contributed to a wealth of knowledge of these drugs. A phylogenetically close relative to DENV, Hepatitis C virus is eradicated from infected individuals with a success rate close to 100% using sofosbuvir-containing combinatorial regimen (Dousson, 2018). Sofosbuvir is an uracil phosphoramidate prodrug converted intracellularly to its 5'-triphosphate form. Its 2'-fluoro, 2'-methyl modified ribose does not impede selection at the HCV NS5b (the RdRp) active site. However, once incorporated and although it contains a 3'-OH, chain termination occurs. It competes with intracellular UTP, and is a non-obligate chain terminator NA promoting accurate chain termination in lieu of UTP.

AT-752, AT-281, and their diastereoisomer bemnifosbuvir (AT-527) are NAs bearing the same 2'-modified ribose and 5'-phosphoramidate motifs as sofosbuvir ~~(Fig. 1A)~~. AT-752 is active against DENV and yellow fever viruses (Good et al., 2021a; Lin et al., 2022), and bemnifosbuvir against both HCV and SARS-CoV2 (Good et al., 2021b, 2020; Shannon et al., 2022). Both drugs are currently under evaluation in clinical trials (https://clinicaltrials.gov/). Neither AT-752 nor bemnifosbuvir carries a natural nucleobase but a *N*-methyl diaminopurine base instead. Upon cell entry, both follow the same metabolism scheme as sofosbuvir up to recover an unsubstituted 5'-monophosphate (Murakami et al., 2014). At that point, though, the ADALP1 enzyme supposedly deaminates the N^6-methyl diaminopurine moiety into guanine, and the resulting 2'-fluoro, 2'-methyl ribose guanosine 5'-monophosphate is converted to the 5'-triphosphate (AT-9010). The corresponding 5'-triphosphate of AT-752, ie., AT-9010, is used as a substrate by the DENV2 NS5 (Good et al., 2021a).

In the present paper, we characterize the in vitro antiviral activity of AT-281, the free base form of the hemi-sulfate salt AT-752, against DENV1-4 and investigate the mechanism of action of the active triphosphate metabolite AT-9010 towards various DENV NS5 domains and full-length products of all four serotypes. We show that AT-752 has a broad spectrum antiviral

activity in cells infected with DENV1-4. In parallel, AT-9010 shows a dual mode of action on flavivirus NS5. AT-9010 binds to the RNA-cap binding site of NS5, inhibiting the RNA 2'-O-MTase activity. We report the crystal structure of AT-9010 in complex with the DENV MTase domain at 1.97 Å resolution, accounting for the observed MTase inhibition. Additionally, AT-9010 is barely incorporated during synthesis of the dinucleotide primer pppAG: although acting as a *bona fide* chain terminator, it does not significantly inhibit pppAG formation. However, AT-9010 competes more efficiently with GTP during RNA elongation than sofosbuvir-5'-triphosphate (SOF-TP) with UTP. We show that AT-9010 substrate and inhibitor properties are valid across all DENV1 - 4. Its guanine base is responsible for its bonified substrate properties relative to sofosbuvir, and thus suggests that AT-752 may achieve broad-spectrum clinical benefit in the control of *Flavivirus* infections.

Results

The nucleoside analogue AT-752 is a purine nucleotide analogue phosphoramidate prodrug which is metabolized intracellularly into AT-9010, a 5'-triphosphate guanosine analogue inhibitor (Good et al., 2020). The natural nucleotide GTP is a substrate used as a building block for many different reactions involving the viral RdRp NS5 along the flaviviral life cycle. Since AT-9010 is a GTP analogue, it could theoretically target both the viral helicase NS3 to fuel motion along RNA and NS5. However, NS3 is able to hydrolyse all NTP unspecifically (Benarroch et al., 2004), and thus, we surmise that it is unlikely that AT-9010 would selectively and efficiently inhibit this process. NS5 likely is the main target of this inhibitor.

GTP is used by NS5 as *i)* the second nucleotide of the strictly conserved pppApG 5'-end of both the (-) and the (+)RNA strands, which can be synthesized even in the absence of a template in the NS5 RdRp active site (Selisko et al., 2012), *ii)* a component of the RNA cap m7GpppApG, *iii)* one of the four nucleotides assembled into the genomic and anti-genomic strand during RNA synthesis of the flavivirus RNA genome.

Figure 1 depicts five putative reactions which could be impeded in the presence of a GTP analogue such as AT-9010, defining possible inhibitory interferences, and thus, a possible mode of action of AT-752/AT-9010 against DENV.

AT-752 is a broad-spectrum DENV inhibitor

We tested the in vitro antiviral activity of AT-281, the free base form of AT-752, using DENV 1–4 that were modified to contain a nano luciferase reporter (see Materials and Methods). Solutions of AT-281 rather than AT-752 were tested to avoid any effect on the pH of the assay that the salt might have had. AT-281 demonstrated potent activity against all four serotypes, with respective EC_{50} values of 0.57, 0.30, 0.75, and 0.57 µM for DENV1, -2, -3 and -4 reporter viruses, respectively, in Huh-7 cells (Fig. 2).

In order to understand the AT-752 molecular basis of action, we set-out to investigate if its active form AT-9010 could compete with GTP in NS5-mediated reactions involved in RNA synthesis of the Dengue RNA genome.

AT-9010 binds to both MTase and RdRp domains of Flavivirus NS5

In order to determine the AT-9010 inhibition effect on all four serotypes DENV1-4, we purified individually three active forms of the DENV NS5 protein, namely the full-length enzyme NS5, and its two functional domains MTase and RdRp. On these three enzymes, we tested their GTP-binding properties in comparison with that of AT-9010 and sinefungin, a known inhibitor of the MTase activity (Lim et al., 2015). Thermal denaturation temperatures were monitored with a thermal shift assay, which records the increase of thermal stability of a given protein/enzyme when bound to its ligand (substrate or inhibitor).

Figure 3 and Table 1 show that for all DENV serotypes, all MTase, RdRp, and NS5 enzymes bind GTP and AT-9010 ($\Delta Tm \geq +2°C$). Sinefungin accordingly does not bind to the RdRp domain (Table 1, panel 2). GTP appears to promote an average 5°C increase in melting temperature to both NS5 and MTase, similar to that of AT-9010. We note that AT-9010 provides to the RdRp domain a greater thermal stability than GTP for all DENV serotypes (≥1°C). Interestingly, there is a functional GTP/AT-9010 RdRp binding site even in the absence of an

RNA primer:template. This finding is consistent with previous work reporting synthesis of the pppApG primer in the presence of Mn^{2+} but absence of an RNA template (Selisko et al., 2012).

In order to determine the mechanism of action of AT-9010 and see on which target its activity predominates, we assayed the different NS5-associated enzymatic activities successively, and evaluated their propensity to be inhibited by AT-9010. We first addressed the MTase domain.

AT-9010 binds to the NS5 MTase domain and inhibits RNA 2'-O-methylation

Inhibition was detected in a 2'-O-MTase assay using the MTase domain of DENV2 NS5 and $^{m7}GpppAC_4$ as a substrate. After mixing DENV2 NS5 MTase with increasing concentrations of compound, MTase activity was measured by a filter-binding assay (FBA). We subsequently performed a dose-response assay to evaluate the inhibition profile of AT-9010. We included the compound AT-9002, which carries a hydroxyl group instead of the fluorine atom at its 2'-position (Fig. 1). The IC_{50} of the compounds, deduced from the Hill slope equation (Y=100/(1+((X/IC_{50})^Hillslope) curve-fitting is 4.1 µM $\pm$ 0.4 for AT-9002 and 29.6 µM $\pm$ 2 for AT-9010 (Fig. 4A). Beside confirming the inhibitory activity of the latter, this indicates that the 2'-OH group probably makes a favorable interaction relevant to the inhibition mechanism of the 2'O-MTase activity.

To understand the molecular basis of the inhibition observed in DENV MTase assays, we performed crystallization studies using the isolated, enzymatically active flavivirus MTase domain. Crystals of DENV3 MTase were grown (Barral et al., 2013), soaked into a solution containing 1 mM AT-9010, and the structure was solved using molecular replacement (Table 2).

Figure 4 detail the atomic contacts of AT-9010 bound to the cap-binding pocket of the DENV3 NS5 MTase domain (Fig. 4B). Unlike for GTP in other structures ((Egloff et al., 2002), for a review, see (Lim et al., 2015)), all three phosphate group densities are perfectly defined, including water molecules making specific contacts with their oxygen substituents (Fig. 4C, D). The AT-9010 molecule is a close mimic of GTP, which binds in a position superimposable to that of AT-9010 (Fig. 4E). Interestingly, the 2'-F, 2'-C-methyl groups that distinguish the analogue from its natural GTP counterpart do not alter the ribose pose. The side chain of Lys14, however, is pointing towards the fluorine atom. Lys14 is clearly the 'ribose check' residue, since in other structures (eg., PDB: 5DTO), it makes simultaneous contacts with the 3'- and the 2'-hydroxyls of the GTP cap, at 2.7 and 2.4 Å, respectively. Here, the fluorine atom does not engage into a hydrogen bond, the distance is kept at 3 Å. Similar to GTP, AT-9002 presumably establishes hydrogen bonds with both its 3'- and 2'-OH, explaining the 10-fold lower IC_{50}. Likewise, the side chain of Asn18 could provide an unfavorable interaction, but its contact atoms are at a distance of 3.3 Å, thus presumably escaping any significant discrimination from GTP at the binding step.

Interestingly, we did not detect any significant inhibition of the RNA-cap N7-guanine MTase activity at concentrations of AT-9010 up to 1mM using an unmethylated $GpppAC_4$ RNA substrate (not shown). This indicates that the GTP/cap-binding mode used when the N7-guanine is poised to receive the methyl group is different from the one observed here, or in other words, N7-methylation by another MTase domain in *trans* is unlikely.

We conclude that AT-9010, whose concentration builds up to hundreds of micromolar in exposed cells (Good et al., 2020), likely interferes with RNA capping through inhibition of the NS5 MTase domain.

AT-9010 does not compete significantly against GTP during dinucleotide primer (pppApG) synthesis

The RdRp domain is able to catalyze both *de novo* primer synthesis and elongation of RNA. We thus investigated how AT-9010 impacts the nucleotide polymerization activities at play at the RdRp active site. To start RNA synthesis from the 3'-end of either (+) genomic or (-) antigenomic strand of the genome, DENV NS5 harbors a *de novo* RNA synthesis activity: the enzyme synthesizes a dinucleotide primer, pppApG, whose sequence is strictly conserved within the *Flavivirus* genus (Selisko et al., 2012). RNA synthesis starts using the very last nucleotide of the genome and antigenome RNA as a template. We have shown before that pppApG is formed in the presence of Mg^{2+} and more efficiently of Mn^{2+} as catalytic ions, the latter even supporting pppApG synthesis in the absence of template. In order to test the effect of AT-9010 on *de novo* initiation we used a short oligonucleotide (AACAGGUUCU) corresponding to the 3'end of the DENV2 genome as a template directing synthesis of pppApG, pppApGpA and pppApGpApA in the presence of ATP and GTP. Figure 5A shows PAGE analysis of reaction products with ATP (with radioactive [γ-^{32}P]-ATP as a tracer), GTP and/or AT-9010 at given concentrations.

As expected in the presence of ATP, GTP and Mn^{2+} we observe the efficient formation of pppApG to pppApGpApA, as well as of pppApA to a minor extent. *De novo* initiation is less efficient in the presence of Mg^{2+} and renders mainly pppApG. When only ATP and AT-9010 are present, with Mn^{2+}, more pppApA is formed but also the dinucleotide pppApAT-9010 which is not further extended. AT-9010 thus acts also in the presence of Mn^{2+} as a *bona fide* chain terminator. We quantified the pppApAT-9010 band formation, compared it to pppApG and found that GTP is around 100 times more efficiently incorporated than AT-9010.

In the presence of Mg^{2+}, we do not observe any formation of pppApAT-9010. When using increasing concentrations of AT-9010 in the presence of 100 µM GTP we observe inhibition of pppApG formation (Fig. 5A). Quantification of product bands (Fig. 5B) allows the determination of IC_{50} values of ~ 315 µM and ~107 µM in the presence of Mn^{2+} and Mg^{2+}, respectively.

We conclude that AT-9010 is strongly discriminated against upon *de novo* initiation by DENV RdRp in comparison to the elongation step (see below). We measured a weak inhibitory effect of AT-9010 on pppApG formation. Our results suggest that the observed anti-*Flavivirus* effect of AT-752 is not significantly based on inhibition of primer pppApG synthesis.

AT-9010 is a non-obligate chain terminator discriminated 9-to-14-fold towards its natural GTP counterpart by NS5 from all four DENV serotypes

A ~12-fold discrimination of AT-9010 relative to GTP has been measured previously at the DENV2 NS5 active site during RNA elongation (Good et al., 2021a). We set up experiments to determine the AT-9010 chain terminator activity spectrum against all four DENV serotypes in experiments measuring competitive nucleotide incorporation as well as single nucleotide incorporation.

For the first set-up, the DENV NS5 of individual serotypes 1-4 was bound to a DENV2 specific primer/template combination corresponding to 10 nt of the 5'end of the genome (P_{10}) annealed to 20 nt of the 3'end of the antigenome (T_{20}) (Good et al., 2021a). This P_{10}/T_{20} combination allows incorporation of GTP at specific positions (here, at position +5, +7, and +8 on the RNA template). The binary complex NS5.P_{10}/T_{20} was challenged with a nucleotide mix containing various ratios of AT-9010 versus GTP. The discrimination against AT-9010 can then

be determined by quantifying at each G /AT-9010 position the percentage of readthrough bands in relation to the percentage of the AT-9010 band multiplied by the concentration ratio of AT-9010 to GTP.

A typical result is shown for DENV4 NS5 in Figure 6A. First, in absence of GTP (three reactions on the left part of the gel) we show that AT-9010 is incorporated and acts as a chain terminator. When increasing AT-9010 concentrations are used in presence of GTP the AT-9010 band at position +5 clearly accumulates (right part of the gel). The quantification of percentage of the readthrough bands versus the AT-9010 band at this position, resulted in a ~14-fold discrimination against AT-9010 versus GTP. The experiments were replicated for all four serotypes (Supplementary Fig. 1A and 1B). Whereas all incorporated AT-9010 in a similar way, determination of discrimination was difficult when the overall activity was less, as for DENV1 and DENV3 NS5. Therefore, we turned to single nucleotide incorporation experiments (see below). Notably, AT-9010 bears chemical resemblance to the activated form of sofosbuvir (2'-F, 2'-C-methyl uracil 5'-triphosphate (SOF-TP)). Sofosbuvir is used as a potent treatment against viral hepatitis caused by the related *Flaviviridae* HCV. SOF, however, shows in infected cells an approximately 20-fold higher IC_{50} activity against DENV compared to HCV (Good et al., 2021a, 2020). We thus repeated this experiment using the same P_{10}/T_{20} system where SOF-TP is potentially incorporated at positions +2 and +6, but as expected did not detect any significant use of SOF-TP (not shown).

In order to determine discrimination of AT-9010 for all four DENV NS5 proteins we used a second set-up of single nucleotide incorporation. We compared the relative incorporation efficiencies of AT-9010 (and AT-9002) relative to GTP at position +1 of a non-specific RNA P_{10}/T_{20} system allowing the incorporation of G as the first nucleotide, using the NAs or GTP as the sole nucleotide at the same concentration. The same design but employing another non-specific RNA P_{10}/T_{20} system allowing the incorporation of U as the first nucleotide was used for SOF-TP and UTP. Figure 6B shows comparative incorporation into RNA of AT-9010 and AT-9002 relative to GTP, as well as SOF-TP relative to UTP using DENV4 NS5 (see Supplementary Figures 2 and 3 for the other serotypes). Steady-state initial incorporation rates were determined (Supplementary Fig.3). The NAs and natural NTP incorporation rates were used to calculate their ratio corresponding to the preference for the natural NTP or the discrimination of NAs relative to their natural nucleotide counterpart (Table 3). None of the NAs are better substrates than their natural nucleotide counterparts. AT-9010 is discriminated against GTP about 11-fold. The efficiency of incorporation of AT-9010 is highest and that of SOF-TP is lowest amongst these three analogues. Being discriminated >100 fold relative to UTP, the pyrimidine base of SOF-TP is probably a key factor driving a loss of potency as a substrate analogue for NS5 compared to AT-9010 (see discussion). Contrary to the MTase inhibition assay, AT-9002 is not performing better than AT-9010 as a RdRp substrate.

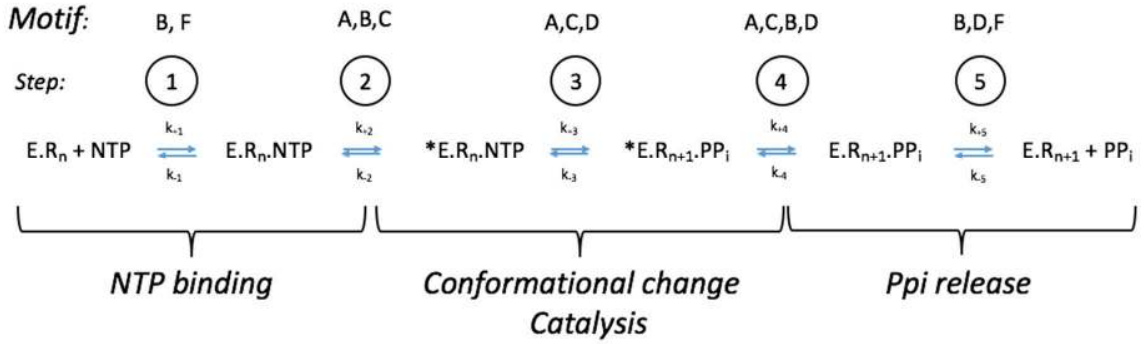
Given that AT-752 can be activated to hundreds of micromolar concentrations of AT-9010 in various cells (Good et al., 2020), this indicates that the NA should be incorporated significantly in the ~11 kb DENV genomes during virus replication, and act as a broad spectrum anti-DENV- and anti-*Flavivirus* agent.

Active-site structural comparison of Hepacivirus, Flavivirus, and Coronavirus RNA polymerases

AT-9010 is a nucleotide analogue active against three different viral RdRps, those of HCV NS5b (Good et al., 2021b, 2020), DENV1 to 4 NS5 (Good et al., 2021b, and this work), and SARS-CoV-2 nsp12 (Shannon et al., 2022). There are commonalities in the active site of these viral RdRps (Peersen, 2019), and they translate into an amino-acid sequence conservation around the catalytic center. For all these RdRps, the chemical reaction allows templated ribonucleotide addition at the RNA 3'-end using a common metal-dependent phosphoryl transfer mechanism.

There are, however, subtle amino acid changes in the conserved Motifs A to G of each viral RdRp, which *in fine* translate into the different NA discriminations measured against their natural NTP. In particular, the inefficient substrate properties of SOF-TP towards both SARS-CoV-2 RdRp (Shannon et al., 2022) and DENV1 to 4 NS5 is puzzling (Potisopon et al., 2017, and this work). SOF-TP is an efficient UTP analogue substrate for the HCV RdRp, which bears significant structural homology with DENV NS5.

Once the polymerase is bound to its dsRNA substrate, nucleotide addition can be represented by a series of events which include three basic steps: NTP binding, conformational change and catalysis, and PPi release (Scheme 1).



Scheme 1. Representation of the kinetic pathway of NTP (or NA-TP) incorporation into RNA.

Unraveling the detailed involvement of the different RdRp motifs in these reactions schemes represents a worthwhile challenge, because it would aid designing better NAs as well as understanding how mutations could confer drug-resistance.

Figure 7 show the comparative models of the HCV, DENV2, and SARS-CoV-2 RdRps in complex with the primer RNA poised for the nucleophilic attack of the primer 3'-OH on the nucleotide α-phosphate. Although HCV and DENV belong to the same *Flaviviridae* family and their RdRps share a significant homology at the amino-acid level, Fig. 7 panel A shows that there are significant differences in conserved motifs involved in nucleotide checking and fidelity.

In (Potisopon et al., 2017), we concentrated in differences of the appropriate closure of Motif A being next to the 2' position of the ribose of the incoming nucleotide. Here we want to add as an obvious difference that HCV NS5b does not have a classic Motif D (Selisko et al., 2018), which is typified by a single K689 and K798 in DENV NS5 and SARS-CoV-2 nsp12 (Fig. 7, panel A and B). Note that HCV NS5b has 2 positively charged amino acids Motif A residues R222 and H223 located in the vicinity of the γ-phosphate of the incoming NTP, whose position is only inferred in panel A, since the crystallized incoming nucleotide is the diphosphate form of sofosbuvir (SOF-DP, panel A). DENV1 to 4 NS5 contain at these positions conserved residues Ala and Gly (Potisopon et al., 2017). Interestingly, HCV is also different from both DENV and SARS-CoV-2 RdRps, in that it carries R48 that does not seem to have any equivalent in either

DENV nor in SARS-CoV-2 RdRps. R48 is close to the β-phosphate of the NTP substrate, which is the phosphate adjacent to the cleaved bond generating the leaving pyrophosphate (Fig. 7 panel A). We also note that the SARS-CoV-2 RdRp is somehow unique in this spatial region, since its Motif F containing residues K551 and R553 forms a peculiar loop unique here to SARS-CoV-2 nsp12, but present also in other CoVs and large RNA genome virus RdRps (Campagnola et al., 2022; Ferron et al., 2021).

Note that the amino acid differences in these otherwise conserved motifs reported in Figure 7 do not concern recognition of the 2'-OH by Motif B, which contains the same amino acids at structurally conserved positions in all three viral RdRps (residues S282, T287 and N291 in the case of HCV NS5b, (Appleby et al., 2015)).

Active site closure is considered to be a main check point of the correctness of the nucleotide substrate. It involves a movement of Motif A towards Motif C (Gong and Peersen, 2010), as well as closure of Motif D lysine onto the leaving pyrophosphate (Yang et al., 2012). Therefore, we surmise that the better SOF-TP incorporation by the HCV RdRp relative to other RdRps is likely due to Motif A and its pseudo-Motif D: in HCV NS5b, the pseudo-Motif D allows catalysis and passing over its fidelity checkpoint for SOF-TP, whereas for DENV NS5 and SARS-CoV-2 nsp12, Motifs A and D (there, in fact, mainly its typical lysine) do not validate the fidelity check for SOF-TP, which would correspond to step 3 in Scheme 1. In contrast, Motifs A and D of DENV1 to 4 NS5 and SARS-CoV-2 nsp12, validate the incorporation of AT-9010 where base pairing seems to allow a better accommodation of the NA even better than AT-9002.

Discussion

The in vitro activities of AT-281 against all four DENV serotypes reported herein using Huh-7 cells infected with reporter viruses (EC_{50} of 0.3-0.75 µM) are consistent with the previously reported EC_{50} values for AT-281 against DENV-2 and -3 (Good et al., 2021a) using Huh-7, BHK-21 and human peripheral blood mononuclear cells infected with unmodified viral strains (EC_{50} of 0.0.48-1.1 µM). Although most transformed cells are incapable of forming AT-9010 to any appreciable extent, these three cell types, as well as all primary mammalian cells and tissues investigated in vitro and in vivo can efficiently activate AT-281 and/or its diastereomer AT-511 to the active triphosphate metabolite (data on file). Additionally, the lack of cytotoxicity of AT-281 was established in these previous assays with 50% cytotoxicity values exceeding the highest concentrations tested (>85 µM in PBMCs and >170 µM in Huh-7 and BHK-21 cells).

We have determined the mechanism of action of AT-752 through the study of the interaction of its GTP analogue AT-9010 with the most conserved flavivirus enzyme, NS5. This protein a bi-functional enzyme expressing an RNA MTase at its N-terminus ~30 KDa domain, and an RNA-dependent RNA polymerase at its C-terminus ~70 KDa domain.

The MTase activity is also bi-functional, as the said domain is able to perform both an RNA cap N7-guanine methylation as well as a 2'-O methylation on the first transcribed nucleotide, a strictly conserved adenosine in the flavivirus genus. Remarkably, AT-9010 is able to selectively inhibit methylation at the 2'-OH but not at the N7-guanine. This selectivity is in line with crystal structures obtained with various RNA and cofactor substrates: If one considers that the SAM donor binding mode is identical for both methylation reactions, the GTP/RNA cap binding site would prevent binding of RNA cap substrates, whereas the N7-guanine poised for methylation supposedly would not overlap with the cap binding site.

We solved the crystal structure of AT-9010 in complex with the DENV3 MTase. It reveals that AT-9010 binds to the GTP/cap binding site in a similar manner as GTP, and its 2'-fluoro, 2'-methyl groups do not clash with any amino acid side chains. We note that the 2'-F removes a positive interaction with a Lys14, accounting for a ~10-fold better inhibition power of AT-9002. The contribution of AT-9010 binding on the MTase cap-binding site to the inhibition of viral replication in infected cells remains to be measured. Indeed, reverse genetics studies have shown that inhibition of 2'-O-MTase activity -independently of N7 MTase activity of NS5- attenuates viral replication in infected cell cultures (Chang et al., 2016). Nevertheless, if the inhibition is limited to the 2'-O-MTase activity we can expect that viral RNA lacking a cap-1 structure will be efficiently detected by the RIG-like receptor RIG-I contributing to an antiviral response involving type -1 interferon secretion . (Daffis et al., 2010). Moreover, it is also possible that the presence of the AT9010 molecule in the cap-binding site of the MTase could affect the capping activity. This possibility was not addressed in this work. If AT-9010 corrupts or inhibits viral RNA capping, we expect that this compound would induce a strong antiviral effect as the incorrectly capped viral RNAs are barely translated into viral proteins. Finally, it should also be noted that even with the modest IC_{50} observed for the MTase, AT-9010 reaches high µM concentration in cells (Good et al., 2020) and could indeed participate to the inhibitor efficiency. Likewise, other steps downstream of the capping reaction might also be altered indeed, such as the co-associated conformational changes and cross-talks between the MTase and RdRp domains (El Sahili and Lescar, 2017; van den Elsen et al., 2021).

In the *Flaviviridae* family, the polymerase activity has been shown to catalyze two kinetically distinct reactions: first, a primer is assembled from ATP and GTP to form the initiating pppApG

dinucleotide (Selisko et al., 2012). This reaction is distributive and requires high concentrations of both nucleotides. Then, as NS5 binds to specific RNA structures (Lee et al., 2021), pppApG is elongated in a processive manner to synthesize full-length genomes of antigenome RNAs (Ackermann and Padmanabhan, 2001). We show that primer synthesis is barely inhibited by AT-9010, suggesting that the active site adopts a more stringent conformation disfavoring the incorporation of AT-9010 into the dinucleotide primer. However, this discrimination decreases significantly when the RdRp enters in its elongation mode. Overall, AT-9010 becomes discriminated over GTP about 14-fold, which should translate into a significant chain-termination effect given that AT-9010 reaches intracellular concentration in the same range as GTP (Good et al., 2020).

Notably, we also observed differences in overall RdRp activity across the four DENV seroptypes, as well as for Zika virus RdRp (not shown), that we interprete being due to serotype - specific sequence requirements for optimal binding to primer/template systems. However, we find discrimination values in the same order of magnitude 7- to 14-fold irrespective of the method used to measure GTP preference, *ie.*, single incorporation or multiple incorporations. The absence of difference between these two methods suggests that pyrophosphorolysis (the reversal of the polymerisation reaction shown in Scheme 1) of the chain-terminating analogue during multiple NTP addition does not occur significantly.

Intriguingly, in the +1 primer extension assay AT-9002 appears to be a 40-fold less efficient substrate than GTP, and 4-fold lower than AT-9010. In the presence of a 2'-C-methyl, the presence of a 2'-OH appears to decrease its substrate properties compared to 2'-F, suggesting that a peculiar, unexpected nucleotide conformation specific to AT-9002 is responsible for its decreased incorporation at the RdRp active site, but not its binding to the MTase active site.

2'-C-methyl GTP (AT-9002) and 2'-C-methyl ATP have been evaluated in enzyme assays by others (Lu et al., 2020). Also, 7-deaza 2'-C-methyl adenosine has shown interesting antiviral properties against flaviviruses, but its clinical evaluation was stopped due to toxicity issues (Arnold et al., 2012). Since bemnifosbuvir nor AT-752 do not show significant toxicity, the modified base or the 2'-F, or both, play a key role in evading toxicity issues when compared to the natural 2'-OH of AT-9002. It remains of great interest in the drug-design field to determine if the observed toxicity is due to the purine base or the 2'-OH of these analogues.

Last, although sofosbuvir 5'-triphosphate shares with AT-9010 the same modified ribose structure, the former is much less efficiently incorporated in RNA than the latter by both DENV and SARS-CoV-2 RdRp ((Shannon et al., 2022), and this work), indicating that a subtle structural difference of both RdRps with the HCV NS5b active site is responsible for this enhanced sofosbuvir 5'-triphosphate discrimination. A serine in Motif B (S282 in HCV) has been proposed to play a central role in checking the NTP 2'-OH prior to incorporation (Appleby et al., 2015). However, the Motif B tetrad (T287, S282, D225, N291 in the case of HCV NS5b, (Appleby et al., 2015) is conserved in all three RdRps studied here. Boehr et al. (Boehr et al., 2019) have proposed that, in poliovirus RdRp, this spatially equivalent serine might not be interfering in the initial NTP binding, but during closure of the active site upon binding of the next NTP because of steric hindrance due to the 2'-C-methyl group. This proposition was confirmed in the SARS-CoV-2 polymerase structure determined in complex with RNA and AT-9010 (Boehr et al., 2019). We thus propose that the present increased discrimination of SOF-TP observed here in DENV and SARS-CoV-2 active sites is due to the Motif D fidelity check-point, which could be more stringent in the case of these latter enzymes carrying a Motif D lysine. Then, to explain the difference between AT-9010 and SOF-TP efficiency in both DENV and

SARS-CoV-2 RdRps, we propose that the superior stacking of the purine relative to the pyrimidine base in the active site of RdRps contributes to the selection process, overcoming AT-9010 Motif D fidelity check enough to yield significant RNA chain-termination. In line with these results, bemnifosbuvir is 10-fold more potent than sofosbuvir in HCV infected cells (Good et al., 2020), and sofosbuvir does not reach a low EC_{50} in DENV or SARS-CoV2 infected cells when compared to HCV infected cells ((Chen et al., 2015) and references therein).

In conclusion, AT-752 shows a mechanism of action consistent with its use as an anti-dengue drug. It would be of interest, given the NS5 conservation across the *Flavivirus* genus, to evaluate it further against other emerging viruses such as *eg*., Kyasanur forest and Omsk viruses (BSL-4 agents), or Tick-Borne Encephalitis virus.

Acknowledgements

The authors acknowledge SOLEIL for provision of synchrotron radiation facilities and we would like to thank beamline teams of Proxima-1 and Proxima-2 for their assistance during data collection.

Competing interests: S.G., K.L., A.M. and J.P.S. are employees of ATEA Pharmaceuticals, Inc. The other authors declare no competing interests.

Funding

This work was supported by grants from the Fondation pour la Recherche Medicale (Aide aux Equipes), and from ATEA Pharmaceuticals. MF and PSO received a post-doctoral fellowship as part of the grant from ATEA Pharmaceuticals. A part of the work was performed on the Aix-Marseille University Antiviral Drug-Design Platform (www.afmb.univ-mrs.fr/facility/plateforme-de-criblage-marseille-luminy/), and benefited from the Agence Nationale pour la Recherche through the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01].

References

Ackermann, M., Padmanabhan, R., 2001. De novo synthesis of RNA by the dengue virus RNA-dependent RNA polymerase exhibits temperature dependence at the initiation but not elongation phase. J. Biol. Chem. 276, 39926–39937. https://doi.org/10.1074/jbc.M104248200

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221. https://doi.org/10.1107/S0907444909052925

Appleby, T.C., Perry, J.K., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D., Wetmore, D.R., McGrath, M.E., Ray, A.S., Sofia, M.J., Swaminathan, S., Edwards, T.E., 2015. Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. Science 347, 771–775. https://doi.org/10.1126/science.1259210

Arnold, J.J., Sharma, S.D., Feng, J.Y., Ray, A.S., Smidansky, E.D., Kireeva, M.L., Cho, A., Perry, J., Vela, J.E., Park, Y., Xu, Y., Tian, Y., Babusis, D., Barauskus, O., Peterson, B.R., Gnatt, A., Kashlev, M., Zhong, W., Cameron, C.E., 2012. Sensitivity of mitochondrial transcription and resistance of RNA polymerase II dependent nuclear transcription to antiviral ribonucleosides. PLoS Pathog 8, e1003030. https://doi.org/10.1371/journal.ppat.1003030

Baker, C., Liu, Y., Zou, J., Muruato, A., Xie, X., Shi, P.-Y., 2020a. Identifying optimal capsid duplication length for the stability of reporter flaviviruses. Emerging Microbes & Infections 9, 2256–2265. https://doi.org/10.1080/22221751.2020.1829994

Baker, C., Xie, X., Zou, J., Muruato, A., Fink, K., Shi, P.-Y., 2020b. Using recombination-dependent lethal mutations to stabilize reporter flaviviruses for rapid serodiagnosis and drug discovery. EBioMedicine 57, 102838. https://doi.org/10.1016/j.ebiom.2020.102838

Barral, K., Sallamand, C., Petzold, C., Coutard, B., Collet, A., Thillier, Y., Zimmermann, J., Vasseur, J.-J., Canard, B., Rohayem, J., Debart, F., Decroly, E., 2013. Development of specific dengue virus 2'-O- and N7-methyltransferase assays for antiviral drug screening. Antiviral Res. 99, 292–300. https://doi.org/10.1016/j.antiviral.2013.06.001

Benarroch, D., Selisko, B., Locatelli, G.A., Maga, G., Romette, J.-L., Canard, B., 2004. The RNA helicase, nucleotide 5'-triphosphatase, and RNA 5'-triphosphatase activities of Dengue virus protein NS3 are Mg2+-dependent and require a functional Walker B motif in the helicase catalytic core. Virology 328, 208–218. https://doi.org/10.1016/j.virol.2004.07.004

Boehr, A.K., Arnold, J.J., Oh, H.S., Cameron, C.E., Boehr, D.D., 2019. 2'-C-methylated nucleotides terminate virus RNA synthesis by preventing active site closure of the viral RNA-dependent RNA polymerase. J Biol Chem 294, 16897–16907. https://doi.org/10.1074/jbc.RA119.010214

Campagnola, G., Govindarajan, V., Pelletier, A., Canard, B., Peersen, O.B., 2022. The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. J Virol 96, e0067122. https://doi.org/10.1128/jvi.00671-22

Chang, D.C., Hoang, L.T., Mohamed Naim, A.N., Dong, H., Schreiber, M.J., Hibberd, M.L., Tan, M.J.A., Shi, P.-Y., 2016. Evasion of early innate immune response by 2'-O-

methylation of dengue genomic RNA. Virology 499, 259–266. https://doi.org/10.1016/j.virol.2016.09.022

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. *MolProbity*: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66, 12–21. https://doi.org/10.1107/S0907444909042073

Chen, Y.-L., Yokokawa, F., Shi, P.-Y., 2015. The search for nucleoside/nucleotide analog inhibitors of dengue virus. Antiviral Res 122, 12–19. https://doi.org/10.1016/j.antiviral.2015.07.010

Daffis, S., Szretter, K.J., Schriewer, J., Li, J., Youn, S., Errett, J., Lin, T.-Y., Schneller, S., Zust, R., Dong, H., Thiel, V., Sen, G.C., Fensterl, V., Klimstra, W.B., Pierson, T.C., Buller, R.M., Gale, M., Shi, P.-Y., Diamond, M.S., 2010. 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. Nature 468, 452–456. https://doi.org/10.1038/nature09489

Doughty, C.T., Yawetz, S., Lyons, J., 2017. Emerging Causes of Arbovirus Encephalitis in North America: Powassan, Chikungunya, and Zika Viruses. Curr Neurol Neurosci Rep 17, 12. https://doi.org/10.1007/s11910-017-0724-3

Dousson, C.B., 2018. Current and future use of nucleo(s)tide prodrugs in the treatment of hepatitis C virus infection. Antivir Chem Chemother 26, 2040206618756430. https://doi.org/10.1177/2040206618756430

Egloff, M.-P., Benarroch, D., Selisko, B., Romette, J.-L., Canard, B., 2002. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. EMBO J. 21, 2757–2768. https://doi.org/10.1093/emboj/21.11.2757

El Sahili, A., Lescar, J., 2017. Dengue Virus Non-Structural Protein 5. Viruses 9, E91. https://doi.org/10.3390/v9040091

Emsley, P., Cowtan, K., 2004. *Coot*: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60, 2126–2132. https://doi.org/10.1107/S0907444904019158

Ferron, F., Sama, B., Decroly, E., Canard, B., 2021. The enzymes for genome size increase and maintenance of large (+)RNA viruses. Trends Biochem Sci 46, 866–877. https://doi.org/10.1016/j.tibs.2021.05.006

Filomatori, C.V., Iglesias, N.G., Villordo, S.M., Alvarez, D.E., Gamarnik, A.V., 2011. RNA sequences and structures required for the recruitment and activity of the dengue virus polymerase. J. Biol. Chem. 286, 6929–6939. https://doi.org/10.1074/jbc.M110.162289

Ghetia, C., Bhatt, P., Mukhopadhyay, C., 2022. Association of dengue virus non-structural-1 protein with disease severity: a brief review. Trans R Soc Trop Med Hyg trac087. https://doi.org/10.1093/trstmh/trac087

Gong, P., Peersen, O.B., 2010. Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 107, 22505–22510. https://doi.org/10.1073/pnas.1007626107

Good, S.S., Moussa, A., Zhou, X.-J., Pietropaolo, K., Sommadossi, J.-P., 2020. Preclinical evaluation of AT-527, a novel guanosine nucleotide prodrug with potent, pan-genotypic activity against hepatitis C virus. PLoS One 15, e0227104. https://doi.org/10.1371/journal.pone.0227104

Good, S.S., Shannon, A., Lin, K., Moussa, A., Julander, J.G., La Colla, P., Collu, G., Canard, B., Sommadossi, J.-P., 2021a. Evaluation of AT-752, a Double Prodrug of a Guanosine Nucleotide Analog with In Vitro and In Vivo Activity against Dengue and Other Flaviviruses. Antimicrob Agents Chemother 65, e0098821. https://doi.org/10.1128/AAC.00988-21

Good, S.S., Westover, J., Jung, K.H., Zhou, X.-J., Moussa, A., La Colla, P., Collu, G., Canard, B., Sommadossi, J.-P., 2021b. AT-527, a double prodrug of a guanosine nucleotide analog, is a potent inhibitor of SARS-CoV-2 in vitro and a promising oral antiviral for treatment of COVID-19. Antimicrob Agents Chemother. https://doi.org/10.1128/AAC.02479-20

Kaptein, S.J.F., Goethals, O., Kiemel, D., Marchand, A., Kesteleyn, B., Bonfanti, J.-F., Bardiot, D., Stoops, B., Jonckers, T.H.M., Dallmeier, K., Geluykens, P., Thys, K., Crabbe, M., Chatel-Chaix, L., Münster, M., Querat, G., Touret, F., de Lamballerie, X., Raboisson, P., Simmen, K., Chaltin, P., Bartenschlager, R., Van Loock, M., Neyts, J., 2021. A pan-serotype dengue virus inhibitor targeting the NS3-NS4B interaction. Nature 598, 504–509. https://doi.org/10.1038/s41586-021-03990-6

Kuczera, D., Assolini, J.P., Tomiotto-Pellissier, F., Pavanelli, W.R., Silveira, G.F., 2018. Highlights for Dengue Immunopathogenesis: Antibody-Dependent Enhancement, Cytokine Storm, and Beyond. J. Interferon Cytokine Res. 38, 69–80. https://doi.org/10.1089/jir.2017.0037

Lantez, V., Dalle, K., Charrel, R., Baronti, C., Canard, B., Coutard, B., 2011. Comparative Production Analysis of Three Phlebovirus Nucleoproteins under Denaturing or Non-Denaturing Conditions for Crystallographic Studies. PLoS Negl Trop Dis 5, e936. https://doi.org/10.1371/journal.pntd.0000936

Lee, E., Bujalowski, P.J., Teramoto, T., Gottipati, K., Scott, S.D., Padmanabhan, R., Choi, K.H., 2021. Structures of flavivirus RNA promoters suggest two binding modes with NS5 polymerase. Nat Commun 12, 2530. https://doi.org/10.1038/s41467-021-22846-1

Lim, S.P., Noble, C.G., Shi, P.-Y., 2015. The dengue virus NS5 protein as a target for drug discovery. Antiviral Res. 119, 57–67. https://doi.org/10.1016/j.antiviral.2015.04.010

Lim, S.P., Sonntag, L.S., Noble, C., Nilar, S.H., Ng, R.H., Zou, G., Monaghan, P., Chung, K.Y., Dong, H., Liu, B., Bodenreider, C., Lee, G., Ding, M., Chan, W.L., Wang, G., Jian, Y.L., Chao, A.T., Lescar, J., Yin, Z., Vedananda, T.R., Keller, T.H., Shi, P.-Y., 2011. Small molecule inhibitors that selectively block dengue virus methyltransferase. J. Biol. Chem. 286, 6233–6240. https://doi.org/10.1074/jbc.M110.179184

Lin, K., Good, S.S., Julander, J.G., Weight, A.E., Moussa, A., Sommadossi, J.-P., 2022. AT-752, a double prodrug of a guanosine nucleotide analog, inhibits yellow fever virus in a hamster model. PLoS Negl Trop Dis 16, e0009937. https://doi.org/10.1371/journal.pntd.0009937

Lu, G., Zhang, X., Zheng, W., Sun, J., Hua, L., Xu, L., Chu, X.-J., Ding, S., Xiong, W., 2020. Development of a Simple In Vitro Assay To Identify and Evaluate Nucleotide Analogs against SARS-CoV-2 RNA-Dependent RNA Polymerase. Antimicrob Agents Chemother 65, e01508-20. https://doi.org/10.1128/AAC.01508-20

Masmejan, S., Musso, D., Vouga, M., Pomar, L., Dashraath, P., Stojanov, M., Panchaud, A., Baud, D., 2020. Zika Virus. Pathogens 9, E898. https://doi.org/10.3390/pathogens9110898

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. *Phaser* crystallographic software. J Appl Crystallogr 40, 658–674. https://doi.org/10.1107/S0021889807021206

Moquin, S.A., Simon, O., Karuna, R., Lakshminarayana, S.B., Yokokawa, F., Wang, F., Saravanan, C., Zhang, J., Day, C.W., Chan, K., Wang, Q.-Y., Lu, S., Dong, H., Wan, K.F., Lim, S.P., Liu, W., Seh, C.C., Chen, Y.-L., Xu, H., Barkan, D.T., Kounde, C.S., Sim, W.L.S., Wang, G., Yeo, H.-Q., Zou, B., Chan, W.L., Ding, M., Song, J.-G., Li, M., Osborne, C., Blasco, F., Sarko, C., Beer, D., Bonamy, G.M.C., Sasseville, V.G., Shi, P.-Y., Diagana, T.T., Yeung, B.K.S., Gu, F., 2021. NITD-688, a pan-serotype inhibitor of the dengue virus NS4B protein, shows favorable pharmacokinetics and efficacy in preclinical animal models. Sci. Transl. Med. 13, eabb2181. https://doi.org/10.1126/scitranslmed.abb2181

Murakami, E., Wang, T., Babusis, D., Lepist, E.-I., Sauer, D., Park, Y., Vela, J.E., Shih, R., Birkus, G., Stefanidis, D., Kim, C.U., Cho, A., Ray, A.S., 2014. Metabolism and pharmacokinetics of the anti-hepatitis C virus nucleotide prodrug GS-6620. Antimicrob Agents Chemother 58, 1943–1951. https://doi.org/10.1128/AAC.02350-13

Paesen, G.C., Collet, A., Sallamand, C., Debart, F., Vasseur, J.-J., Canard, B., Decroly, E., Grimes, J.M., 2015. X-ray structure and activities of an essential Mononegavirales L-protein domain. Nat Commun 6, 8749. https://doi.org/10.1038/ncomms9749

Peersen, O.B., 2019. A Comprehensive Superposition of Viral Polymerase Structures. Viruses 11. https://doi.org/10.3390/v11080745

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25, 1605–1612. https://doi.org/10.1002/jcc.20084

Pierson, T.C., Diamond, M.S., 2020. The continued threat of emerging flaviviruses. Nat Microbiol 5, 796–812. https://doi.org/10.1038/s41564-020-0714-0

Potisopon S, Ferron F, Fattorini V, Selisko B, Canard B. 2017. Substrate selectivity of Dengue and Zika virus NS5 polymerase towards 2'-modified nucleotide analogues. Antiviral Res. 2017 Apr;140:25-36. doi: 10.1016/j.antiviral.2016.12.021.

Ramanathan, A., Robb, G.B., Chan, S.-H., 2016. mRNA capping: biological functions and applications. Nucleic Acids Res. 44, 7511–7526. https://doi.org/10.1093/nar/gkw551

Ray, D., Shah, A., Tilgner, M., Guo, Y., Zhao, Y., Dong, H., Deas, T.S., Zhou, Y., Li, H., Shi, P.-Y., 2006. West Nile virus 5'-cap structure is formed by sequential guanine N-7 and ribose 2'-O methylations by nonstructural protein 5. J. Virol. 80, 8362–8370. https://doi.org/10.1128/JVI.00814-06

Rivera, L., Biswal, S., Sáez-Llorens, X., Reynales, H., López-Medina, E., Borja-Tabora, C., Bravo, L., Sirivichayakul, C., Kosalaraksa, P., Martinez Vargas, L., Yu, D., Watanaveeradej, V., Espinoza, F., Dietze, R., Fernando, L., Wickramasinghe, P., Duarte MoreiraJr, E., Fernando, A.D., Gunasekera, D., Luz, K., Venâncioda Cunha, R., Rauscher, M., Zent, O., Liu, M., Hoffman, E., LeFevre, I., Tricou, V., Wallace, D., Alera, M., Borkowski, A., 2022. Three-year Efficacy and Safety of Takeda's Dengue Vaccine Candidate (TAK-003). Clinical Infectious Diseases 75, 107–117. https://doi.org/10.1093/cid/ciab864

Selisko, B., Dutartre, H., Guillemot, J.-C., Debarnot, C., Benarroch, D., Khromykh, A., Desprès, P., Egloff, M.-P., Canard, B., 2006. Comparative mechanistic studies of de novo RNA

synthesis by flavivirus RNA-dependent RNA polymerases. Virology 351, 145–158. https://doi.org/10.1016/j.virol.2006.03.026

Selisko, B., Papageorgiou, N., Ferron, F., Canard, B., 2018. Structural and Functional Basis of the Fidelity of Nucleotide Selection by Flavivirus RNA-Dependent RNA Polymerases. Viruses 10. https://doi.org/10.3390/v10020059

Selisko, B., Potisopon, S., Agred, R., Priet, S., Varlet, I., Thillier, Y., Sallamand, C., Debart, F., Vasseur, J.-J., Canard, B., 2012. Molecular basis for nucleotide conservation at the ends of the dengue virus genome. PLoS Pathog. 8, e1002912. https://doi.org/10.1371/journal.ppat.1002912

Shannon, A., Fattorini, V., Sama, B., Selisko, B., Feracci, M., Falcou, C., Gauffre, P., El Kazzi, P., Delpal, A., Decroly, E., Alvarez, K., Eydoux, C., Guillemot, J.-C., Moussa, A., Good, S.S., La Colla, P., Lin, K., Sommadossi, J.-P., Zhu, Y., Yan, X., Shi, H., Ferron, F., Canard, B., 2022. A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase. Nature Communications 13, 621. https://doi.org/10.1038/s41467-022-28113-1

Shannon, A., Selisko, B., Le, N.-T.-T., Huchting, J., Touret, F., Piorkowski, G., Fattorini, V., Ferron, F., Decroly, E., Meier, C., Coutard, B., Peersen, O., Canard, B., 2020. Rapid incorporation of Favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. Nat Commun 11, 4682. https://doi.org/10.1038/s41467-020-18463-z

van den Elsen, K., Quek, J.P., Luo, D., 2021. Molecular Insights into the Flavivirus Replication Complex. Viruses 13, 956. https://doi.org/10.3390/v13060956

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., Bricogne, G., 2011. Data processing and analysis with the *autoPROC* toolbox. Acta Crystallogr D Biol Crystallogr 67, 293–302. https://doi.org/10.1107/S0907444911007773

Wang, Y., Xie, X., Shi, P.-Y., 2022. Flavivirus NS4B protein: Structure, function, and antiviral discovery. Antiviral Research 207, 105423. https://doi.org/10.1016/j.antiviral.2022.105423

Yang, X., Smidansky, E.D., Maksimchuk, K.R., Lum, D., Welch, J.L., Arnold, J.J., Cameron, C.E., Boehr, D.D., 2012. Motif D of viral RNA-dependent RNA polymerases determines efficiency and fidelity of nucleotide addition. Structure 20, 1519–1527. https://doi.org/10.1016/j.str.2012.06.012

Material and methods

5'-triphosphate nucleosides, dinucleotide synthesis, and nucleic acid sequences

AT-9010, 5'-triphosphate 2'-fluoro-2'-C-methyl uridine, and Remdesivir 5'-triphosphate were from NuBlocks LLC, Oceanside, CA, USA. Other NTPs were purchased from GE Healthcare. Chemical synthesis of 5'-triphosphate (TP) dinucleotides pppUpU, pppUpG, pppGpU and pppApU was performed on an ABI 394 synthesizer (Applied Biosystems) and purified as described in (Shannon et al., 2020). The accession numbers of NS5 sequences used here are available at www.ncbi.nlm.nih.gov under the following accession numbers: DENV1_NS5: NP_722465; DENV2_NS5: NP_739590; DENV3_NS5: YP_001531176; DENV4_NS5: NP_740325 .

Antiviral assays

Transcripts of the dengue viruses (DENV1 WP, DENV2 NGC, DENV3 VN32, DENV4 MY01) were modified with a nano luciferase reporter between the 5'UTR and capsid gene as previously described (Baker et al., 2020b, 2020a) to create tagged viruses that are stable with cell culture passaging while producing a robust luciferase signal after inoculation for efficient screening of antiviral activity. Eight 2-fold serial dilutions of AT-281 were mixed with each reporter virus (MOIs of 0.1 for DENV1, DENV2 and DENV3; MOI of 0.001 for DENV4) and added to plates containing Huh-7 cells (RRID: CVCL 0336) that were seeded at 1×10^4 cells per well in a 96-well plate the previous day in Dulbecco's Modified Eagle Medium (Invitrogen, Carlsbad, CA) with 2% fetal bovine serum (Hyclone, Logan, UT). Cells were washed 48 h post infection three times with phosphate-buffered saline, followed by addition of NanoGlo® substrate diluted 1:100 in NanoGlo® Assay Buffer (Promega, Madison, WI). Luciferase activities were read by a BioTek Cytation 5 plate reader after 3 minutes. Data were analyzed with GraphPad Prism 9 software. Luciferase activities were normalized to DMSO treatment samples which were set to 100%. Error bars represent mean ± SD. Results are representative of three independent experiments with each one analyzed in triplicate.

Expression of DENV NS5, RdRp and MTase domains

Full length DENV NS5 and RdRp domains used in this study were expressed under the control of a T7-promoter in pET28a vectors in Escherichia coli (*E. coli*) NEB C2566 cells (New England Biolabs) carrying the pRARE2LacI (Novagen) plasmid. Proteins were expressed overnight at 17°C in TB (with 25 µg/ml Kanamycin and 17 µg/ml Chloramphenicol), following induction with 100 µM IPTG at an OD_{600nm} of 0.8 – 1. EtOH was added to 2% (v/v) final concentration.
DENV MTases were expressed under the control of a T7-promoter in pDEST17 vectors in Escherichia coli (*E. coli*)) NEB C2566 cells (New England Biolabs) carrying the pRARE2LacI (Novagen) plasmid. Proteins were expressed overnight at 17°C in TB (with 100 µg/mL Ampicillin and 17 µg/mL Chloramphenicol), following induction with 200 µM IPTG at an OD_{600nm} of 0.8 – 1. EtOH was added to 2% (v/v) final concentration.

Purification of DENV NS5

The cells were disrupted by sonication on ice in a lysis buffer (50 mM NaP pH 8, 1 M NaCl, 20% glycerol) supplemented with, 1.0 mg/mL Lysozyme, 22 µg/mL DNase, 1.6 % igepal, 0.5 mM TCEP and a complete protease inhibitor cocktail (COC) from Roche. Protein from the soluble fraction was loaded onto TALON® Superflow™ cobalt-based IMAC resin (Cytiva), washed 5 times with lysis buffer, 10 times with lysis buffer (igepal free) prior to elution with a lysis buffer supplemented with 250 mM imidazole and 250 mM glycine. Imidazole was removed by an overnight dialysis in a dialysis buffer (50 mM NaP pH 8, 20 % glycerol, 150 mM NaCl, 250 mM glycine and 0.5 mM TCEP) before to perform a last step purification using a size exclusion chromatography (SEC) HiLoad® 16/600 Superdex® 200 pg column in a buffer (10 mM HEPES pH 8, 300 mM NaCl, 10% glycerol and 0.5 mM TCEP). Protein was then concentrated to 10 to 15 $mg.mL^{-1}$ and stored at -80°C after a last dialysis in the SEC buffer supplemented with 40 % of glycerol.

Purification of DENV RdRp domains

The cells were disrupted by sonication on ice in a lysis buffer (50 mM NaP pH 8, 1 M NaCl, 20% glycerol) supplemented with, 1.0 mg/mL Lysozyme, 22 µg/mL DNase, 1.6 % igepal, 0.5 mM TCEP and a complete protease inhibitor cocktail (COC) from Roche. Protein from the soluble fraction were loaded onto TALON® Superflow™ cobalt-based IMAC resin (Cytiva), washed 5 times with lysis buffer, 10 times with lysis buffer (igepal free) prior to elution with a lysis buffer supplemented with 250 mM imidazole and 250 mM glycine. Imidazole was removed by an overnight dialysis in a dialysis buffer (50 mM NaP pH 8, 20 % glycerol, 150 mM NaCl, 250 mM glycine and 0.5 mM TCEP) before to perform a last step purification using a GE Hi-trap Heparine column. Protein was then concentrated to 10 to 15 $mg.mL^{-1}$ and stored at -80°C after a last dialysis in a final buffer containing 20 mM HEPES pH 8, 300 mM NaCl, 40 % of glycerol and 0.5 mM TCEP.

Purification of DENV MTase domains

The cells were disrupted by sonication on ice in a lysis buffer (50mM HEPES pH 7.5, 10 % glycerol, 500 mM NaCl and 5 mM Imidazole) supplemented with 0.2 mM Benzamidine, 1 mg/mL Lysozyme, 22 µg/mL DNase and 0.5 mM TCEP. Protein from the soluble fraction were loaded onto TALON® Superflow™ cobalt-based IMAC resin (Cytiva) and washed with lysis buffer (1M NaCl) prior to elution with 250 mM imidazole. The protein was finally purified with a SEC (HiLoad® 16/600 Superdex® 200 pg) in a buffer of 50 mM HEPES pH 8, 300 mM NaCl and 0.5 mM TCEP. Protein was then concentrated to 10 to 15 $mg.mL^{-1}$ and stored at -80°C after a last dialysis in a final buffer containing 20 mM HEPES pH 8, 300 mM NaCl, 40 % of glycerol and 0.5 mM TCEP.

Thermal shift assay

The thermal stability of the DENV1 to 4 proteins and their interaction with compounds were evaluated by using a thermal shift assay. Reactions were performed in white frame star PCR 96-well plates (Greiner ref 044705) with a final volume of 20 µl. The best conditions were found by using a 20 mM HEPES buffer, pH 7.5 containing 50 mM NaCl, 5 mM DTT, 10% glycerol and 1M Sorbitol.

Volumes of 0.4 µl of GTP, sinefungin or AT-9010 were disposed in the plate (100 µM final concentration). Water was used as a control without compound. 2 µM of NS5, 2 µM of RdRp domains or 6µM of MTase domains were mixed in the buffer with 2.5X, 0.75X or 1.5X of dye, respectively (final concentrations). The mix of protein and Protein thermal shift kit from ThermoFisher (ref 4461146) was prepared and 19.6 µl were dispensed into each well. The plate was sealed with adhesive film and then submitted to a run on a CFX96 RT-PCR (Bio-Rad). The program started by a 1-min equilibration phase at 20°C, followed by a temperature gradient of 0.5°C/30 s, from 20°C to 95°C, recording fluorescence every 0.5°C. Fluorescence was recorded using the FRET channel. Every experiment was performed in triplicate. Curves were plotted and analyzed to determine the melting temperature T_m by using the Boltzmann sigmoidal equation with the GraphPad Prism software.

pppApG Primer synthesis

Reactions were done in 50 mM HEPES buffer, pH 8.0 containing 10 mM KCl, 10 mM DTT and 10 µM template 5'-AACAGGUUCU-3' (corresponding to the 3' end of the DENV2 genome), 1 µM DENV2 NS5, 100 µM non-labeled NTPs and increasing concentrations of AT-9010, and either 5 mM $MgCl_2$ or 2 mM $MnCl_2$, as given in the Figure. Radiolabeled [γ-^{32}P]-ATP was used at 0.3 µCi per µl reaction volume (3000 Ci/mmol, Hartmann Analytics). Reactions were started by addition of a mixture of NTPs and catalytic ions. At given time points samples were taken, reactions stopped by adding an equal volume of FBD stop solution (formamide, 10 mM EDTA). Reaction products were separated using sequencing gels of 20% acrylamide-bisacrylamide (19:1), 7 M urea with TTE buffer (89 mM Tris pH 8.0, 28 mM taurine (2-aminoethanesulfonic acid), 0.5 mM EDTA). RNA product bands were visualized using an Amersham™ Typhoon™ Biomolecular Imager (Cytiva). The intensity of each band was quantified using the ImageQuant software (Cytiva). For IC_{50} measurements, values were normalized and fitted with Prism (GraphPad software) using the following equation: Y=100/(1+((X/IC_{50})^Hillslope)). IC_{50} is defined as the inhibitory compound concentration that causes 50 % reduction in enzyme activity. Reactions were done either four times (Mn) or twice (Mg), independently.

MTase activity assay

The transfer of tritiated methyl from [^{3}H]-SAM onto RNA substrate was monitored by filter-binding assay (FBA), performed according to the method described previously (Paesen et al., 2015). The recombinant methyltransferase domain of the non- structural protein NS5 from DENV serotype 2 (DENV NS5 MTase) corresponding to residues 1–296 of NS5 was expressed and purified as described previously (Egloff et al., 2002). FBA was carried out in reaction mixture [40 mM Tris-HCl (pH 8.0), 1 mM DTT, 2 µM SAM and 0.1 µM 3H-SAM (Perkin Elmer)] in the presence of 0.7 µM ^{m}GpppAC4 synthetic RNA and DENV2 NS5MTase protein (500 nM). The enzyme was first mixed with the compound (.5 to 1000 µM) suspended in water before the addition of RNA substrate and SAM and then incubated at 30°C. Reaction mixtures were stopped after 30 min by their 10-fold dilution in ice-cold water. Samples were transferred to diethylaminoethyl (DEAE) filtermat (Perkin Elmer) using a Filtermat Harvester (Packard Instruments). The RNA-retaining mats were washed twice with 10 mM ammonium formate pH 8.0, twice with water and once with ethanol. They were soaked with scintillation fluid (Perkin Elmer), and ^{3}H-methyl transfer to the RNA substrates was determined using a Wallac MicroBeta

TriLux Liquid Scintillation Counter (Perkin Elmer). For IC_{50} measurements, values were normalized and fitted with Prism (GraphPad software) using the following equation: Y=100/(1+((X/IC_{50})^Hillslope)). IC_{50} is defined as the inhibitory compound concentration that causes 50 % reduction in enzyme activity.

Incorporation of nucleotide analogues into RNA

DENV NS5 proteins were pre-incubated with the annealed P_{10}/T_{20} RNA (either the DENV2-specific combination to test multiple NTP incorporations: 6Fam-5'-AGUUGUUAGU-3'/5'-GUCCACGUAGACUAACAACU-3', or the non-specific combinations to study single nucleotide incorporations: Cy5-5'-GUCAUUCUCC-3'/5'-UAGCUUCUUAGGAGAAUG-3' AC for U incorporation or Cy5-5'-GUCAUUCUCC-3'/5'-UAGCUUCUUCGGAGAAUGAC-3' for G incorporation) in an assembly buffer containing 20 mM HEPES pH 7.5, 10% glycerol, 5 mM $MgCl_2$, and 5 mM DTT for 10 min at 30°C to create an active RNA elongation complex.
Multiple nucleotide incorporation was tested as follows. Reactions were started by adding AT-9010 with either all four NTPs or UTP, ATP, and CTP only. Final concentrations were 1 µM DENV NS5, 0.25 µM P_{10}/T_{20}, 100 µM of each NTP, and between 10 and 625 µM AT-9010 in a final buffer containing 20 mM HEPES pH 7.5, 15% glycerol, 5 mM $MgCl_2$, and 5 mM DTT.
Single nucleotide incorporation was tested as follows. Reactions were started by adding the corresponding NTP or analogue. Final concentrations were 1 µM NS5, 0.25 µM P_{10}/T_{20}, 100 µM NTP or analogue in a final buffer containing 20 mM HEPES pH 7.5, 15% glycerol, 5 mM $MgCl_2$, and 5 mM DTT. Reactions were quenched at designated time points in 3× volume FBD stop solution (formamide, 10 mM EDTA) and products were separated using sequencing gels of 20% acrylamide-bisacrylamide (19:1), 7 M urea with TBE buffer. RNA product bands were visualized using an Amersham™ Typhoon™ Biomolecular Imager (Cytiva). The intensity of each band was quantified using the ImageQuant software (Cytiva).

Crystallographic and modeling studies

The construct used for crystallization was the same as described in (Barral et al., 2013) with a cleavable Hexahistidine-Thioredoxin tag at the N-terminus of DENV3 NS5 coding sequence (amino acids region 1-277) in plasmid pMcox20A. Transformed E. coli T7 express cells (New England BioLabs) were grown at 37°C in LB medium supplemented with 100µg/mL ampicillin, to an OD_{600nm} of 0.6-0.8. Expression of the recombinant protein was induced by the addition of 100µM IPTG, and incubation was continued for a further 16h at 17°C. The cells were harvested by centrifugation at 8000g for 10 min at 4°C and stored at -80°C. The purification of the protein and the tag removal was performed in non-denaturing conditions as described in (Lantez et al., 2011) by using TALON® Superflow™ cobalt-based IMAC resin (Cytiva) for affinity chromatography. The final size exclusion chromatography step was performed using a Superdex S200 HiLoad 16/60 column with a buffer containing 20 mM HEPES pH 7.5, 200 mM NaCl, 10% glycerol and 2 mM DTT.
Crystallization conditions were adapted from (Lim et al., 2011) using the sitting-drop vapor diffusion method using crystallization buffer. All crystals were grown at 293.15 K, using a 1 : 2 ratio of protein (10 mg/mL) to precipitant solution (22% PEG 8000, 200 mM NaCl, 20 mM Tri-Sodium Citrate, 100 mM Tris pH 8.5). Crystals grew in 4 days and were soaked overnight with 1

mM of AT-9010 (final concentration). All crystals were cryo-protected with reservoir solution supplemented with 10% glycerol, and flash-frozen in liquid nitrogen at 100 K.
Data collection and diffraction data were collected at Soleil synchrotron. Original native data set was collected on Proxima2. Data set was processed and analysed with autoPROC toolbox (Vonrhein et al., 2011). Structure was solved by molecular replacement using PHASER (McCoy et al., 2007) and PDB 4CTK as a reference model. As previously observed the crystal Asymmetric Unit is composed of 2 chains, however the additional positive density blobs found in both chain is not equivalent. In one chain the extra density corresponds to a single molecule of SAM while in the other chain the positive density corresponds to one SAH and an AT-9010 molecules. The model was build using COOT (Emsley and Cowtan, 2004) and refined at 1.90 Å using PHENIX (Adams et al., 2010). The structure was checked and confirmed to have good stereochemistry according to MOLPROBITY (Chen et al., 2010). Data collection and refinement statistics are listed in Table 2. Structural analysis and figures was done using UCSF CHIMERA (Pettersen et al., 2004).

Table 1. Thermal shift analysis of four serotypes of DENV NS5 domains (MTase and RdRp) and full-length NS5 in the presence of GTP, the MTase inhibitor sinefungin, and the GTP analogue AT-9010.

	ΔT_m(°C)	DENV1	DENV2	DENV3	DENV4
MTase domain	---	---	---	---	---
	GTP	4.9 ± 0.1	5.7 ± 0.1	4.1 ± 0.1	4.0 ± 0.1
	Sinefungin	1.9 ± 0.1	2.9 ± 0.1	2.1 ± 0.1	2.8 ± 0.1
	AT-9010	3.8 ± 0.1	4.0 ± 0.1	3.5 ± 0.1	2.9 ± 0.1
RdRp domain	---	---	---	---	---
	GTP	4.5 ± 0.2	2.4 ± 0.1	2.2 ± 0.1	3.3 ± 0.1
	Sinefungin	-0.1 ± 0.1	-0.8 ± 0.1	0.0 ± 0.1	-0.1 ± 0.1
	AT-9010	5.6 ± 0.1	3.8 ± 0.1	3.2 ± 0.1	4.1 ± 0.1
NS5	---	---	---	---	---
	GTP	5.8 ± 0.1	5.7 ± 0.1	4.8 ± 0.1	3.9 ± 0.1
	Sinefungin	1.2 ± 0.1	1.4 ± 0.1	1.8 ± 0.1	1.4 ± 0.1
	AT-9010	6.1 ± 0.1	6.0 ± 0.1	5.0 ± 0.1	4.8 ± 0.1

Table 2. Crystallographic data and statistic for the DENV3 MTase/AT-9010 complex. Statistics for the highest-resolution shell are shown in parentheses.

Data processing	
	DENV3 MTase complexed with soaked **AT-9010**
Wavelength (Å)	0.9801
Space group	P 21 21 2
– a, b, c (Å)	61.26, 184.13, 52.15
– α, β, γ (°)	90.00, 90.00, 90.00
Resolution range (Å)	45.38 - 1.9 (1.97 - 1.9)
Total No. of reflections	583677 (27381)
No. of unique reflections	45354 (2284)
Completeness (%)	96.0 (97.8)
Multiplicity	12.9 (12.0)
I/σ(I)	18.5 (4.7)
R meas	0.093 (0.46)
R merge	0.089 (0.44)
R-pim	0.026 (0.13)
CC1/2	0.99 (0.94)
Wilson B-factor (Å)	24.6
Structure solution & refinement	
No. of reflections, working set	45347 (4551)
No. of reflections, test set (%)	2195 (217), 5.0 %
R_{work}	0.20 (0.27)
R_{free}	0.24 (0.33)
No. of non-H atoms	4589
-Protein	4162
-Ligand	65
-Water	362
R.M.S. deviations	Protein residues 516
-Bonds (Å)	0.013
-Angles (°)	1.21
Average B-factors (Å^2)	29.38
-Protein	28.68
-Ligand	41.71
-Water	35.22
Ramachandran Plot	
-Favoured (%)	98.04
-Allowed (%)	1.96
-Outliers (%)	0.00
Clashcore	5.21
PDB code	8BCR

Table 3. Preference for GTP over AT-9010, AT-9002, and of UTP over SOF-TP for the four serotypes of DENV NS5 in a single nucleotide incorporation assay. Incorporation rates (R_{NTP}) of individual 5'-triphosphate analogues were determined as described in 'Material and Methods', and preference for GTP over AT-9010 or AT-9002 and UTP over SOF-TP is reported as $P = R_{GTP}/R_{AT\text{-}9010}$, $P = R_{GTP}/R_{AT\text{-}9002}$, $P = R_{UTP}/R_{SOF\text{-}TP}$.

Preference	**DENV1**	**DENV2**	**DENV3**	**DENV4**
GTP > AT-9010	7.7 ± 0.5	9.3 ± 0.1	11.6 ± 1.3	11.2 ± 0.7
GTP > AT-9002	19.4 ± 2.8	18.2 ± 0.2	17.5 ± 1.5	41.5 ± 7.3
UTP > SOF-TP	> 100	> 100	> 100	> 100

Table 4. Conservation and identification of essential amino acids in contact with the phosphates of the incoming nucleotide in viral RdRps HCV NS5b, DENV2 NS5, and SARS-CoV2 nsp12. The Protein Data Bank (PDB, www.rcsb.org/) codes used are 4WTG for HCV, 5ZQK for DENV2, and 7ED5 for SARS-CoV-2 nsp12.

Virus Contacts at	HCV	DENV2	SARS-CoV-2
α- phosphate	R158	R472	R555
β- phosphate	R48	-	R553*
γ- phosphate	K51, R222, H223	K358	K551*
Motif D residue	-	K689	K798
Other features	H223 corresponds to SARS Motif D K621	-	K551 and R553 CoV unique Motif F K621 corresponds to HCV H223

Figure legends

Figure 1. Cartoon representing the flavivirus NS5 bi-functional MTase and RdRp enzyme, with potentially inhibited enzyme activities. The NS5 protein is represented from N- to C-terminus, with the conserved sequence motifs A-to-G reported along NS5. Structures of the two guanosine analogues AT-9010 and AT-9002 along with sofosbuvir-5'-triphosphate studied here are represented below the NS5 protein, together with relevant enzyme activities.

Figure 2. *In vitro* activity of AT-281 against DENV1–4 reporters. Serial dilutions of AT-281 and virus modified with a nano luciferase reporter (MOIs of 0.1 for DENV1, DENV2 and DENV3; MOI of 0.001 for DENV4) were added to 96-well plates seeded with $1x10^4$ Huh-7 cells per well. After incubation for 48 h, cells were washed and EC_{50} values were determined by luciferase activities. Data were normalized to DMSO treatment samples which were set to 100%. Error bars represent mean ± SD. Results are representative of three independent experiments with each one analyzed in triplicate

Figure 3. Thermal stability shift experiments showing the stabilizing effect of GTP, AT-9010 or Sinefungin on the DENV serotype 1 to 4 MTases, RdRp and NS5 protein. The values are reported in delta Tm (ΔTm (°C)) using the proteins alone as a baseline.

Figure 4. Functional and structural basis of inhibition of the DENV MTase by AT-9010. A) Dose-response determination using a filter binding assay (FBA). Reactions were incubated at 30°C during 30 min and the enzymatic activity was determined using FBA as described in the Material and Methods section.
B) Overall structure of DENV3 MTase represented in blue ribbons complexed with SAH and AT-9010 represented as sticks. AT-9010 is located at the Cap binding site. C) Omit map at 1σ of the AT-9010. D) Close view of the AT-9010 binding in the GTP binding site. Interacting residues are labelled and shown as sticks. E) Structural comparison of the position of a GTP (pink) and AT9010 (light brown) bound in the Cap binding site. Cap binding site structures belong to DENV2 (in pink - PDB 2P1D)) and DENV3 (in light brown) MTases and are conserved.

Figure 5. AT-9010 incorporation and inhibition during *de novo* initiation reaction by DENV2 NS5.
A) PAGE analysis of *de novo* initiation reactions using a 10-nt template corresponding to the 3' end of the genome (AACAGGUUCU) resulting in the formation of pppApG, pppApGpA and pppApGpApA over time. GTP was either replaced by AT9010 or AT9010 was added in increasing concentrations to the reactions containing 100 µM GTP. B) Inhibition of *de novo* product formation by AT-9010 based on the quantification of product bands of four (Mn) or two (Mg) independent reactions; the corresponding IC_{50} values are given.

Figure 6. Discrimination of AT-9010, AT-9002, and SOF-TP by DENV4 NS5 against their natural nucleotide competitors GTP, GTP, and UTP, respectively. A) Primer extension assay allowing multiple insertion of ATP, CTP, and UTP in the absence, the presence of GTP; or the presence of a mixture of GTP and AT-9010 as indicated below the auto-fluorograph of the gel. The nature of regular inserted Watson-Crick nucleotides is indicated in the right of the gel. B)

Primer extension assay allowing incorporation of a single nucleotide GTP, AT-9010, AT-9002, UTP or SOF-TP. Quantification of band-products greater in size than the G/AT (+5) band relative to the G/AT (+5) band were measured using a Typhoon FluorImager, and allowed to calculate the discrimination value. The incorporation rates for AT-901 and AT9002 are compared to that of GTP, and to that of UTP in the case of SOF-TP. Similar results for other DENV1 - 3 NS5s are reported in Supplementary Figures 1-3.

Figure 7. Comparison of RdRp active sites of HCV, DENV, and SARS-CoV-2 RdRps. A) The HCV NS5b (PDB: 4WTG) was superimposed onto DENV2 NS5 (PDB: 5ZQK). Amino acid side chains in the vicinity of Motif D and as well as the nucleotide phosphates are represented. The Motif D region is represented in orange for HCV. Metal A and B are represented with purple and orange spheres in the vicinity of the Motif A and C catalytic aspartates, respectively. Unique amino acid labels discussed in the text are represented in red. B) The SARS-CoV-2 nsp12 structure (PDB:) was superimposed onto DENV NS5 (PDB: 5ZQK). The Motif D Lysine amino acid side chain is shown in red for SARS-CoV-2 nsp12 and grey for DENV NS5, respectively. Essential amino acids contacting the nucleotide phosphates are represented. A unique Metal B is represented with a green sphere in the vicinity of the Motif A and C catalytic aspartates, respectively (Shannon et al., 2022). Unique amino acid labels discussed in the text are represented in red.

Scheme 1. Representation of the kinetic pathway of NTP (or NA-TP) incorporation into RNA. The cartoon starts on the left with the viral RdRp bound to its RNA substrate. The main individual steps are numbered 1 to 5. The RdRp Motifs A to G involved in this scheme are indicated on top, as reported in the literature and reviewed in Selisko et al, 2018. E: enzyme; R_n: RNA of n nucleotides in length; PPi: inorganic pyrophosphate. *indicates an enzyme undergoing a conformational change.

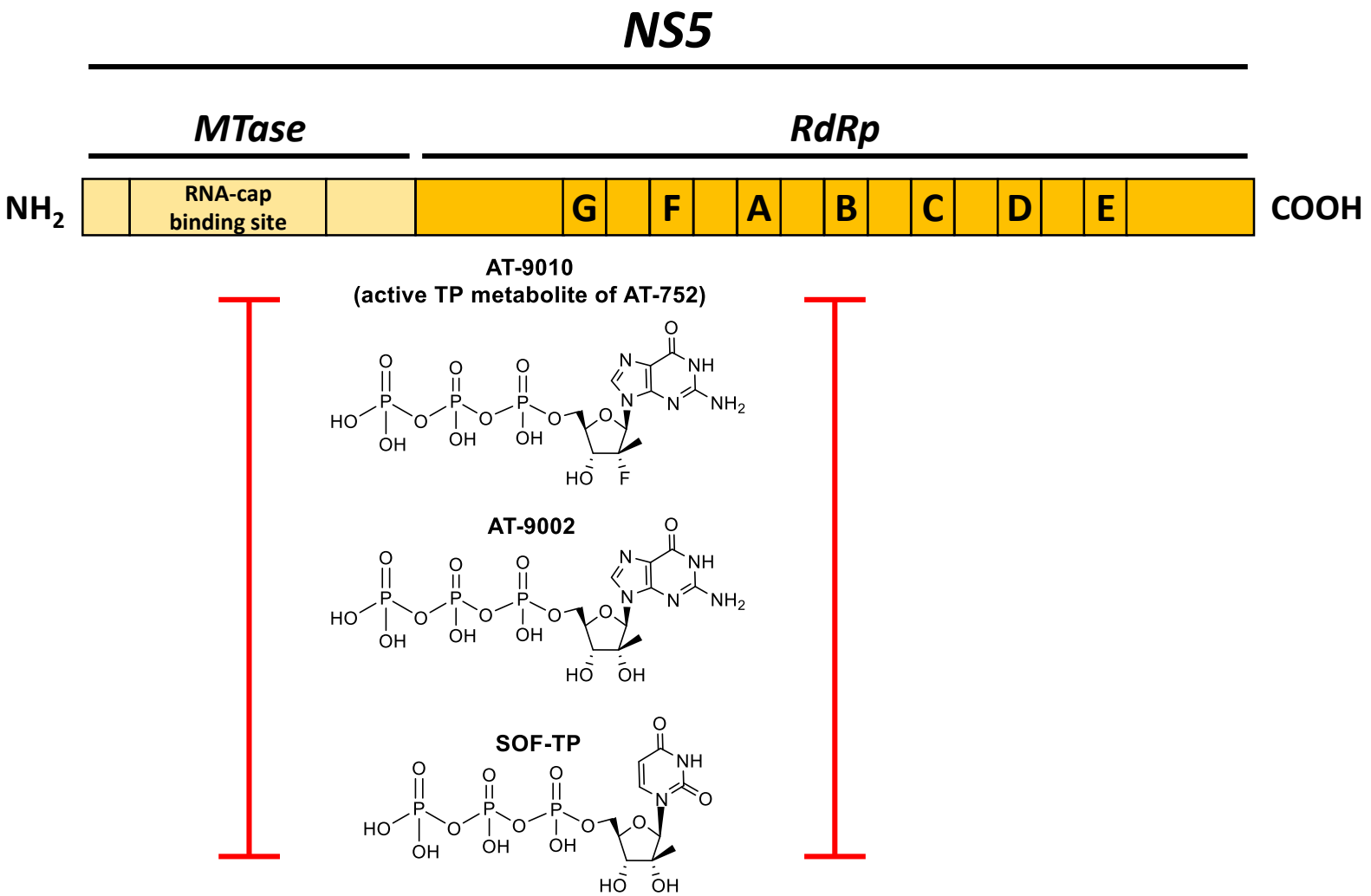

NS5
MTase
RdRp
NH2
RNA-cap binding site
G
F
A
B
C
D
E
COOH
AT-9010
(active TP metabolite of AT-752)
AT-9002
SOF-TP
GTP/RNA cap binding
2'-O MTase activity
N7 MTase Activity
pppApG primer synthesis (de novo)
RNA synthesis (elongation)

Figure 1

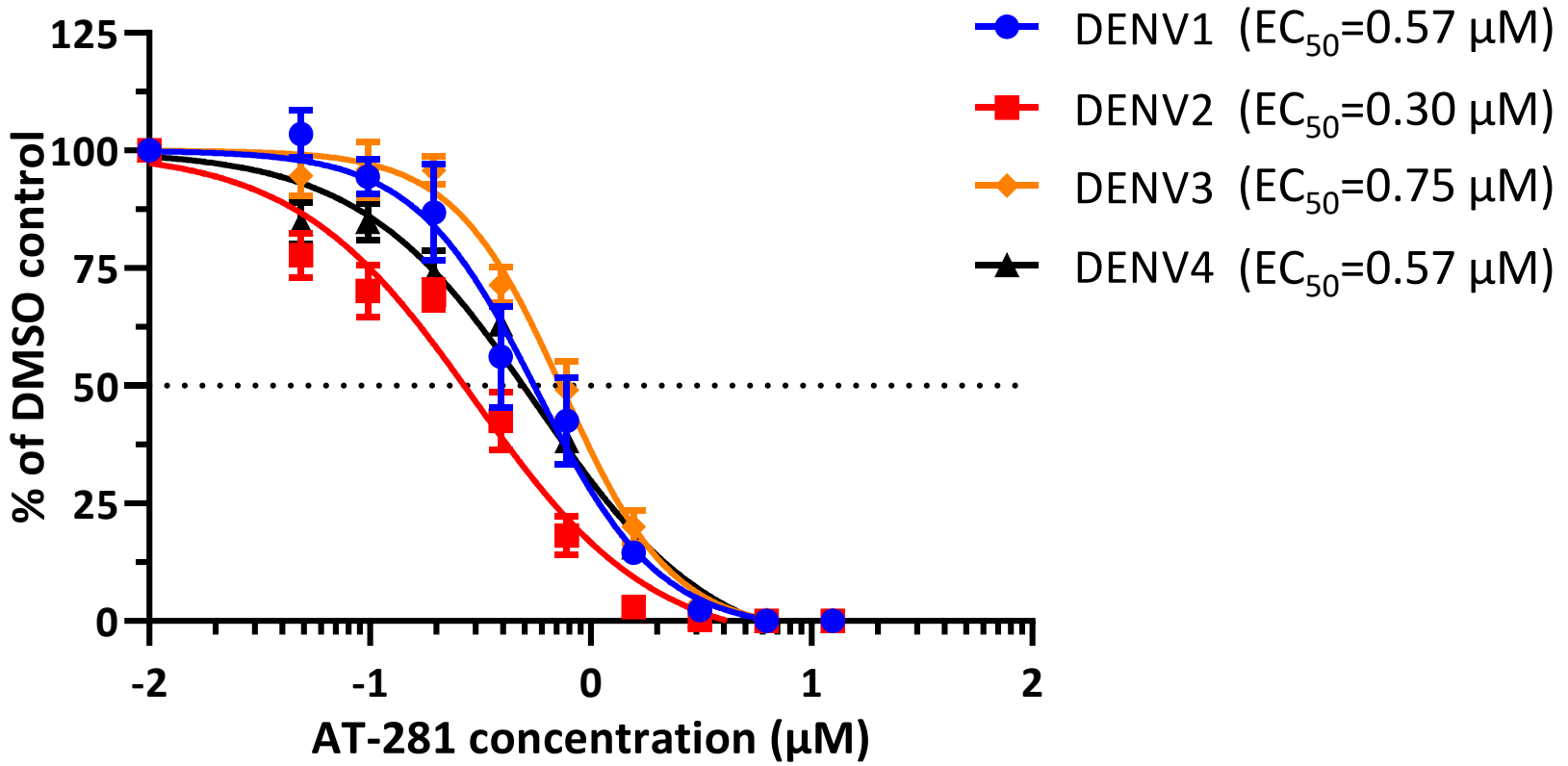
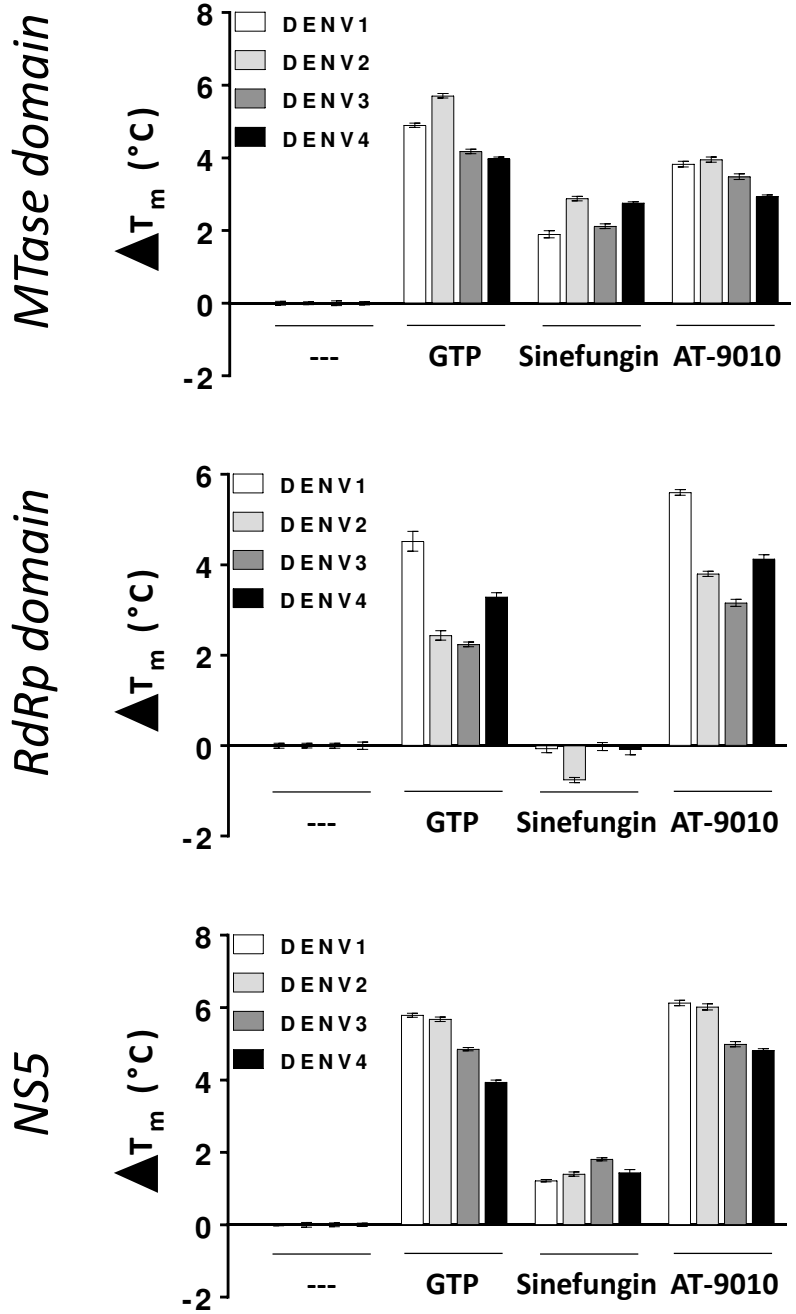

125
100
75
50
25
0
% of DMSO control
-2
-1
0
1
2
AT-281 concentration (µM)
DENV1 (EC50=0.57 µM)
DENV2 (EC50=0.30 µM)
DENV3 (EC50=0.75 µM)
DENV4 (EC50=0.57 µM)

Figure 2


MTase domain
RdRp domain
NS5
DENV1
DENV2
DENV3
DENV4
▲Tm (°C)
8
6
4
2
0
-2

GTP
Sinefungin
AT-9010

Figure 3

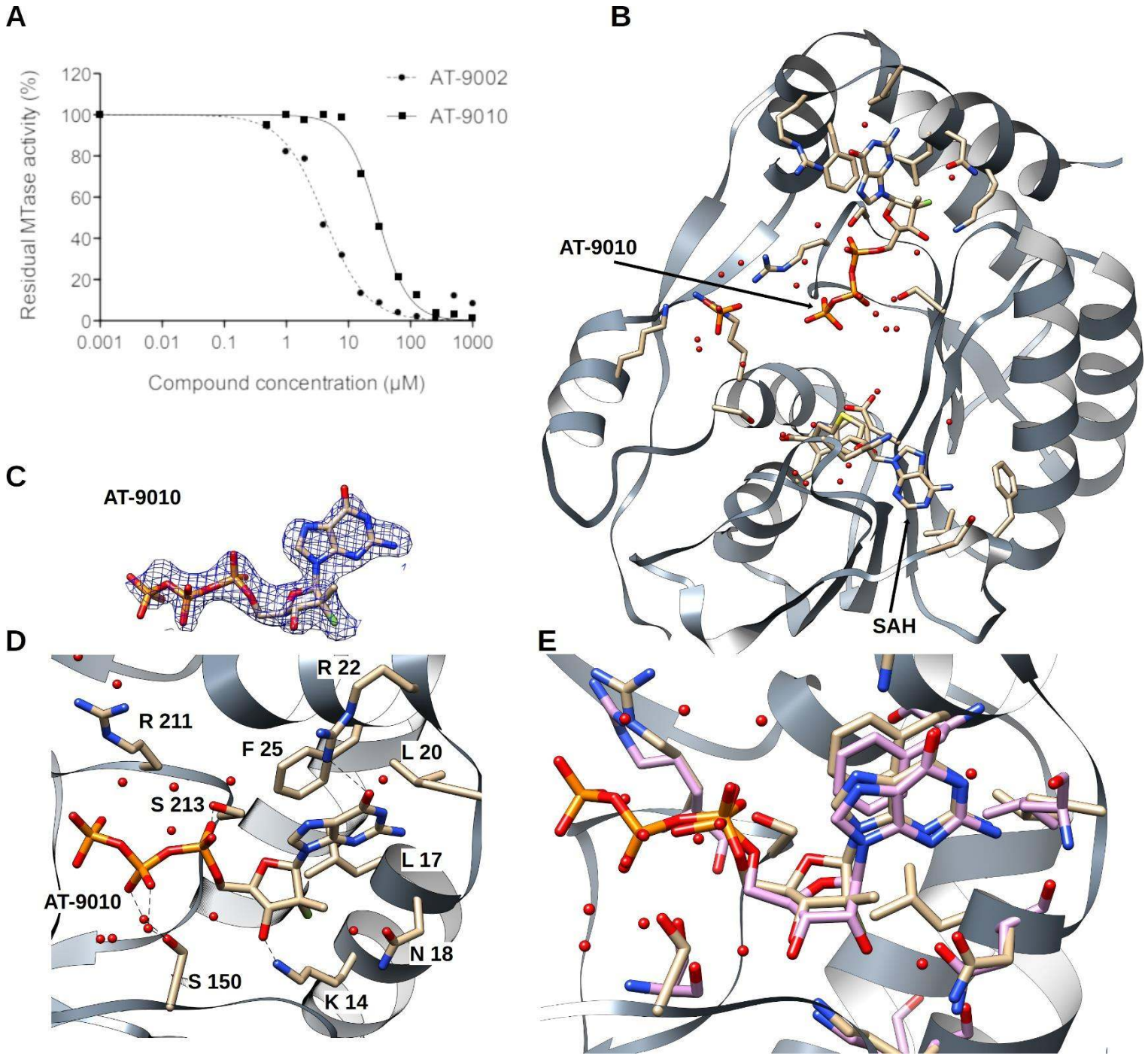

A
Residual MTase activity (%)
120
100
80
60
40
20
0
0.001
0.01
0.1
1
10
100
1000
Compound concentration (μM)
AT-9002
AT-9010
B
AT-9010
SAH
C
AT-9010
D
R 22
R 211
F 25
L 20
S 213
L 17
AT-9010
N 18
S 150
K 14
E

Figure 4

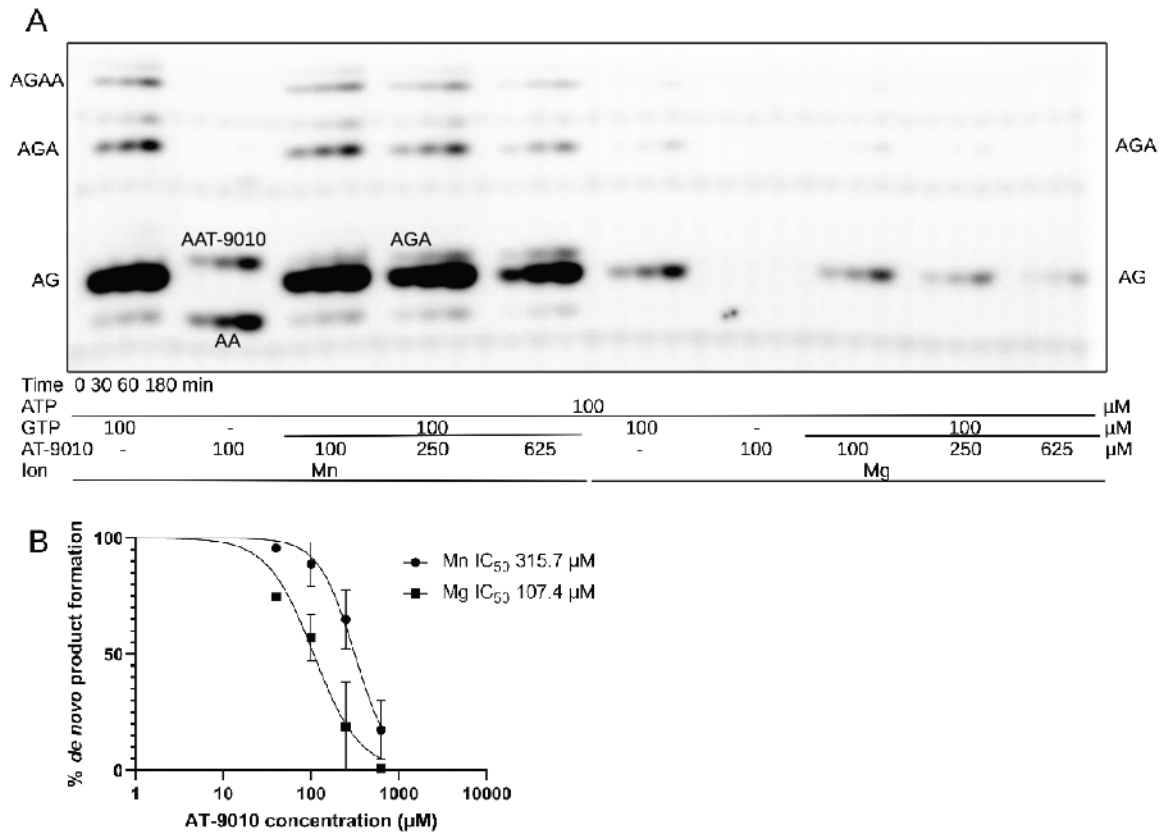

A
AGAA
AGA
AG
AAT-9010
AGA
AA
AGA
AG
Time 0 30 60 180 min
ATP 100 µM
GTP 100 - 100 100 - 100 µM
AT-9010 - 100 100 250 625 - 100 100 250 625 µM
Ion Mn Mg
B
% de novo product formation
100
50
0
1
10
100
1000
10000
AT-9010 concentration (µM)
Mn IC50 315.7 µM
Mg IC50 107.4 µM

Figure 5

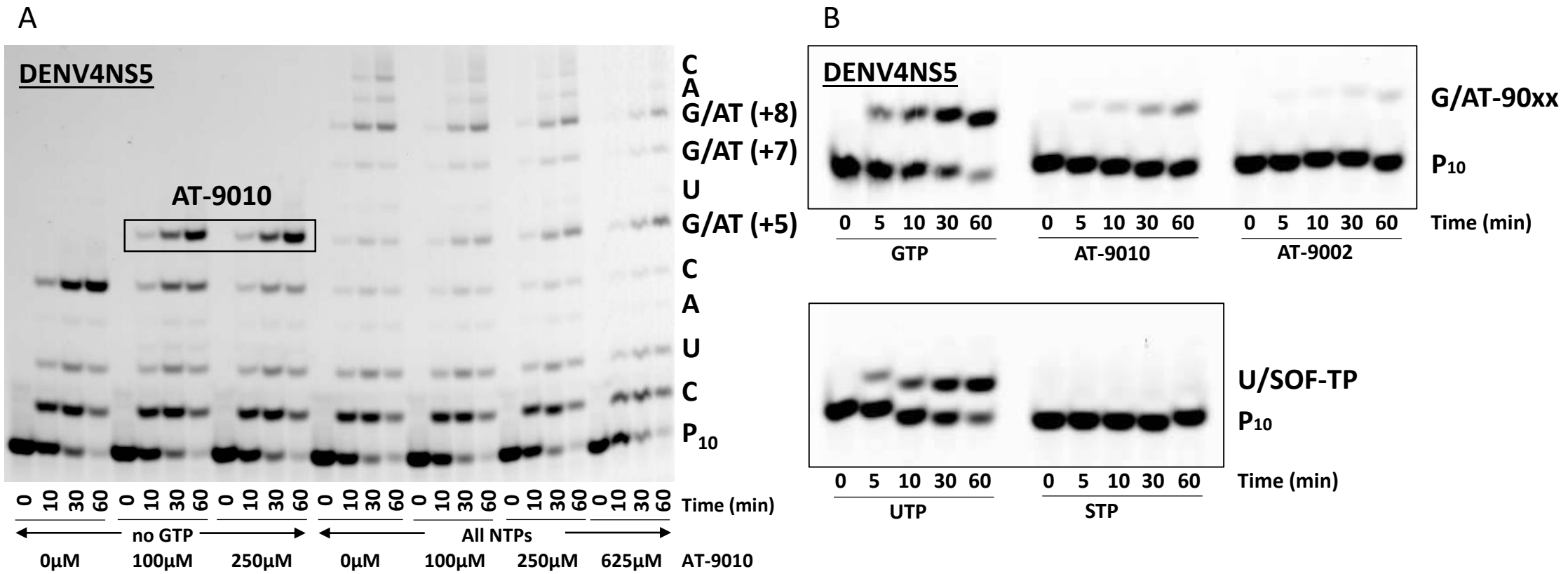
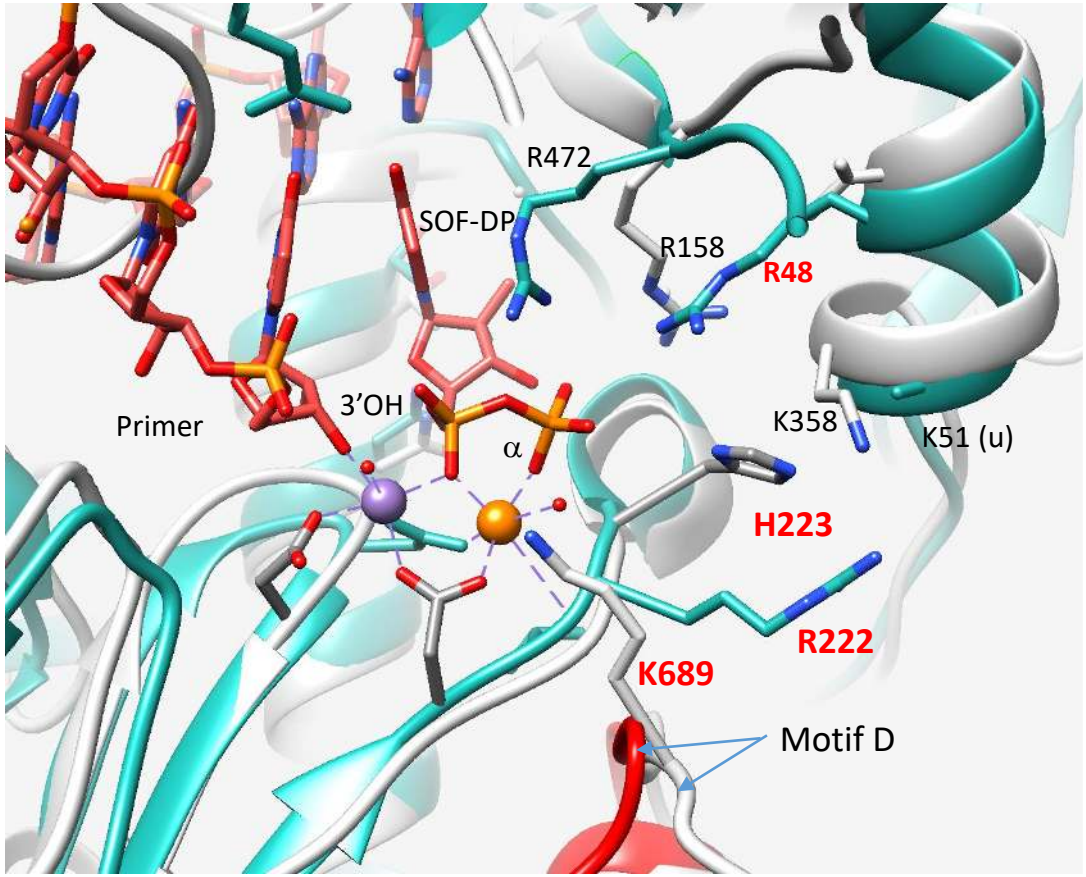

A
DENV4NS5
AT-9010
C
A
G/AT (+8)
G/AT (+7)
U
G/AT (+5)
C
A
U
C
P10
0 10 30 60 0 10 30 60 0 10 30 60 0 10 30 60 0 10 30 60 0 10 30 60 0 10 30 60
Time (min)
no GTP
All NTPs
0μM 100μM 250μM 0μM 100μM 250μM 625μM AT-9010
B
DENV4NS5
G/AT-90xx
P10
0 5 10 30 60 0 5 10 30 60 0 5 10 30 60
Time (min)
GTP AT-9010 AT-9002
U/SOF-TP
P10
0 5 10 30 60 0 5 10 30 60
Time (min)
UTP STP

Figure 6

A HCV - DENV



B DENV-SARS-CoV2

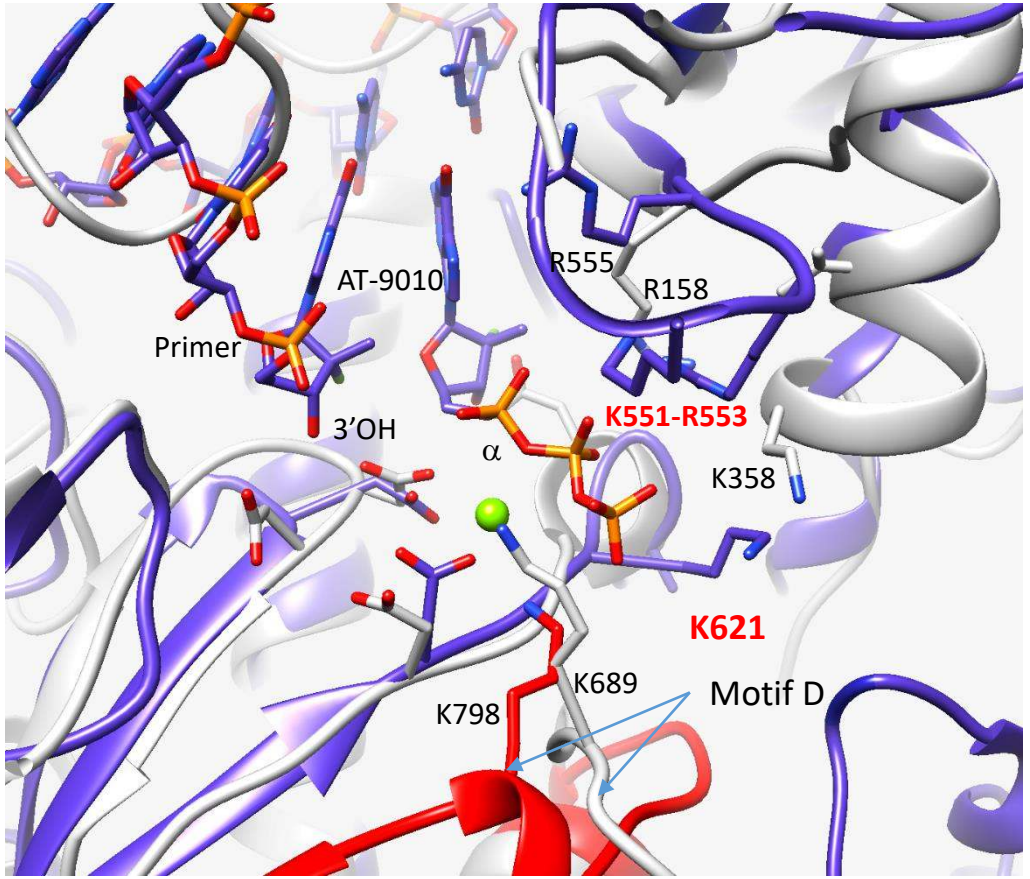
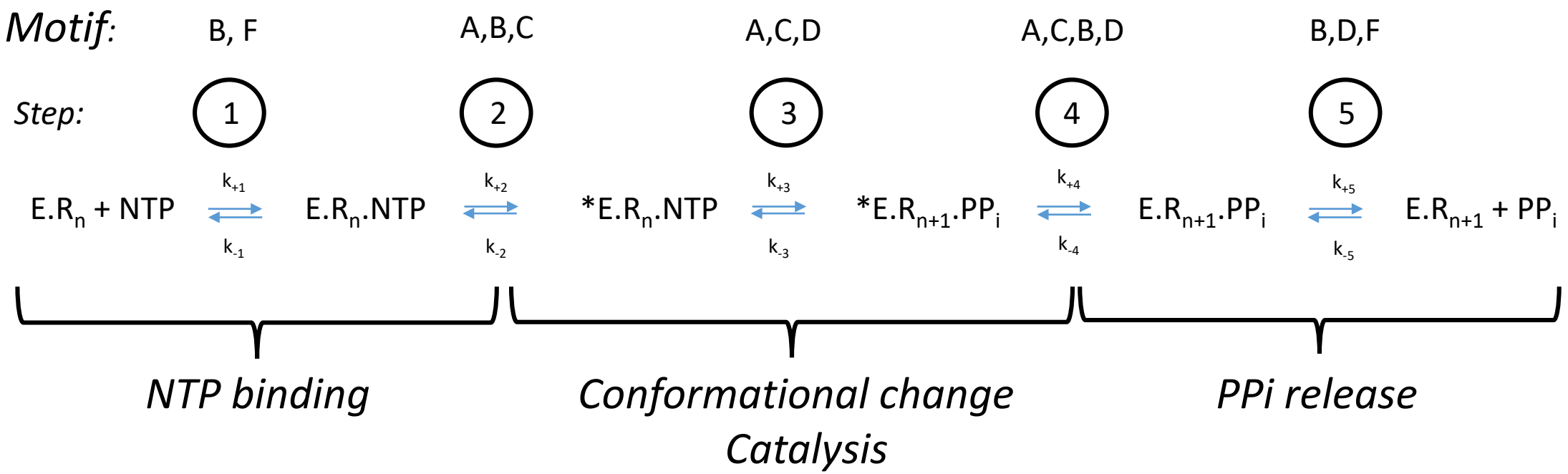


Figure 7


Motif:
B, F
A,B,C
A,C,D
A,C,B,D
B,D,F
Step:
1
2
3
4
5
E.Rn + NTP
kf1
kr1
E.Rn.NTP
kf2
kr2
*E.Rn.NTP
kf3
kr3
*E.Rn+1.PPi
kf4
kr4
E.Rn+1.PPi
kf5
kr5
E.Rn+1 + PPi
NTP binding
Conformational change
Catalysis
PPi release

Scheme 1

A3 : A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome

11186–11198 *Nucleic Acids Research, 2022, Vol. 50, No. 19* *Published online 21 October 2022*
https://doi.org/10.1093/nar/gkac876

A second type of N7-guanine RNA cap methyltransferase in an unusual locus of a large RNA virus genome

Ashleigh Shannon [1,†], Bhawna Sama [1,†], Pierre Gauffre [1], Théo Guez [2], Françoise Debart [2], Jean-Jacques Vasseur [2], Etienne Decroly [1], Bruno Canard [1,3] and François Ferron [1,3,*]

[1]Aix-Marseille Université and CNRS, Laboratoire Architecture et Fonction des Macromolécules Biologiques, UMR 7257, 13009, Marseille, France, [2]IBMM, University of Montpellier, CNRS, ENSCM, Montpellier, France and [3]European Virus Bioinformatics Center, Leutragraben 1, 07743 Jena, Germany

Received September 12, 2022; Editorial Decision September 16, 2022; Accepted September 29, 2022

ABSTRACT

The order *Nidovirales* is a diverse group of (+)RNA viruses, with a common genome organization and conserved set of replicative and editing enzymes. In particular, RNA methyltransferases play a central role in mRNA stability and immune escape. However, their presence and distribution in different *Nidovirales* families is not homogeneous. In *Coronaviridae,* the best characterized family, two distinct methytransferases perform methylation of the N7-guanine and 2′-OH of the RNA-cap to generate a cap-1 structure (m7GpppNm). The genes of both of these enzymes are located in the ORF1b genomic region. While 2′-*O*-MTases can be identified for most other families based on conservation of both sequence motifs and genetic loci, identification of the N7-guanine methyltransferase has proved more challenging. Recently, we identified a putative N7-MTase domain in the ORF1a region (N7-MT-1a) of certain members of the large genome *Tobaniviridae* family. Here, we demonstrate that this domain indeed harbors N7-specific methyltransferase activity. We present its structure as the first N7-specific Rossmann-fold (RF) MTase identified for (+)RNA viruses, making it remarkably different from that of the known *Coronaviridae* ORF1b N7-MTase gene. We discuss the evolutionary implications of such an appearance in this unexpected location in the genome, which introduces a split-off in the classification of *Tobaniviridae*.

INTRODUCTION

The *Nidovirales* order encompasses a remarkably diverse group of enveloped, positive-sense RNA viruses with genome sizes ranging from ∼11 to 41 kb. The order is currently classified into 14 families, with members infecting a range of hosts (eight vertebrate and six invertebrate families) (1,2). At one end of the size spectrum are the small-genome *Arteriviridae* members, with genomic RNAs ranging in size from ∼11 to 15 kb. At the other end of the spectrum, *Coronaviridae* genomes reach >32 kb in size, topped only by a few members of the (currently) under populated families such as the *Mononiviridae* (>41 kb). Despite these significant size-differences, these viruses are classified together based on their relatively conserved genome organization and mRNA transcriptional scheme. In the majority of nidoviruses, the first two-thirds of the genome contains two large overlapping open reading frames (ORF1a and 1b). Translation of the ORFs yields two polyproteins (pp1a and pp1ab) containing between 12 and 16 non-structural proteins involved in viral replication and transcription. The 3′ end of the genome codes for a nested set of subgenomic mRNAs (sg mRNAs) directing the synthesis of structural and accessory proteins. The name of the order comes from *Nido* (*latin*: nest), which refers to this common 'nested' gene expression pattern allowing differential expression levels of non-structural versus structural and accessory proteins (3).

Due to their unusual size, large *Nidovirales* RNA genomes and transcripts need both genetic and biophysical stability in order to faithfully transmit and express their genetic content. Genetic stability is generally ensured by the replication/transcription complex (RTC), composed of a core RNA-dependent RNA polymerase (nsp12) flanked with processivity factors nsp7 and nsp8 (4). The RTC also recruits nsp14, a 3'-to-5′ exonuclease (ExoN) which corrects

*To whom correspondence should be addressed. Tel: +33 4 91828628; Email: francois.ferron@univ-amu.fr
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

© The Author(s) 2022. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

replication errors (5). This mismatch repair system, unique to large genome nidoviruses, ensures accurate replication and transfer of genomic information, and is thus essential for genetic stability (6).

Many +RNA viruses have additionally evolved mechanisms to protect their RNAs by the incorporation of a 5′ cap. As is the case for eukaryotic mRNAs, this 5′ cap allows recognition and translation of viral mRNA by host ribosomes, and additionally protects transcripts from 5′-to-3′ exonucleolytic degradation. Viral RNA capping is thus essential not only for genome stability and translation, but also for innate-immune escape since the viral RNA cap cannot be distinguished from its cellular counterpart (7–11).

For the nidoviruses, the presence of an RNA cap has been reported for three distinct viruses within the order (12–15), but high-resolution structural analysis of nidovirus cap structures is still lacking. Further indication of the presence of a 5′ type-1 cap is indirectly provided by the identification of key enzymes required in the conventional, eukaryotic capping pathway (16). Among these enzymes are i) an RNA triphosphatase (RTPase), which is responsible for hydrolysis of the 5′-triphosphate of the viral RNA into a 5′-diphosphate, ii) a guanylyltransferase (GTase), which covalently transfers a GMP moiety (originating from GTP) to the 5′-diphosphate, generating a 5′5′ triphosphate bridge (GpppN) and iii) one or two methyltransferases (MTases), that subsequently methylate the cap using *S*-adenosylmethionine (AdoMet or SAM) as the methyl donor, converting it to *S*-adenosyl-homocysteine (AdoHcy or SAH). Methylation generally occurs in a sequential order, with the N7-guanine methylated first for generation of the ^{m}GpppN-RNA (type-0 cap), followed by the 2′-hydroxyl position of the first transcribed nucleotide, yielding ^{m}GpppN$_{m}$-RNA (type-1 cap).

Most large RNA viruses (>17 kb) of positive genome polarity generally encode at least one MTase gene, and it has been proposed that the correlation between RNA genome size and presence of an MTase reflects a beneficial coevolution (17). Five distinct structural classes of enzymes that catalyze SAM-dependent methylation have been described, with most viral MTases (along with most DNA methyltransferases), falling into the largest, Class I group (18). Class I MTases are characterized by the presence of a Rossmann-fold (RF) domain, a super-secondary structure adopted by a diverse group of dinucleotide binding enzymes. It is generally characterized by an alternating βαβ architecutre, made up of of a central β-sheet with up to seven β-strands, surrounded by α-helices. The glycine-rich motif I (GxGxG) follows the first β-strand, and is involved in SAM binding.

For *Coronaviridae,* most of the enzymes required in this capping pathway have now been identified, including the N7-guanine and 2′-*O* MTases, which reside in nsp14 and nsp16, respectively (19–21). The RTPase resides in nsp13 (22) while more recently, the GTase activity was attributed to the nidovirus RdRp-associated nucleotidyl transferase (NiRAN) domain, located at the N-terminus of nsp12 (23–26). However, the precise mechanism of action of this enzyme, including the requirement for viral cofactors and specific substrate requirements, is still somewhat speculative. For most other nidoviral families, the capping pathway and enzymes have been poorly characterized. Capping enzymes which are homologous to those in the *C*oronaviridae genome have been identified for several families (19–22,25,27,28). In contrast, for the the small-genome *Arteriviridae* family, none of the enzymes required in the capping pathway appear to be present, suggesting that the presence of an RNA cap is not conserved throughout the entire order.

Recently, we performed a large-scale genomic analysis of the order *Nidovirales* in order to clarify the presence of MTase domains across different families (29). The 2′*O* MTase can be clearly identified for most of these viruses based on a conserved genomic location (encoded at the end of ORF1b), and can be further characterized by their RF structure and K–D–K–E catalytic tetrad. Unlike 2′*O*-MTases, RF N7-MTases lack a specific catalytic signature sequence. As such, determining their methylation-site specificity is more challenging. Furthermore, the N7-MTase domain residing in nsp14 of CoVs shows very limited sequence conservation with other N7-MTases, and lacks the βαβ architecture characteristic of RF enzymes. While nsp14-like, non-RF N7-MTases can also be identified for several other nidoviruses, this enzyme appears to be lacking entirely for many of the families. Interestingly however, we and others have identified a putative N7-MTase in pp1a (N7-MT-1a) of several members of the large-genome *Tobaniviridae* family (29–31).

Here we confirm that this domain is indeed the missing N7-MTase. We present its crystal structure at 1.7 Å resolution, revealing that it adopts a canonical RF secondary structure. This makes it the first N7-specific RF MTase identified for any positive-stranded RNA virus to date. Its closest homologues are the N7-MTases of the microsporidian parasite *Encephalitozoon cuniculi*, humans and two large DNA viruses—despite extremely limited sequence identity with any of these MTases. In regards to the genomic position, sequence and predicted secondary structure, this domain is remarkably different from the known nidovirus nsp14 N7-MTase gene, illustrating an outstanding genetic diversity within a single order of viruses. Intriguingly, this domain only appears to be present in non-mammalian infecting (fish or reptile host) tobanivirus members (29). The characterization of this new N7-MTase will not only aid with further clarification within this viral order, but will assist in the identification and assignment of MTases across all domains of life.

MATERIALS AND METHODS

Plasmid construction

An *Escherichia coli* codon optimized sequence for the region spanning Met^{1333} to Ile^{1668} of White Bream virus (WBV, NCBI Reference YP_803214.1) was synthesized by GenScript, with EcoRI and HindIII restriction enzyme sequences flanking the 5′ and 3′ ends respectively. The synthesized gene was subcloned into a pQE30 expression vector, under the control of a T5 promoter, incorporating a 5′ hexahistidine sequence. Deletion PCR mutagenesis with outward facing primers was subsequently performed us-

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

ing Q5 site-directed mutagenesis (New England Biolabs), to delete regions from the 5′ end region. Final recombinant plasmids contained domains starting from Met^{1333}, Thr^{1342}, Thr^{1350}, Ser^{1366}, Ala^{1358}, Asp^{1374} or Ser^{1389}, spanning to Ile^{1668}. A final C-terminal truncation was additionally performed for expression of a 260 amino acid (∼30 kDa) domain spanning from Asp^{1374}–Asp^{1633}. This construct was subsequently used for site-directed mutagenesis using outwards facing, overlapping primers and High Fidelity Phusion Polymerase (NEB).

Protein expression and purification

Protein was expressed in C2523 (pLacI) *E. coli* overnight at 18°C from cultures induced with 0.4mM isopropyl-β-D-thiogalactopyranoside and 2% ethanol once the OD_{600} reaction 0.8. Bacterial pellets were resuspended in lysis buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 2 mM BME, 1 mg/ml lysozyme, 20 μg/ml DNase, 1 mM PMSF), incubated for 30 min before lysis via sonication (amplitude 40%, 2 s on, 2 s off for 1 min). The soluble lysate was purified with affinity chromatography using HisPur Cobalt resin (Thermo Scientific). The bound lysate was washed with 2 column volumes wash buffer (20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 20 mM imidazole, 2 mM BME) and eluted with 150 mM imidazole (20 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 150 mM imidazole, 2 mM BME). The purified protein was subsequently concentrated and loaded onto a Superdex 200 gel filtration column (GE Healthcare), eluted in gel filtration buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 10% glycerol, 2.5 mM TCEP).

Synthesis of RNA substrates

RNA synthesis was previously described in (32). RNA sequences were chemically synthesized on solid support using an ABI 394 oligonucleotides synthesizer. After RNA elongation with 2′-*O*-pivaloyloxymethyl phosphoramidite ribonucleotides (33) and 2′-*O*-methyl phosphoramidite ribonucleotides (Chemgenes, USA), the 5′-hydroxyl group was phosphorylated and the resulting H-phosphonate derivative was oxidized and activated into a phosphoroimidazolidate derivative (34) to react with guanosine diphosphate yielding Gppp-RNA (35). After deprotection and release from the solid support, GpppRNAs were purified by IEX-HPLC and they were characterized by MALDI-TOF spectrometry. N7-methylation of the purified GpppRNA was performed enzymatically using N7-hMTase to give mGppp-RNA (35). All sequences used for analysis are listed in Supplementary Table S1.

Filter binding assays for methylation activity

Methylation activity was performed using filter binding assays, with reactions prepared in 96-well plates. For standard assays, final buffer conditions were 40 mM Tris pH 8, 1 mM DTT, 0.7 μM RNA, 1.9 μM cold AdoMet, supplemented with 0.1 μM ^{3}H-AdoMet. Initial reactions to assess potential activity were run with final enzyme concentrations ranging from 10 nM – 500 nM, with subsequent assays run with 10–500 nM enzyme, as stated in figure legends. Reactions were quenched at various time points by diluting five-fold in a solution of 200 μM AdoHcy, then further diluted to a 300 ul total volume in ice-cold H2O. A Filtermat Harvester (Packard Instruments) was used to transfer the reactions to 8 × 12 filtermats (GF/C, Perkin Elmer) coated with a 0.3% (w/v) solution of polyethylenimine (PEI, Sigma-Aldrich) for RNA binding. Filtermats were washed twice with 10 mM ammonium formate pH 8.0, twice with water and once with ethanol, prior to soaking in liquid scintillation fluid. RNA methylated with ^{3}H-AdoMet was measured using a Wallace MicroBeta TriLux Liquid Scintillation Counter (Perking Elmer).

Thin layer chromatography (TLC)

Triphosphorylated RNA (pppG-RNA_{12} or pppA-RNA_{12}, 10 μM) were capped using the vaccinia virus capping enzyme (New England Biolabs) in the presence of 2 μCi α-^{32}P-GTP (Perkin Elmer) to yield radiolabeled G*pppN capped RNA (where * represents the radioactive α-^{32}P). Duplicate reactions were run in the presence of 0.1 mM AdoMet, for creation of N7-guanine methylated cap controls (^{m}G*ppppN). Additional reactions were additionally prepared using $pppN_m$-RNA_{12} (pre-methylated at the 2′*O* position), in the presence or absence of AdoMet, for generation of G*$pppN_m$ and ^{m}G*$pppN_m$ cap controls. Capped RNA was incubated with 1 μM (final concentration) of the purified WB MTase enzyme in a final reaction containing 40 mM Tris, pH 8, 40 μM AdoMet, and 1 μM capped RNA. RNA was digested with Nuclease P1 (New England Biolabs) for 3–4 h at 37°C, and loaded onto polyethylenimine cellulose TLC sheets (Macherey Nagel), along with cap controls. Samples were separated using 0.65 M LiCl as mobile phase. TLC sheets were exposed and visualized using using an Amersham Typhoon BiomolecularImager (GE Healthcare).

Crystallization and structure determination

Crystallization conditions were initially identified using the sitting-drop vapor diffusion method using INDEX and PEGII 96 reagent screens (Hampton Research). Conditions were optimized for the WBV 260 construct (WBV-260, Asp^{1374}–Asp^{1633}), mixed with AdoMet (New England Biosciences) at a 1:1.2 ratio prior to crystallization. All crystals were grown at 293.15 K, using a 1:1 ratio of protein (8 mg/mL) to precipitant solution (0.1 M Tris pH 7.7, 20.3% PEG 3350, 0.2 M MgCl and 10% isopropanol), and grew in around 5 days as long needles/urchins. The presence of isopropanol drastically improved the size and diffraction of crystals. In our attempt to phase, several crystals were soaked in Tantalum Bromide and harvested while green, according to manufacturer protocol (Jena Bioscience). All crystals were cryo- protected with reservoir solution supplemented with 20% PEG400, and flash-frozen in liquid nitrogen at 100 K. Data collection and diffraction data were collected at Soleil synchrotron, from several crystals. Original

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

Table 1. Data collection and refinement statistics

	White Bream N7-MTase Proxima 2 | PDB: 7Z05	**White Bream N7-MTase** Proxima 1 | PDB : 7Z2J
Wavelength	0.992	1.255
Resolution range	44.93–2.33 (2.41–2.33)	39.07–1.66 (1.72–1.66)
Space group	*C* 1 2 1	*P* 1 21 1
Unit cell	105.22, 49.787, 53.93 90.00, 97.72, 90.00	41.58, 49.02, 53.76 90.00, 110.04, 90.00
Total reflections	80 601 (7962)	159 906 (14 462)
Unique reflections	11 969 (1168)	24 131 (2383)
Multiplicity	6.7 (6.8)	6.6 (6.1)
Completeness (%)	99.62 (99.57)	99.81 (99.04)
Mean *I*/sigma(*I*)	11.93 (2.28)	16.16 (2.41)
Wilson *B*-factor	39.11	30.35
***R*-merge**	0.115 (1.08)	0.056 (0.64)
***R*-meas**	0.12 (1.17)	0.061 (0.70)
***R*-pim**	0.04817 (0.44)	0.02 (0.28)
CC1/2	0.99 (0.874)	0.99 (0.91)
CC*	0.99 (0.966)	1 (0.97)
Reflections used in refinement	11942 (1164)	24129 (2383)
Reflections used for *R*-free	582 (52)	1168 (121)
***R*-work**	0.23 (0.36)	0.19 (0.38)
***R*-free**	0.24 (0.35)	0.20 (0.42)
CC(work)	0.94 (0.85)	0.94 (0.32)
CC(free)	0.94 (0.86)	0.93 (0.37)
Number of non-hydrogen atoms	2013	1835
Macromolecules	1924	1780
Ligands	15	0
Solvent	74	55
Protein residues	240	228
RMS(bonds)	0.012	0.014
RMS(angles)	1.64	1.93
Ramachandran favored (%)	92.83	97.33
Ramachandran allowed (%)	6.33	2.22
Ramachandran outliers (%)	0.84	0.44
Rotamer outliers (%)	6.91	3.03
Clashscore	12.40	1.13
Average *B*-factor	55.43	38.89
Macromolecules	55.55	38.58
Ligands	60.03	
Solvent	51.57	49.14

Statistics for the highest-resolution shell are shown in parentheses.

native and Sulfure SAD data set were collected on Proxima2, under the assumption that we could use experimental phasing to solve the structure. Further Sulfure SAD and Tantalum SAD data were collected on Proxima1. Data sets were processed individually and analysed with autoPROC toolbox (36). For crystals soaked in Tantalum Bromide, we obtained crystals in C2 and P21 space groups, however no significant anomalous signal was detected. Distant homologous structures failed to give phased solution, ultimately, a model was generated using AlphaFold2 (37) and structure was solved by molecular replacement using PHASER (38). The first model was rebuild in COOT (39). and refined at 2.63 Å using PHENIX (40). That model was used as solution for molecular replacement using PHASER and structure was refined using BUSTER (41,42) at 1.66 Å. Both structures were checked and confirmed to have good stereochemistry according to MOLPROBITY (43). Data collection and refinement statistics are listed in Table 1.

Structural analysis and sequence comparison of N7-MTase domain

Homologous structure searches using the DALI and PDBefold servers (44,45) retrieved the N7-MTases of the large DNA virus, African Swine Fever virus (ASFV, PDB 7D8U), *E. cuniculi* (PDB 1Z3C), human (PDB 3BGV) and Vaccinia virus (PDB 2VDW) as the closest homologues (sequence identity between 14 and 16%, *z*-score ranging from 17 to 20). Structural superimpositions of these structures was done manually using UCSF CHIMERA (46). From the structural superimposition, we generated a multiple sequence alignment (MSA) and visualized with each corresponding secondary structure in ESPript (47). Secondary structure analysis of WB N7-MTase was done using ENDscript (47). For motif identification, each sequence from the structural MSA was used as a BLAST input search against NR_50_1Nov (MPI TOOLKIT) (48). All retrieved sequences were merged in SEAVIEW (49) and sequences with 90% identity and above were discarded to avoid cross-referencing. The 102 remaining sequences were aligned according to the structural MSA profile. Visualization of the motifs was done using WebLogo (50).

Sequence comparison of Tobaniviridae family members

The presence/absence of the N7-MT-1a domain was compared via MSA using Clustal Omega (51) on SeaView for the 23 *Tobaniviridae* members (Supplementary Table S2), as recognised under the latest ICTV proposal (https://ictv.global/ICTV/proposals/2021.005S.R.Nidovirales.zip). The MSA was then manually modified (respecting amino acid chemistry and with no random gap introduction) based on the most conserved motifs/residues, determined from the extended structural MSA profile. Secondary structure prediction using PHD (52) were used to refine the alignment, in particular the conservation of the α helix and β strand secondary elements, corresponding to the characteristic β1–α2–β2 RF signature fold. The final alignment was visualized together with secondary structures of WBV, CSV, HpoToV_33749, VCSTV-A, MVNV, and SerTV-C18 using ESPript. The specific tobani- N7-MTase key motifs were visualized using WebLogo.

RESULTS

Expression of a 311 amino acid domain of the putative N7-MTase in White Bream virus yields a soluble, active enzyme

To confirm the recent identification of a putative N7-MTase domain in the pp1a of non-mammalian tobaniviruses, we first performed expression screens for one of the members of this family, White Bream virus (WBV). Alignments with the vaccinia virus D1 subunit N7-MTase (VV-D1) and other confirmed N7-MTases reveals several conserved domains, however the the natural N- and C-terminal borders, and importance of these regions for activity was unclear. As such, several lengths of the putative MTase, ranging from 260 to 336 amino acids in length were expressed and purified. All expressed domains include the five conserved motifs identified by structural alignment with the vaccinia virus D1 subunit N7-MTase (VV-D1). High yields of soluble protein (~20–30 mg/l culture) can be obtained from

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

a 311 amino acid domain spanning residues Ala[1358]–Ile[1668] (WBV-311) (Figure 1A). Gene constructs longer than this were insoluble. Expression of three N-terminally truncated domains, starting at Ser[1366] (WBV-303), Asp[1374] (WBV-295) or Ser[1389] (WBV-280) produces similar yields of soluble proteins. However, all four purified expression products appear as a double-band at their approximate molecular weights (31–34 kDa) when analyzed by SDS-PAGE (Figure 1B), suggesting that the protein is partially cleaved at the C-terminus during expression.

The soluble protein products were tested by filter binding assays (FBA) for their ability to methylate a short (13-mer) GpppG-capped RNA substrate using radioactive ^{3}H-AdoMet as the methyl donor. High levels of methyltransfer are observed for three out of the four soluble proteins, with no activity for controls lacking either RNA or protein (Figure 1C). WBV-295 is the most active product, and was subsequently selected to further characterize MTase activity. The shortest product (WBV-280) is >100-fold less active (25-fold higher enzyme concentration used in assay), indicating that the 16 N-terminal residues (Asp[1374]–Ser[1383]) of WBV-295 are functionally important. These results confirm that the coding sequence encompassed by Asp[1374]–Ile[1668] in the ORF1a of WBV harbors methyltransferase activity.

The WBV-MTase harbors N7-guanine specific activity

Methylation specificity was determined using FBA with a series of capped RNAs (RNA_{13}, Supplementary Table S1), pre-methylated at different positions of the cap structure (^{m}GpppN, GpppN$_m$, ^{m}GpppN$_m$) (Figure 1D). High levels of methylation are observed for any substrate with an unmethylated cap structure (Gppp*N*). The first nucleotide of the WBV genome is a G, however both GpppG and GpppA capped RNA substrates are methylated equally well (Figure 1D). Likewise, no difference in activity is observed when using a different RNA (GpppG-RNA_9, Supplementary Table S1), indicating that neither the identity of the first nucleotide nor the RNA sequence influences substrate recognition and/or methylation (Supplementary Figure S1a). Activity is blocked by pre-methylation at the N7-guanine position (^{m}GpppG-RNA, Figure 1D). Methylation is also drastically reduced for substrates carrying a 2′*O*-methylated ribose (Figure 1D). These results suggest that the WBV MTase specifically targets the N7-guanine site, and that N7-methylation precedes ribose 2′*O*-methylation, as is the case for other viruses (21,53). To confirm that the methylation activity was specifically targeted to the N7 position of the guanine, and not the less conventional N3 position, we performed TLC migration analysis using P^{32} radiolabelled α-GTP as the cap donor. Following incubation of either GpppG- or GpppA-RNA with the WBV MTase, caps migrate to the expected ^{m7}GpppN position (Figure 1E, Supplementary Figure S1b). No change in migration is observed for caps pre-methylated at the N7 position, while incubation with GpppN$_m$-RNA yields products that migrated to the ^{m7}GpppG$_m$ cap-1 structure, consistent with N7-guanine specificity. In contrast to the FBA, 100% of the GpppG$_m$ substrate is converted to the ^{m7}GpppG$_m$ when analyzed via TLC. This was attributed to the higher concentrations of enzyme and substrate used for TLC analysis, in addition to the longer incubation times. FBA analysis with increasing enzyme concentrations, similar to that used in TLC, confirmed this to be the case (Supplementary Figure S1c, d). We therefore conclude that while an unmethylated cap structure (GpppN) is the preferred substrate for N7-guanine methylation, GpppG$_m$ capped RNAs are also able to be methylated, albeit at a slower rate.

For bi-functional MTases, the preference for N7 or 2′*O* methylation is pH-dependent (32,54). For WBV, N7-methylation is optimal at pH 8–8.5, with no 2′*O* methylation at any pH, suggesting that this domain does not have dual functionality (Figure 1F). Some MTases also harbor internal 2′O methylation activity (32). For the WBV-MTase, no activity is observed with ^{m}GpppG$_m$-RNA substrates (Figure 1D). In addition, capped RNAs containing partial, or complete internal 2′*O*-ribose methylations (Gppp*N*(N_m)) are highly methylated, confirming that this MTase does not target internal residues (Figure 1D). The SARS-CoV nsp14 N7-MTase can also utilize GTP as an acceptor substrate for methylation (55). On the contrary, the WBV N7-MTase is unable to methylate GTP only (data not shown). The WBV MTase is therefore RNA Cap-guanine specific, with a strong preference for methylation of an open (GpppN) cap structure.

Structural overview of the *Tobaniviridae* N7-MT-1a

Crystals of the WBV N7-MT-1a were obtained using the WBV-295 construct, however they grew as thin needles that diffracted poorly (∼6 Å). A C-terminally truncated construct (Asp[1374]–Asp[1633], WBV-260) yielded a clean product of expected molecular weight, possessing activity levels equivalent to WBV-295 (Supplementary Figure S2). The WBV-260 protein crystallized in two different space groups, C2 and $P2_1$, with one monomer per asymmetric unit, diffracting to 2.6 and 1.7 Å, respectively (Table 1). The difference is likely due to soaking with tantalum bromide (C2 crystals), creating an additional constraint for the crystal. The two structures are almost identical, with the exception of the last flexible C-terminal helix that is not visible in the Tantalum-soaked crystal. AdoHcy, the by-product of the methylation reaction, is found in the active-site of both structures.

The central core of the WBV N7-MT-1a adopts a canonical Rossmann-fold (RF), with an alternating βαβ structural motif, characteristic of the class I family of MTases. This core structure is formed by a central, seven-stranded β-sheet (ordered β3, β2, β1, β4, β5, β10, β9), flanked by α-helices on each side (α1, α2, α3 on one side, and α4, α5/α6 and α7 on the other). The final β-strand (β10) is inserted in an anti-parallel orientation between the fifth and seventh strands (β5 and β9) of the β-sheet. In addition to this central fold, two other structural features are present; a flexible '3β-flap' domain comprised of a short α-helical loop followed by three anti-parallel β-strands (β1 and β6–β8) and an α-helical bundle (β2, α8 and α9) wrapping around α1, herein referred to as the α-bundle region (Figure 2).

Structural comparison with identified homologs

A structure similarity search retrieved the N7-MTase of the large DNA virus, African Swine Fever virus (ASFV, PDB

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

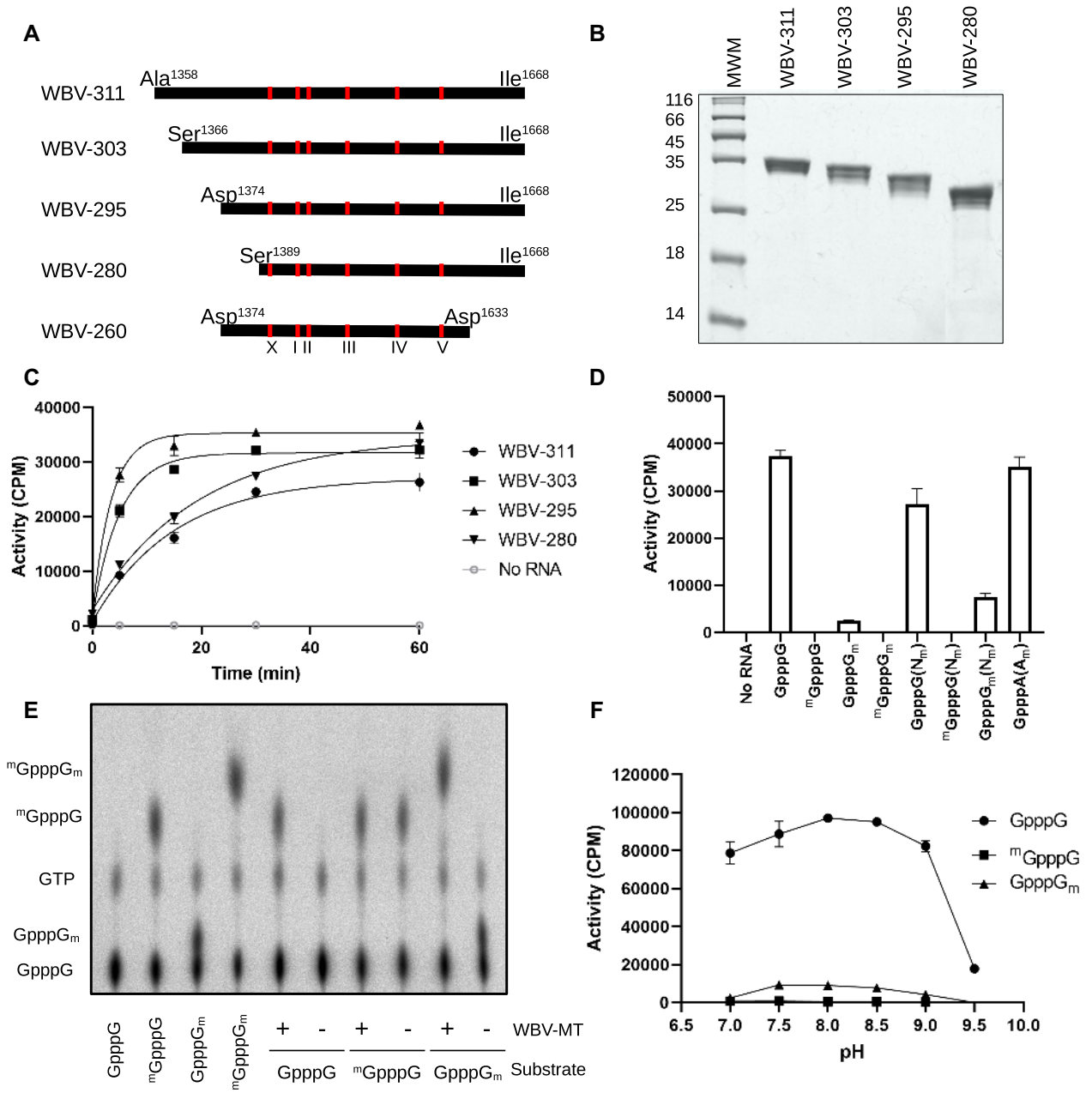


Figure 1. Cap-dependent methyltransferase activity of the WBV N7-MT-1a domain. For (C), (D) and (F), the transfer of tritiated methyl groups from AdoMet to capped RNA substrates (Table S1) was measured through filter binding assays and shown as counts per minute (CPM). (**A**) Schematic of the different constructs of N7-MT-1a. Limits indicate the starting amino acid and numbering in respect to 1a polyprotein, with approximate location of conserved domains (I, II, III, IV, V and X) shown. (**B**) SDS-PAGE analysis of different constructs. (**C**) Timecourse comparing the activity of different lengths of the putative N7-MT-1a domain from WBV. For the WBV-280 construct, final enzyme concentration in assays was 500 nM. For all other constructs, final enzyme concentration was 20 nM. (**D**) Methylation of different capped RNA_{13} substrates premethylated at the guanine N7 position (^{m}GpppN) or 2′*O* ribose groups of the first nucleotide (GpppN_m), or internal residues of the RNA (Gppp$N$(N_m)), where N is all residues. Activity represents single, 30 minute timepoint measured for the WBV-295 product, and is shown as the mean ± SD ($n = 3$). (**E**) Thin-layer chromatography analysis showing migration of differently methylated GpppG cap structures. The first four lanes show migration of cap controls, prepared with the commercial, vaccinia virus N7-MTase. The remainder of gel shows modification of different capped RNAs (substrate) in the presence (+) or absence (−) of the WBV N7-MT-1a. (**F**) Methylation activity of the WBV-295 product at different pH on different capped RNA_9 substrates. The plotted values represent a single, 30 min timepoint shown as the mean ± SD ($n = 3$).

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

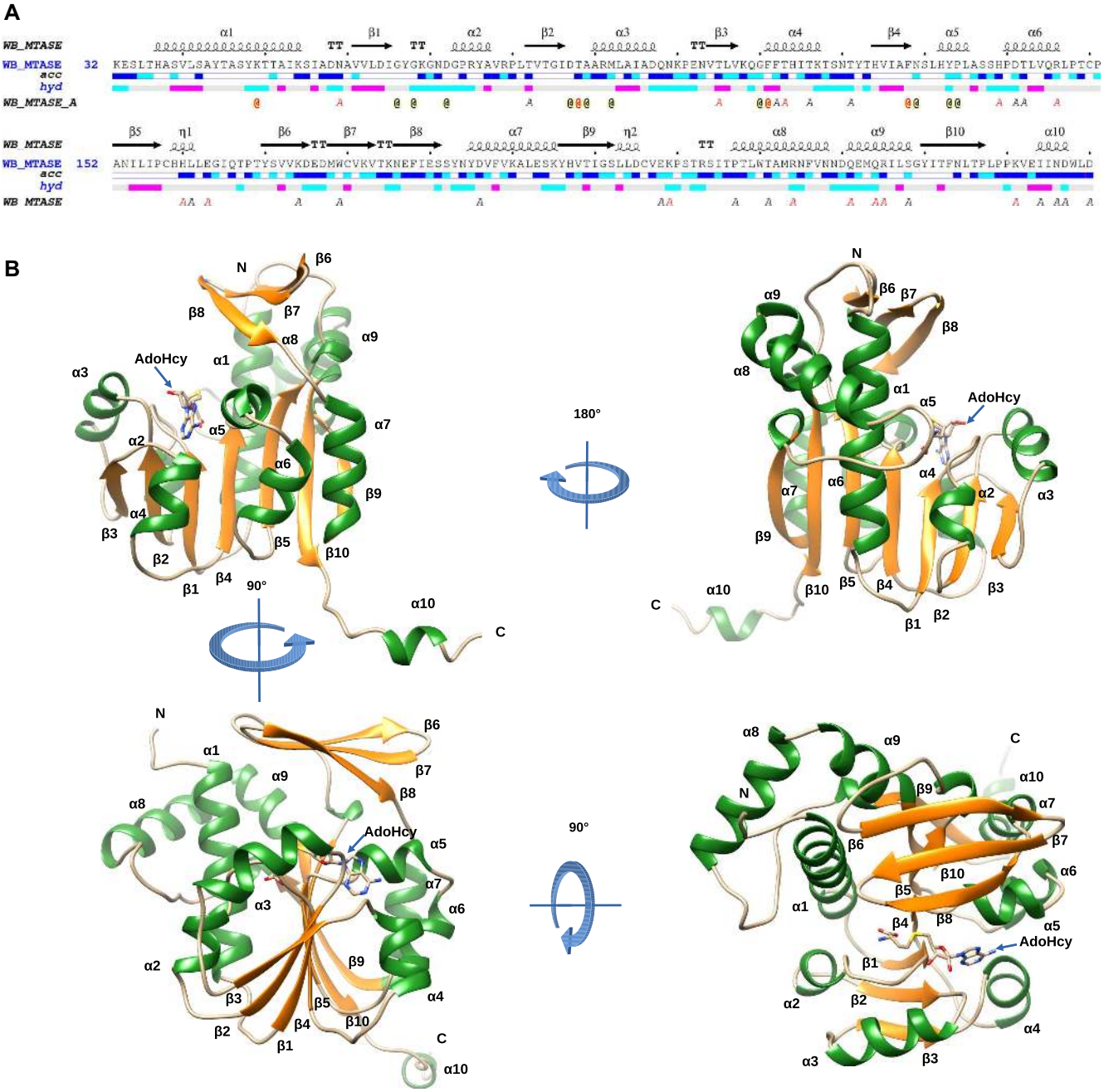


Figure 2. Sequence and structure of WBV N7-MT-1a. (**A**) Sequence of the resolved N7-MT-1a domain with the corresponding structure elements shown above, and solvent accessibility (acc; blue-exposed, cyan-intermediate, white-buried) and residues relative hydrophobicity (hyd; pink-hydrophobic, grey-intermediate, cyan-hydrophilic). Residues interacting with SAH are marked with an @. (**B**) Ribbon model of N7-MT-1a domain colored by secondary structures: loop (tan), α-helix (green), β-strand (gold). Secondary structures are labeled according to (A).

7D8U) as the closest homolog, with 16% sequence identity (*z*-score of 20.1 and a 2.6 Å rmsd) (56). Despite low sequence similarity (<15%) the N7-MTases from the microsporidian parasite *Encephalitozoon cuniculi,* Ecm1 (PDB ID: 1Z3C, 2.45 Å rmsd), VV-D1 (PDB ID: 2DVW and 4CKB, rmsd of 2.45 Å) and Hcm1 from humans (PDB ID: 5E8J, 2.3 Å rmsd) were additionally identified as potential homologues (Supplementary Figure S3). Superimposition shows the most significant structural variations are observed in the N-terminal domain, as well as in the 3β-flap and α-bundle regions.

The first 22 residues of the WBV N7-MT-1a are unresolved in the structure. This region is important for the methylation activity, as the WBV-280 construct (lacking residues Asp1 to Ser16) is > 100-fold less active than its longer counterpart, WBV-295 (Figure 1C). Comparatively, the N-terminal 40 amino acids (aa) are also missing from the Ecm1 structure, with at least 10 of these residues known

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

to be important for activity (PDB ID: 1Z3C) (57). The structure of the Vaccinia virus D1 N7-MTase shows that the equivalent region (residues 545–560), folds over the AdoMet binding site, directly contributing to MTase activity (PDB ID: 2VDW). In contrast, the N-terminus of ASFV forms a β-strand, which does not contribute to the active-site, but rather stabilizes the crystal packing. Together, these results suggest that the N-terminal domains of these N7-MTases are highly flexible, yet important for catalytic activity and/or structural stability.

A comparison of the 3β-flap domains of these five MTases shows the WBV N7-MT-1a flap-structure is most similar to that of Ecm1, which is also formed by an α-helix and three-stranded antiparallel β-sheet (Supplementary Figure S3). Nevertheless, the WBV N7-MT-1a domain is more compact, as both the ŋ1 helix and the loop between the first two β-strands are shorter. Likewise, the human, ASFV and vaccinia virus N7-MTases also contain longer helix and loop elements, as well as a fourth antiparallel *β*-strand, making their flap domains even larger.

The α-bundle domain is the most structurally diverse region of the MTase. While all related MTases possess a similar domain between β9 and β10 of the central β-sheet, the length, position, and number of α-helices vary substantially between these structures. The human N7-MTase contains an additional insertion in the α-bundle region (referred to as the lobe), which is absent in other structures, including WBV N7-MT-1a. The observed variability of the 3β-flap and α-bundle region is indicative of their dynamic role in the accommodation and stabilization of the RNA GpppG-cap structure during cap 0 synthesis and as such are structural markers of the N7-MTase fold.

A sequence homology search using the structural alignment of the five N7-MTases retrieved a pool of 120 diverse sequences, of which 6 conserved sequence motifs involved either in AdoMet binding and catalytic transfer (motifs X, I, III, IV) or in the GpppG-cap structure binding (motifs III, IV, V) can be defined (Figure 3, discussed below).

AdoMet/AdoHcy active-site pocket

As with other, class I MTases, AdoHcy binds in a deep, predominantly negatively charged groove (Figure 4A), formed by the loops between β1–α2, β2–α3, β3–α4 and β4–α5. The AdoHcy adopts an *anti*- conformation, similar to that seen in the Ecm1, Hmc1 and ASFV MTase structures. Like most other RF MTases, these enzymes share a conserved, glycine-rich motif I (residues 58–62 in WBV-260) located in the β1–α2 loop, that accommodates AdoMet binding. In contrast, for the VV-D1 MTase the adenosine nucleoside is bound in the rarer *syn*-conformation, likely due to the side chain of an aspartic acid (D598), that clashes with the adenosine base in the *anti*-conformation (Supplementary Figure S3).

Several highly conserved residues contribute to the stabilization of the methyl donor in the active site (Figures 3A and 4A, B). The carboxylate group of AdoHcy is coordinated by K39 of motif X (located in α1), and its amine group is stabilized by a water molecule bond with D62 of motif I. D62 is located in the β1–α2 loop region, just downstream of the conserved glycine-rich motif. The ribose ring is stabilized by a hydrogen bond with the side chain of D77 of motif II. Finally, Y123 of motif III has a hydrophobic interaction with the ribose and purine group. Alanine mutagenesis of several conserved residues located around the AdoMet binding site, including K39, D77 and D54 completely abolish activity (Figure 3B). A D62A mutant was insoluble and thus could not be tested, suggesting that this residue also plays a critical role in the folding and stability of the domain. Mutation of the semi-conserved lysine in motif I (K59A) reduces activity to ∼20% compared to the wild-type enzyme. The residual activity suggests that it does not directly contribute to AdoMet binding, but rather may play a role in the methyltransfer reaction. Overall, these results show that the AdoMet binding pocket of the WBV N7-MT-1a is well conserved with other RF N7-MTases.

Predicted cap binding site and coordination

The cap binding site can be inferred by the motif conservation (III, IV, V) and by analogy with the Ecm1 structure, complexed with ^{m7}GpppG (PDB ID: 1RI2). The site is an open, moderately charged pocket, lined on one side by residues from α1 and α9 of the α-bundle (Figure 4A). The floor is formed by β4 and β5, topped by the 3β-flap domain. A hydrophobic pocket (Y240, F188, L152, P147, I146) ensures general positioning of the cap base of the GpppG, while guanine base selectivity is ensured by hydrogen bonding to conserved residues E180 (motif IV), H122 (motif III) and Y123 (motif V), whose side chain atoms contact the guanine O6, N1 or N2 atoms (Figure 4B, C). While an H122A mutant was insoluble and thus not tested, mutation of E180 and Y240 to alanine abolishes methyltransferase activity, confirming the importance of these residues for cap binding (Figure 3). The only mutant that retains a low level of activity is K59A, located in motif I. This residue is not predicted to perform direct contacts with the cap structure or AdoHcy, but likely aids in providing a positive electrostatic charge for the positioning of the triphosphate of the cap structure.

The presence of the N7-MTase domain in non-mammalian infecting tobaniviruses

Using the newly defined WBV N7-MT-1a domain, we performed a MSA to look for the presence of this enzyme in other *Tobaniviridae* family members. This family is composed of 23 members which infect a range of mammalian (bovine, equine, porcine, etc.) and non-mammalian (various fish and reptiles) hosts. Remarkably, the N7-MT-1a domain appears to be present only in the 18 non-mammalian *Tobaniviridae* members (supplementary Figure 4). These viruses currently fall under the subfamilies *Serpentovirinae* and *Piscanivirinae*. While there is still some ambiguity for the precise location of N- and C-terminal domains, the characteristic secondary elements of RF enzymes (βaβ-fold) and sequence motifs could be easily identified for all of these viruses. We conclude that N7-MT-1a is a genetic marker for non-mammalian *Tobaniviridae*.

DISCUSSION

Capping of viral RNA is a common process employed by many virus families to mimic the host eukaryotic mRNA, which is essential for RNA stability, translation, and innate-immune escape (7–10). The cytoplasmic lifestyle of these

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

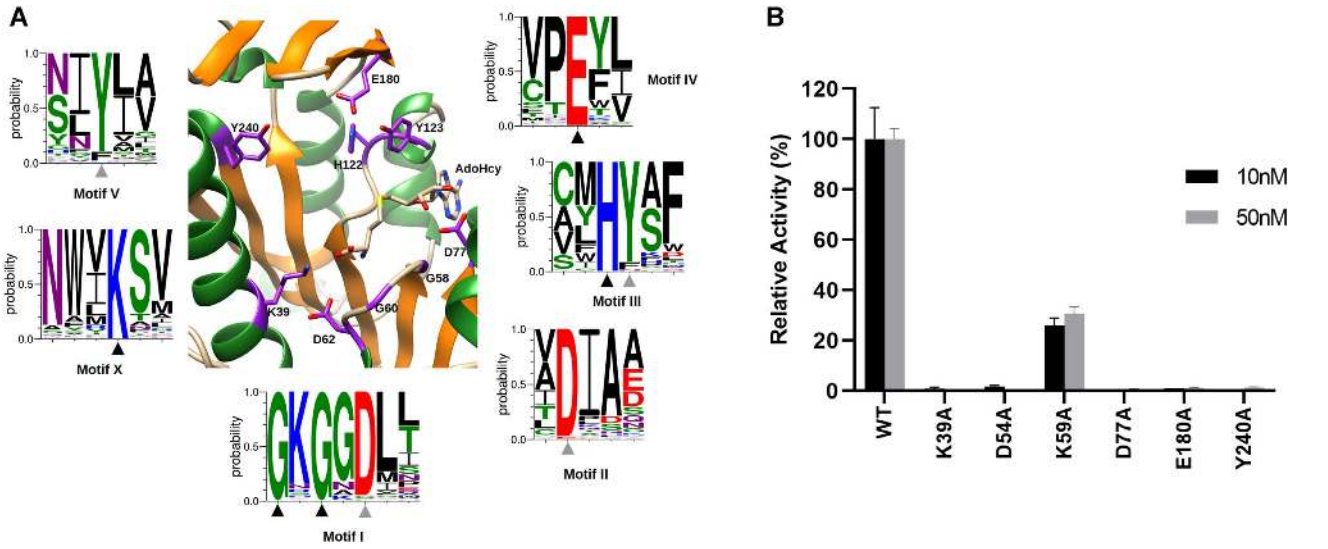


Figure 3. (**A**) Stuctural core conservation analysis. Central, catalytic site of the WBV N7-MT-1a in ribbons (same color code as in Figure 2) with side-chains of key residues shown in purple. WebLogo analysis of the six conserved motifs are shown around the structure, based on a structural MSA with other N7-guanine specific MTases. Key residues are indicated with a triangle (black if identical, grey if similar). (**B**) Activity of various WMV-260 mutants. The transfer of tritiated methyl groups from AdoMet to GpppG-RNA_{13} was measured through filter binding assays. Activity represents single, 30 min timepoint at two enzyme concentrations, and is shown as the mean ± SD relative to the activity of the wild-type enzyme ($n = 3$).

viruses requires them to carry their own capping machinery. Positive-sense RNA viruses vary considerably in their capping pathways and enzyme requirements (16). With the exception of the large-genome Nidoviruses, most +RNA viruses carry either a single methyltransferase domain, or none at all. For example, viruses of the *Picornaviridae* and *Caliciviridae* families do not possess any capping enzymes, but rather covalently attach a small viral protein (known as the VpG) to the 5′ end of the genome which functions as a cap-substitute (58). In the case of alphaviruses, the nsp1 protein harbors both GTase and MTase activity, first mediating the N7- methylation of GTP and subsequent transfer to RNA for formation of a cap-*O* structure (m^7GpppN) (59). In contrast the Flavivirus MTase on NS5 is able to perform both N7- and 2′-*O* methylation of the RNA cap (54,60–64). Genome size and structural constraints for these viruses apparently meet the activity level requirements to process the single capped RNA species, that is, their (+)RNA genomes. The cap methylation demand for large-genome nidoviruses is thus probably much greater, and may have driven the evolutionary acquisition of two separate MTases to cope for high methylation needs on the numerous mRNA species (genomic and subgenomic RNAs).

For *Coroniviridae*, the cap *N*7- and 2′*O*- methyltransferases have been well-characterized, residing in nsp14 and nsp16 of pp1ab, respectively (5,19–21). Nsp16, located at the very end of ORF1b, is easily discernible as a 2′*O* MTase due to its K–D–K–E catalytic sequence motif and conserved RF structure. A homologous 2′*O* MTase domain can be mapped at the same genomic location for most of the other families in the *Nidovirales* order, with the exception of several small-genome nidoviruses, including the *Arteriviridae* family. In contrast, the identification of the N7-MTase domain has proved more challenging, and was seemingly lacking for several nidovirus families. In the CoV family, the N7-MTase domain has a distinct, non-RF structure (5). In most other nidoviral families however, this N7-MTase located at the nsp14 locus appears to be either completely absent, or its presence is uncertain due to lack of homology. Here, we confirmed the presence of a Class-1, N7-specific MTase in pp1a of WBV, a member of the *Tobaniviridae* family. The WBV N7-MT-1a exhibits a canonical RF βαβ structural motif, similar to that of diverse N7-MTases, including that of *E. cuniculi,* humans and two large dsDNA viruses, despite low sequence similarity. This proved challenging for molecular replacement, with all sequence based structural predictions (with the exception of AlphaFold2) failing to converge toward either a consensus model or a partial model suitable for phasing the data. In terms of sequence, structure and substrate-specificity, there is no homologous MTase currently identified for any other +RNA virus to date. It is also remarkably different from the ORF1b N7-MTase gene of CoVs, suggesting that these MTases were acquired through separate evolutionary events.

Using the newly defined WBV N7-MT-1a domain, we revisited the *Tobaniviridae* family. Remarkably, this domain is only present in the 18 non-mammalian host *Tobaniviridae* members. The acquisition of such a critical gene is a considerably rare event. Combined with its unexpected genomic location in ORF1a, we believe that the presence of this MTase is significant enough to warrant its consideration as a genetic marker for this family of viruses. Consequently, we propose that the other four mammalian-infecting Tobanivirus members be reassigned to a new, separate nidovirus family.

The genomic location of the N7-MTase in ORF1a is surprising, and demonstrates the plasticity of nidovirus genomes (Supplementary Figure S5). In general, the organization of genes encoded on ORF1a and ORF1b can be

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

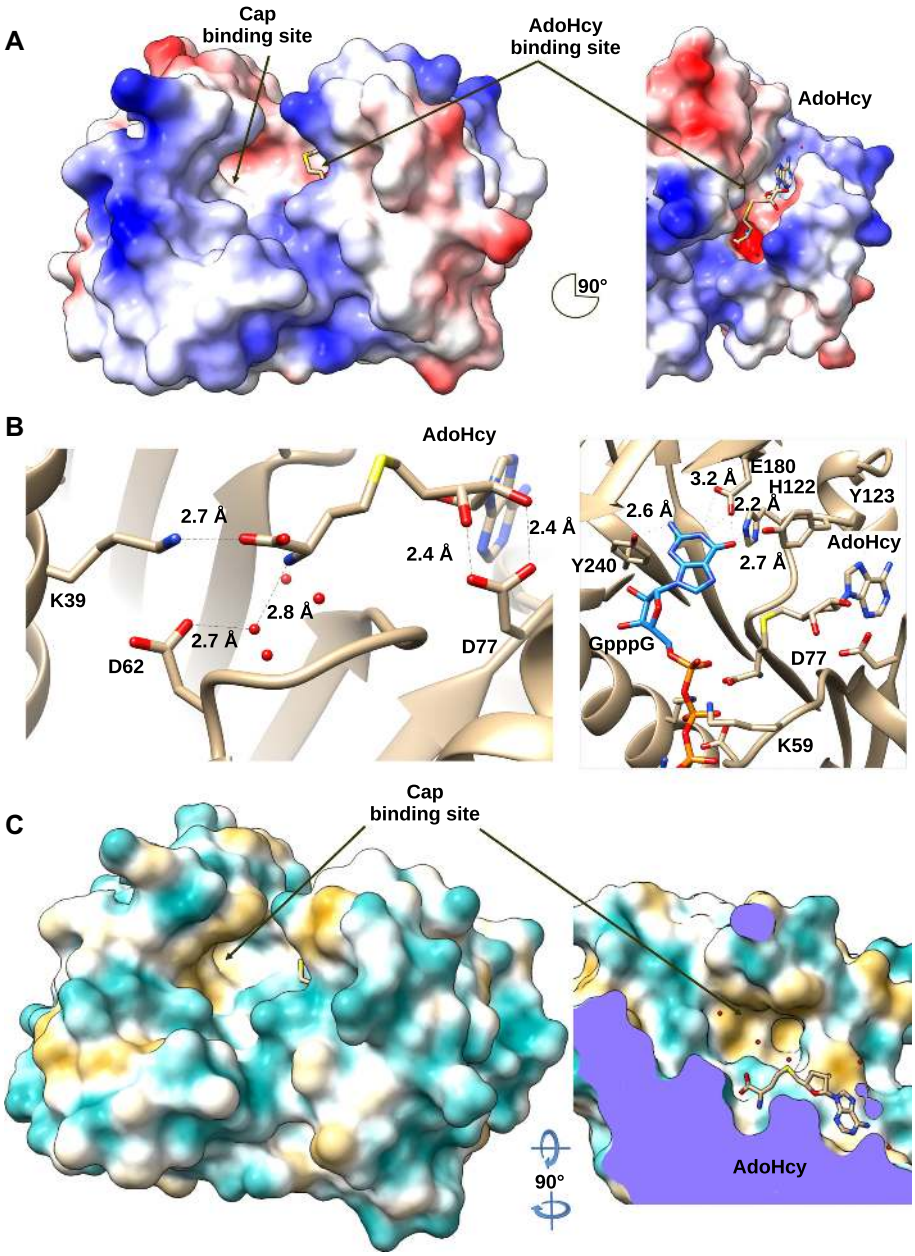


Figure 4. (**A**) Electrostatic surface potential of WBV N7-MT-1a, calculated by Adaptive Poisson–Boltzmann Solver, from –10 (red) to +10 (blue) kT/e. Left panel shows the cap binding site while the right panel is rotated 90° and shows the AdoHcy in AdoMet binding site. (**B**) Left panel: AdoMet binding site, with conserved residues shown in sticks, showing interactions with AdoHcy. Right panel: Predicted stabilization of GpppG, docked from PDB:1RI2 in WBV N7-MT-1a cap binding site. (**C**) Hydrophobic surface representation of WBV N7-MT-1a (yellow is hydrophobic, cyan is aliphatic). Left panel shows the cap binding site. The right panel is rotated 90° (on x and y axes) and sliced up to the AdoMet binding site.

summarized as follows: ORF1a codes for scaffolding proteins for the replication complex, protein activators (protease and regulatory cofactors), and enzymes involved in antagonizing the innate immune system; ORF1b codes for RNA synthesis, editing or degradation enzymes, including nucleotide transferase, polymerase, NTPase-helicase, nucleases and methyltransferases (65). This division of functional tasks is likely used to regulate the copy numbers of specific genes. The proteins expressed from ORF1a are expected to be present at higher concentrations compared to those from ORF1b, whose expression is controlled by a −1 ribosomal frameshift. This raises the following questions: what is the benefit (or requirement) for these tobaniviruses to express higher levels of their N7-MTase? Furthermore, why is the same not true for the 2′*O* MTase, which maintains its conserved genomic location at the end of ORF1b? While we cannot definitively answer these questions, we speculate that it is related to the reptilian/fish host specificity of these viruses. More specifically, it has been shown that there is an inverse correlation between levels of DNA methylation and body temperature. In reptiles and fishes (cold-blooded animals), methylation levels are higher than those of warm-

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

blooded mammals (66,67). It is therefore tempting to hypothesize that similar processes are at play regarding viral RNA cap methylation. Early and efficient methylation of the cap structure may be critical for escape from cellular immune sensors and for RNA stability. The acquisition of an N7-MTase gene in ORF1a is therefore in accordance with higher expression levels early during the viral replication cycle.

Alternatively, the tobanivirus N7-MT-1a may additionally recognize and methylate other substrates (distinct from the 5′ cap), potentially explaining the need for higher copy numbers of this enzyme. Both cellular and viral mRNAs have been shown to be internally methylated, most commonly at the m^6A, m^5C, m^1G and ribose 2′-*O*-positions. In the case of viral RNA, these modifications may have either pro- or anti-viral effects, and may be involved in regulating viral replication, gene expression, and/or overcoming the antiviral response (7–10). Here, we did not observe internal methylation of short, synthetic RNAs in biochemical assays. While it is still possible that the WBV MTase recognizes a specific sequence or secondary-structure in the viral RNA for methylation, this seems unlikely from a functional point of view, given the specificity for methylation at the N7-position of the cap guanine.

Cellular MTases have additionally been shown to methylate various other RNA substrates, including transfer RNA (tRNA) at various bases and positions (68). A structural comparison of the WBV N7-MT-1a with class I, RF tRNA N-MTases shows distinct RF topologies, including the lack of 3β-flap and α-helical bundle domains, which would otherwise overlap the tRNA substrate binding-site (Supplementary Figure S6) (69). Methylation is also not limited to RNA substrates, with protein methyltransferases (PMTs) able to perform N-methylation of arginine and lysine side-chains, most commonly during histone modification (70). Of note, one of these PMTs was identified as the best HHsearch hit for N7-MT-1a of two tobaniviruses, including WBV (30). While putative functions of this domain have been proposed based on sequence similarity alone (29,30), the statistical nature of the previous work stands as a prediction until demonstrated or disproved experimentally. The structure presented here clearly shows that the PMT structure is highly distinct from that of WBV (71) reflecting the critical adaptation of the catalytic site and confirming that the WB MTase structure is unambiguously an RNA MTase, as supported by the biochemical data.

Therefore, while we cannot definitively rule out the possibility that the tobanivirus N7-MT-1a recognizes another RNA substrate, the lack of a N7-MTase in ORF1b, the structural homology with known N7-guanine RNA MTases, the absence of homology with other protein or tRNA MTases and the strong N7-guanine specific cap activity, substantiates that the WBV pp1a MTase is the missing N7-guanine MTase involved in the RNA capping pathway.

In conclusion, we have demonstrated the presence of an N7-specific MTase domain in ORF1a of non-mammalian tobaniviruses, highlighting the amazing plasticity of the *Nidovirales* genome. Remarkably, this MTase is different from the ORF1b N7-MTase present in *Coronaviridae*. In the group of large RNA viruses, these two MTases were therefore likely acquired during separate events. Its structure belongs to the Rossman-fold family of enzymes, making it the first N7-specific RF MTase identified for (+)RNA viruses. Furthermore, its genomic location in ORF1a may also reflect a necessity for higher levels of methylation in cold-blooded hosts, supported by the fact that this domain is missing in tobaniviruses infecting mammalian hosts. Based on this, we propose that the *Tobaniviridae* family be further refined using this enzyme as a genetic marker.

DATA AVAILABILITY

Reference alignment corresponding to *Tobaniviridae* N7-MT-1a domain was deposited to UniProt to update accession numbers polyprotein 1a (Q008X5) or 1a-1b (Q008X6).

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data bank under accession number 7Z05 and 7Z2J.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Authors thank the intern Bill Makem-Tamekem for his technical assistance in protein and crystals production. The authors acknowledge SOLEIL for provision of synchrotron radiation facilities and we would like to thank beamline teams of Proxima-1 and Proxima-2 for their assistance during data collection.

FUNDING

Fondation pour la Recherche Médicale (FRM) [aide aux équipes]; Agence Nationale pour la Recherche through the French Infrastructure for Integrated Structural Biology (FRISBI) [ANR-10-INSB-05-01]; Fondation Infection Méditerranée [Bursary to B.S.]. Funding for open access charge: Fondation pour la Recherche Médicale.
Conflict of interest statement. None declared.

REFERENCES

1. Walker,P.J., Siddell,S.G., Lefkowitz,E.J., Mushegian,A.R., Adriaenssens,E.M., Alfenas-Zerbini,P., Dempsey,D.M., Dutilh,B.E., García,M.L., Curtis Hendrickson,R. *et al.* (2022) Recent changes to virus taxonomy ratified by the international committee on taxonomy of viruses (2022). *Arch. Virol.*, https://doi.org/10.1007/s00705-022-05516-5.
2. Gulyaeva,A.A. and Gorbalenya,A.E. (2021) A nidovirus perspective on SARS-CoV-2. *Biochem. Biophys. Res. Commun.*, **538**, 24–34.
3. Pasternak,A.O., Spaan,W.J.M. and Snijder,E.J.Y. (2006) Nidovirus transcription: how to make sense…?*J. Gen. Virol.*, **87**, 1403–1421.
4. Subissi,L., Posthuma,C.C., Collet,A., Zevenhoven-Dobbe,J.C., Gorbalenya,A.E., Decroly,E., Snijder,E.J., Canard,B. and Imbert,I. (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E3900–E3909.
5. Ferron,F., Subissi,L., Silveira De Morais,A.T., Le,N.T.T., Sevajol,M., Gluais,L., Decroly,E., Vonrhein,C., Bricogne,G., Canard,B. *et al.* (2018) Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **115**, E162–E171.

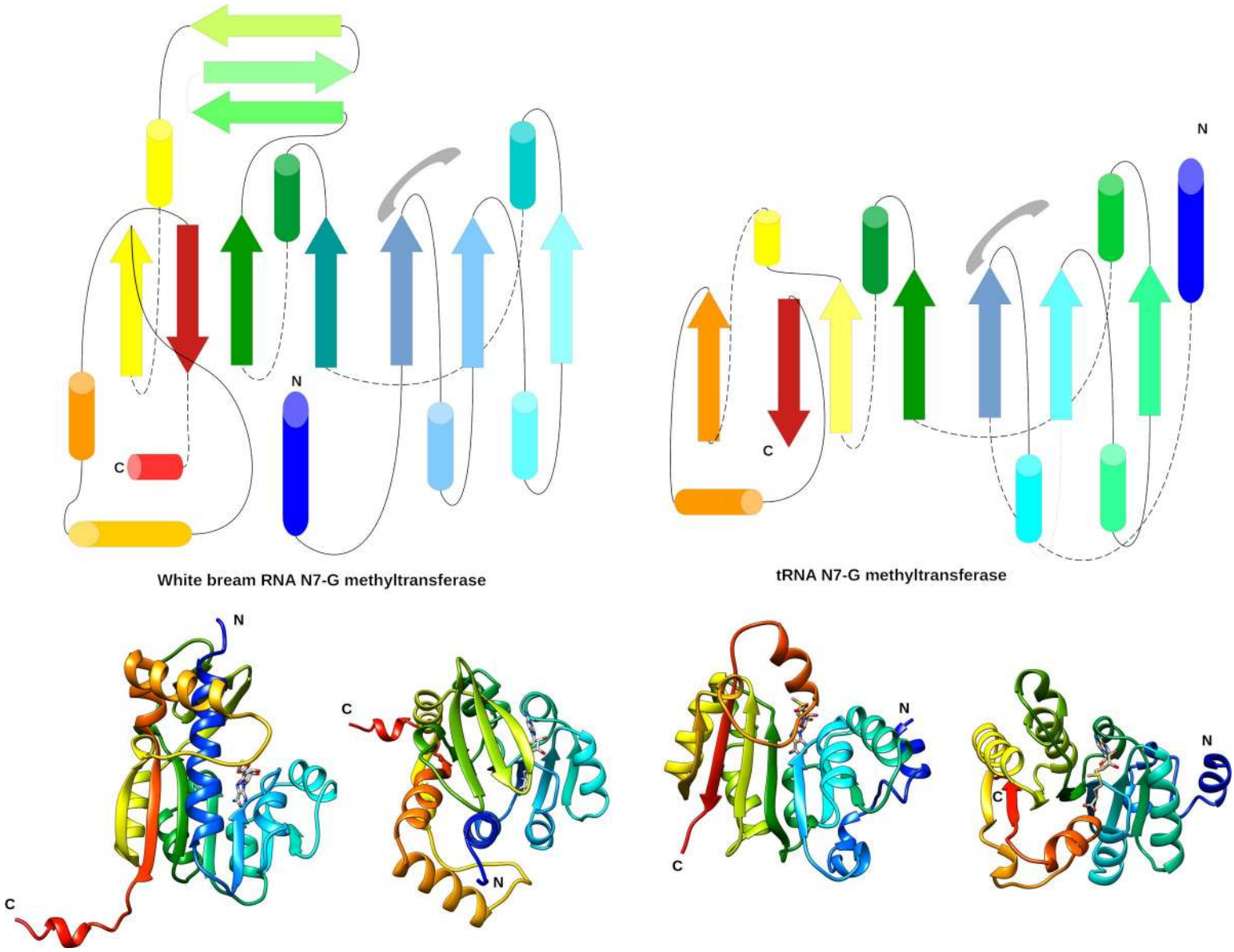
Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

6. Eckerle,L.D., Becker,M.M., Halpin,R.A., Li,K., Venter,E., Lu,X., Scherbakova,S., Graham,R.L., Baric,R.S., Stockwell,T.B. *et al.* (2010) Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog.*, **6**, e1000896.
7. Daffis,S., Szretter,K.J., Schriewer,J., Li,J., Youn,S., Errett,J., Lin,T.-Y., Schneller,S., Zust,R., Dong,H. *et al.* (2010) 2′-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature*, **468**, 452–456.
8. Li,P., Liu,Y., Song,R., Zhao,L., Yang,J., Lu,F. and Cao,X. (2022) RNA 2′-O-methyltransferase fibrillarin facilitates virus entry into macrophages through inhibiting type i interferon response. *Front. Immunol.*, **13**, 793582.
9. Ringeard,M., Marchand,V., Decroly,E., Motorin,Y. and Bennasser,Y. (2019) FTSJ3 is an RNA 2′-O-methyltransferase recruited by HIV to avoid innate immune sensing. *Nature*, **565**, 500–504.
10. Züst,R., Cervantes-Barragan,L., Habjan,M., Maier,R., Neuman,B.W., Ziebuhr,J., Szretter,K.J., Baker,S.C., Barchet,W., Diamond,M.S. *et al.* (2011) Ribose 2′-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor mda5. *Nat. Immunol.*, **12**, 137–143.
11. Ramanathan,A., Robb,G.B. and Chan,S.-H. (2016) mRNA capping: biological functions and applications. *Nucleic Acids Res.*, **44**, 7511–7526.
12. van Vliet,A.L.W., Smits,S.L., Rottier,P.J.M. and de Groot,R.J. (2002) Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus. *EMBO J.*, **21**, 6571–6580.
13. Lai,M.M., Patton,C.D. and Stohlman,S.A. (1982) Further characterization of mRNA's of mouse hepatitis virus: presence of common 5′-end nucleotides. *J. Virol.*, **41**, 557–565.
14. Lai,M.M. and Stohlman,S.A. (1981) Comparative analysis of RNA genomes of mouse hepatitis viruses. *J. Virol.*, **38**, 661–670.
15. Sagripanti,J.L., Zandomeni,R.O. and Weinmann,R. (1986) The cap structure of simian hemorrhagic fever virion RNA. *Virology*, **151**, 146–150.
16. Decroly,E., Ferron,F., Lescar,J. and Canard,B. (2012) Conventional and unconventional mechanisms for capping viral mRNA. *Nat. Rev. Microbiol.*, **10**, 51–65.
17. Ferron,F., Sama,B., Decroly,E. and Canard,B. (2021) The enzymes for genome size increase and maintenance of large (+)RNA viruses. *Trends Biochem. Sci.*, **46**, 866–877.
18. Schubert,H.L., Blumenthal,R.M. and Cheng,X. (2003) Many paths to methyltransfer: a chronicle of convergence. *Trends Biochem. Sci.*, **28**, 329–335.
19. Chen,Y., Cai,H., Pan,J., Xiang,N., Tien,P., Ahola,T. and Guo,D. (2009) Functional screen reveals SARS coronavirus nonstructural protein nsp14 as a novel cap N7 methyltransferase. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 3484–3489.
20. Bouvet,M., Debarnot,C., Imbert,I., Selisko,B., Snijder,E.J., Canard,B. and Decroly,E. (2010) In vitro reconstitution of SARS-Coronavirus mRNA cap methylation. *PLoS Pathog.*, **6**, e1000863.
21. Decroly,E., Imbert,I., Coutard,B., Bouvet,M., Selisko,B., Alvarez,K., Gorbalenya,A.E., Snijder,E.J. and Canard,B. (2008) Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2′O)-methyltransferase activity. *J. Virol.*, **82**, 8071–8084.
22. Ivanov,K.A., Thiel,V., Dobbe,J.C., Meer,Y.vander, Snijder,E.J. and Ziebuhr,J. (2004) Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J. Virol.*, **78**, 5619–5632.
23. Yan,L., Ge,J., Zheng,L., Zhang,Y., Gao,Y., Wang,T., Huang,Y., Yang,Y., Gao,S., Li,M. *et al.* (2021) Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell*, **184**, 184–193.
24. Park,G.J., Osinski,A., Hernandez,G., Eitson,J.L., Majumdar,A., Tonelli,M., Henzler-Wildman,K., Pawłowski,K., Chen,Z., Li,Y. *et al.* (2022) The mechanism of RNA capping by SARS-CoV-2. Nature, **609**, 793–800.
25. Walker,A.P., Fan,H., Keown,J.R., Knight,M.L., Grimes,J.M. and Fodor,E. (2021) The SARS-CoV-2 RNA polymerase is a viral RNA capping enzyme. *Nucleic Acids Res.*, **49**, 13019–13030.
26. Lehmann,K.C., Gulyaeva,A., Zevenhoven-Dobbe,J.C., Janssen,G.M.C., Ruben,M., Overkleeft,H.S., van Veelen,P.A., Samborskiy,D.V., Kravchenko,A.A., Leontovich,A.M. *et al.* (2015) Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. *Nucleic Acids Res.*, **43**, 8416–8434.
27. Ivanov,K.A. and Ziebuhr,J. (2004) Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5′-triphosphatase activities. *J. Virol.*, **78**, 7833–7838.
28. Yan,L., Yang,Y., Li,M., Zhang,Y., Zheng,L., Ge,J., Huang,Y.C., Liu,Z., Wang,T., Gao,S. *et al.* (2021) Coupling of N7-methyltransferase and 3′-5′ exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell*, **184**, 3474–3485.
29. Ferron,F., Debat,H.J., Shannon,A., Decroly,E. and Canard,B. (2020) A N7-guanine RNA cap methyltransferase signature-sequence as a genetic marker of large genome, non-mammalian tobaniviridae. *NAR Genomics Bioinforma.*, **2**, lqz022.
30. Gulyaeva,A.A., Sigorskih,A.I., Ocheredko,E.S., Samborskiy,D.V. and Gorbalenya,A.E. (2020) LAMPA, LArge multidomain protein annotator, and its application to RNA virus polyproteins. *Bioinformatics*, **36**, 2731–2739.
31. Bukhari,K., Mulley,G., Gulyaeva,A.A., Zhao,L., Shu,G., Jiang,J. and Neuman,B.W. (2018) Description and initial characterization of metatranscriptomic nidovirus-like genomes from the proposed new family abyssoviridae, and from a sister group to the coronavirinae, the proposed genus alphaletovirus. *Virology*, **524**, 160–171.
32. Martin,B., Coutard,B., Guez,T., Paesen,G.C., Canard,B., Debart,F., Vasseur,J.-J., Grimes,J.M. and Decroly,E. (2018) The methyltransferase domain of the sudan ebolavirus l protein specifically targets internal adenosines of RNA substrates, in addition to the cap structure. *Nucleic Acids Res.*, **46**, 7902–7912.
33. Lavergne,T., Bertrand,J.-R., Vasseur,J.-J. and Debart,F. (2008) A base-labile group for 2′-OH protection of ribonucleosides: a major challenge for RNA synthesis. *Chem. – Eur. J.*, **14**, 9135–9138.
34. Zlatev,I., Lavergne,T., Debart,F., Vasseur,J.-J., Manoharan,M. and Morvan,F. (2010) Efficient solid-phase chemical synthesis of 5′-Triphosphates of DNA, RNA, and their analogues. *Org. Lett.*, **12**, 2190–2193.
35. Thillier,Y., Decroly,E., Morvan,F., Canard,B., Vasseur,J.-J. and Debart,F. (2012) Synthesis of 5′ cap-0 and cap-1 RNAs using solid-phase chemistry coupled with enzymatic methylation by human (guanine-N7)-methyl transferase. *RNA*, **18**, 856–868.
36. Vonrhein,C., Flensburg,C., Keller,P., Sharff,A., Smart,O., Paciorek,W., Womack,T. and Bricogne,G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 293–302.
37. Jumper,J., Evans,R., Pritzel,A., Green,T., Figurnov,M., Ronneberger,O., Tunyasuvunakool,K., Bates,R., Žídek,A., Potapenko,A. *et al.* (2021) Highly accurate protein structure prediction with alphafold. *Nature*, **596**, 583–589.
38. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
39. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
40. Liebschner,D., Afonine,P.V., Baker,M.L., Bunkóczi,G., Chen,V.B., Croll,T.I., Hintze,B., Hung,L.-W., Jain,S., McCoy,A.J. *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in *phenix*. *Acta Crystallogr. Sect. Struct. Biol.*, **75**, 861–877.
41. Blanc,E., Roversi,P., Vonrhein,C., Flensburg,C., Lea,S.M. and Bricogne,G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2210–2221.
42. Smart,O.S., Womack,T.O., Flensburg,C., Keller,P., Paciorek,W., Sharff,A., Vonrhein,C. and Bricogne,G. (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in *BUSTER*. *Acta Crystallogr. D Biol. Crystallogr.*, **68**, 368–380.
43. Chen,V.B., Arendall,W.B., Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024

macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.
44. Krissinel,E. and Henrick,K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2256–2268.
45. Holm,L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Res.*, **50**, W210–W215.
46. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
47. Robert,X. and Gouet,P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320–W324.
48. Gabler,F., Nam,S., Till,S., Mirdita,M., Steinegger,M., Söding,J., Lupas,A.N. and Alva,V. (2020) Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinforma.*, **72**, e108.
49. Gouy,M., Guindon,S. and Gascuel,O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.*, **27**, 221–224.
50. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.
51. Madeira,F., Pearce,M., Tivey,A.R.N., Basutkar,P., Lee,J., Edbali,O., Madhusoodanan,N., Kolesnikov,A. and Lopez,R. (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.*, **50**, W276–W279.
52. Rost,B. and Sander,C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.*, **232**, 584–599.
53. Ray,D., Shah,A., Tilgner,M., Guo,Y., Zhao,Y., Dong,H., Deas,T.S., Zhou,Y., Li,H. and Shi,P.-Y. (2006) West nile virus 5′-cap structure is formed by sequential guanine N-7 and ribose 2′-O methylations by nonstructural protein 5. *J. Virol.*, **80**, 8362–8370.
54. Zhou,Y., Ray,D., Zhao,Y., Dong,H., Ren,S., Li,Z., Guo,Y., Bernard,K.A., Shi,P.-Y. and Li,H. (2007) Structure and function of flavivirus NS5 methyltransferase. *J. Virol.*, **81**, 3891–3903.
55. Jin,X., Chen,Y., Sun,Y., Zeng,C., Wang,Y., Tao,J., Wu,A., Yu,X., Zhang,Z., Tian,J. *et al.* (2013) Characterization of the guanine-N7 methyltransferase activity of coronavirus nsp14 on nucleotide GTP. *Virus Res.*, **176**, 45–52.
56. Du,X., Gao,Z.-Q., Geng,Z., Dong,Y.-H. and Zhang,H. (2020) Structure and biochemical characteristics of the methyltransferase domain of RNA capping enzyme from african swine fever virus. *J. Virol.*, **95**, e02029-20.
57. Fabrega,C., Hausmann,S., Shen,V., Shuman,S. and Lima,C.D. (2004) Structure and mechanism of mRNA cap (guanine-N7) methyltransferase. *Mol. Cell*, **13**, 77–89.
58. Goodfellow,I. (2011) The genome-linked protein VPg of vertebrate viruses — a multifaceted protein. *Curr. Opin. Virol.*, **1**, 355–362.
59. Ahola,T. and Kaariainen,L. (1995) Reaction in alphavirus mRNA capping: formation of a covalent complex of nonstructural protein nsPl with 7-methyl-GMP. *Proc. Natl. Acad. Sci. U.S.A.*, **92**, 507–511.
60. Egloff,M.-P., Benarroch,D., Selisko,B., Romette,J.-L. and Canard,B. (2002) An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J.*, **21**, 2757–2768.
61. Malet,H., Egloff,M.-P., Selisko,B., Butcher,R.E., Wright,P.J., Roberts,M., Gruez,A., Sulzenbacher,G., Vonrhein,C., Bricogne,G. *et al.* (2007) Crystal structure of the RNA polymerase domain of the west nile virus non-structural protein 5. *J. Biol. Chem.*, **282**, 10678–10689.
62. Mastrangelo,E., Bollati,M., Milani,M., Selisko,B., Peyrane,F., Canard,B., Grard,G., de Lamballerie,X. and Bolognesi,M. (2007) Structural bases for substrate recognition and activity in meaban virus nucleoside-2′-O-methyltransferase. *Protein Sci.*, **16**, 1133–1145.
63. Assenberg,R., Ren,J., Verma,A., Walter,T.S., Alderton,D., Hurrelbrink,R.J., Fuller,S.D., Bressanelli,S., Owens,R.J., Stuart,D.I. *et al.* (2007) Crystal structure of the murray valley encephalitis virus NS5 methyltransferase domain in complex with cap analogues. *J. Gen. Virol.*, **88**, 2228–2236.
64. Duan,W., Song,H., Wang,H., Chai,Y., Su,C., Qi,J., Shi,Y. and Gao,G.F. (2017) The crystal structure of zika virus NS5 reveals conserved drug targets. *EMBO J.*, **36**, 919–933.
65. Lauber,C., Goeman,J.J., Parquet,M.delC., Nga,P.T., Snijder,E.J., Morita,K. and Gorbalenya,A.E. (2013) The footprint of genome architecture in the largest genome expansion in RNA viruses. *PLoS Pathog.*, **9**, e1003500.
66. Varriale,A. and Bernardi,G. (2006) DNA methylation and body temperature in fishes. *Gene*, **385**, 111–121.
67. Varriale,A. and Bernardi,G. (2006) DNA methylation in reptiles. *Gene*, **385**, 122–127.
68. Swinehart,W.E. and Jackman,J.E. (2015) Diversity in mechanism and function of tRNA methyltransferases. *RNA Biol.*, **12**, 398–411.
69. Goto-Ito,S., Ito,T., Kuratani,M., Bessho,Y. and Yokoyama,S. (2009) Tertiary structure checkpoint at anticodon loop modification in tRNA functional maturation. *Nat. Struct. Mol. Biol.*, **16**, 1109–1115.
70. Kaniskan,H.Ü., Martini,M.L. and Jin,J. (2018) Inhibitors of protein methyltransferases and demethylases. *Chem. Rev.*, **118**, 989–1068.
71. Sun,L., Wang,M., Lv,Z., Yang,N., Liu,Y., Bao,S., Gong,W. and Xu,R.-M. (2011) Structural insights into protein arginine symmetric dimethylation by PRMT5. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 20538–20543.

Downloaded from https://academic.oup.com/nar/article/50/19/11186/6761739 by guest on 06 October 2024



Supplementary Figure 6. Topology of WBV N7-MT-1a (top left) and tRNA N7-G methyltransferase (top right), with structures shown below. Position of SAM binding-site indicated with grey semicircle.

b. Financement du projet de thèse

Ce projet doctoral a été autofinancé par notre laboratoire pour une durée totale de 4 années grâce aux financements des projets européens SCORE (#101003627) et CARE (#101005077). Ce projet a pu également être mené à bien grâce au soutien de la Fondation pour la Recherche Médicale (Aide aux Équipes), ainsi que de la French Infrastructure for Integrated Structural Biology (FRISBI). Deux bourses m'ont aussi permis de présenter ce travail au 34th European Crystallographic Meeting, grâce au soutien de l'Association Française de Cristallographie (AFC) et de l'European Crystallographic Association (ECA). Nous tenons à les remercier pour leur financement et leur soutien.

c. Figures supplémentaires

C1 : Tableaux des Essais d’expression à haut débit

	Cellules				Plasmide			
N°	Nom Cell	LacI ?	Plasmide en plus ?	Résistance	Plasmide testé	Tag	Position tag	Résistance
1	Tuner (DE3) – pRARE2/LacI	oui	pRARE2/LacI	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
2	BL21-PGTF2	-	pGT-f2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
3	C2523-RARE2	-	pRARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
4	C2523-RARElacI	oui	pRARE2/LacI	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
5	C2523-PGTF2	-	pGT-f2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
6	C3037	oui	MiniF lacIq	CmmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
7	ROS2	-	pRARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
8	ROS2(DE3)-pLysI	Peut etre..	pLysI/RARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR

	Culture							
N°	Milieu utilisé	Double résistance pour le même antibio	Antibios Utilisé	Température avant induction	Temps avant induction	Do d'induction	Température après induction	Temps de culture
1	LB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
2	LB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
3	LB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
4	LB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
5	LB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
6	LB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
7	LB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
8	LB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h

	Purification						Résultat Purif	
N°	Do Elution	Mg de proteine	Concentration en mg/ml	Quantité par L de culture théorique	Rendement	Précipitation de l'élution ?	Contamination sur gel	Choix final
1	1,269	1,994	0,797	33,227	+++	oui	++	2
2	1,548	2,515	1,006	41,918	+++	oui	++	2
3	1,618	2,651	1,060	44,177	+++	oui	++	2
4	0,298	0,166	0,066	2,761	-	-	---	-4
5	2,001	3,372	1,349	56,194	+++	oui	++	2
6	2,86	4,989	1,995	83,145	+++	oui	++	2
7	2,492	4,296	1,718	71,599	+++	oui	++	2
8	2,952	5,162	2,065	86,032	+++	oui	++	2

Annexe C1 : Tableaux des Essais d'expression à haut débit (De 1 à 8).
Voir Gels SDS-PAGES correspondant en annexes C2.

	Cellules				Plasmide			
N°	Nom Cell	LacI ?	Plasmide en plus ?	Résistance	Plasmide testé	Tag	Position tag	Résistance
9	Tuner (DE3) – pRARE2/LacI	oui	pRARE2/LacI	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
10	BL21-PGTF2	-	pGT-f2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
11	C2523-RARE	-	pRARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
12	C2523-RARElacI	oui	pRARE2/LacI	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
13	C2523-PGTF2	-	pGT-f2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
14	C3037	oui	MiniF lacIq	CmmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
15	ROS2	-	pRARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
16	ROS2(DE3)-pLysI	Peut etre..	pLysI/RARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR

	Culture							
N°	Milieu utilisé	Double résistance pour le même antibio	Antibios Utilisé	Température avant induction	Temps avant induction	Do d'induction	Température après induction	Temps de culture
9	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
10	TB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
11	TB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
12	TB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
13	TB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
14	TB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
15	TB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
16	TB	oui, CmR	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h

	Purification						Résultat Purif	
N°	Do Elution	Mg de proteine	Concentration en mg/ml	Quantité par L de culture théorique	Rendement	Précipitation de l'élution ?	Contamination sur gel	Choix final
9	0,552	0,644	0,258	10,730	+	a peine	---	-3
10	0,682	0,889	0,355	14,809	+	a peine	---	-3
11	0,419	0,393	0,157	6,557	-	a peine	---	-4
12	0,99	1,468	0,587	24,473	++	a peine	---	-2
13	0,545	0,631	0,252	10,511	+	a peine	--	-2
14	0,738	0,994	0,398	16,566	+	a peine	--	-2
15	0,766	1,047	0,419	17,445	+	a peine	--	-2
16	0,776	1,066	0,426	17,759	+	a peine	--	-2

Annexe C1 : Tableaux des Essais d'expression à haut débit (De 8 à 16).
Voir Gels SDS-PAGES correspondant en annexes C2.

	Cellules				Plasmide			
N°	Nom Cell	LacI ?	Plasmide en plus ?	Résistance	Plasmide testé	Tag	Position tag	Résistance
17	Tuner (DE3) – pRARE2/LacI	oui	pRARE2/LacI	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
18	BL21-PGTF2	-	pGT-f2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
19	C2523-RARE	-	pRARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
20	C2523-RARElacI	oui	pRARE2/LacI	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
21	C2523-PGTF2	-	pGT-f2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
22	C3037	oui	MiniF lacIq	CmmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
23	ROS2	-	pRARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR
24	ROS2(DE3)-pLysI	Peut etre..	pLysI/RARE2	CmR	pQE30 :: H6_Tev_NSP12 MERS	6His	Nter	AmpR & CmR

	Culture							
N°	Milieu utilisé	Double résistance pour le même antibio	Antibios Utilisé	Température avant induction	Temps avant induction	Do d'induction	Température après induction	Temps de culture
17	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
18	NZ	oui, CmR	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
19	NZ	oui, CmR	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
20	NZ	oui, CmR	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
21	NZ	oui, CmR	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
22	NZ	oui, CmR	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
23	NZ	oui, CmR	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
24	NZ	oui, CmR	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h

	Purification						Résultat Purif	
N°	Do Elution	Mg de proteine	Concentration en mg/ml	Quantité par L de culture théorique	Rendement	Précipitation de l'élution ?	Contamination sur gel	Choix final
17	1,214	1,890	0,756	31,501	+++	un peu	--	-1
18	1,661	2,732	1,093	45,526	+++	un peu	--	-1
19	1,481	2,393	0,957	39,878	+++	un peu	--	-1
20	0,638	0,806	0,322	13,429	+	un peu	---	-1
21	2,039	3,443	1,377	57,386	+++	oui	---	-3
22	1,601	2,619	1,047	43,643	+++	un peu	--	-1
23	1,366	2,176	0,870	36,270	+++	un peu	--	-1
24	1,464	2,361	0,944	39,345	+++	un peu	--	-1

Annexe C1 : Tableaux des Essais d'expression à haut débit (De 17 à 24).
Voir Gels SDS-PAGES correspondant en annexes C2.

	Cellules				Plasmide			
N°	Nom Cell	LacI ?	Plasmide en plus ?	Résistance	Plasmide testé	Tag	Position tag	Résistance
25	BL21	-	-	-	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
26	C2523	-	-	-	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
27	C2530	-	-	-	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
28	Tuner (DE3) – pRARE2/LacI	oui	pRARE2/LacI	CmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
29	BL21-PGTF2	-	pGT-f2	CmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
30	C2523-RARE	-	pRARE2	CmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
31	C2523-RARElacI	oui	pRARE2/LacI	CmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
32	C2523-PGTF2	-	pGT-f2	CmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
33	C3037	oui	MiniF lacIq	CmmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
34	ROS2	-	pRARE2	CmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
35	ROS2(DE3)-pLysI	Peut etre..	pLysI/RARE2	CmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR
36	ROS2(DE3)-pLysS	Peut etre..	pLysS/RARE2	CmR	pJ404 :: NSP12TevHB MERS	6His	Cter	AmpR

	Culture							
N°	Milieu utilisé	Double résistance pour le même antibio	Antibios Utilisé	Température avant induction	Temps avant induction	Do d'induction	Température après induction	Temps de culture
25	TB	-	AmpR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
26	TB	-	AmpR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
27	TB	-	AmpR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
28	TB	-	AmpR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
29	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
30	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
31	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
32	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
33	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
34	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
35	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h
36	TB	-	AmpR & CmR	25°C	Environs 5h	Environ 0,4	25°C	Environ 16h

	Purification						Résultat Purif	
N°	Do Elution	Mg de proteine	Concentration en mg/ml	Quantité par L de culture théorique	Rendement	Précipitation de l'élution ?	Contamination sur gel	Choix final
25	0,652	0,831	0,333	13,858	+	-	--	1
26	0,514	0,572	0,229	9,531	+	-	---	0
27	0,433	0,419	0,168	6,991	-	-	---	-2
28	0,685	0,894	0,357	14,892	+	-	--	1
29	0,543	0,626	0,251	10,440	+	-	--	1
30	0,555	0,649	0,260	10,816	+	-	---	0
31	0,614	0,760	0,304	12,666	+	-	---	0
32	0,629	0,788	0,315	13,136	+	-	---	0
33	0,677	0,878	0,351	14,641	+	-	---	0
34	0,566	0,670	0,268	11,161	+	-	---	0
35	0,518	0,579	0,232	9,656	+	-	---	0
36	0,462	0,474	0,190	7,901	-	-	---	-1

Annexe C1 : Tableaux des Essais d'expression à haut débit (De 25 à 36).
Voir Gels SDS-PAGES correspondant en annexes C2.

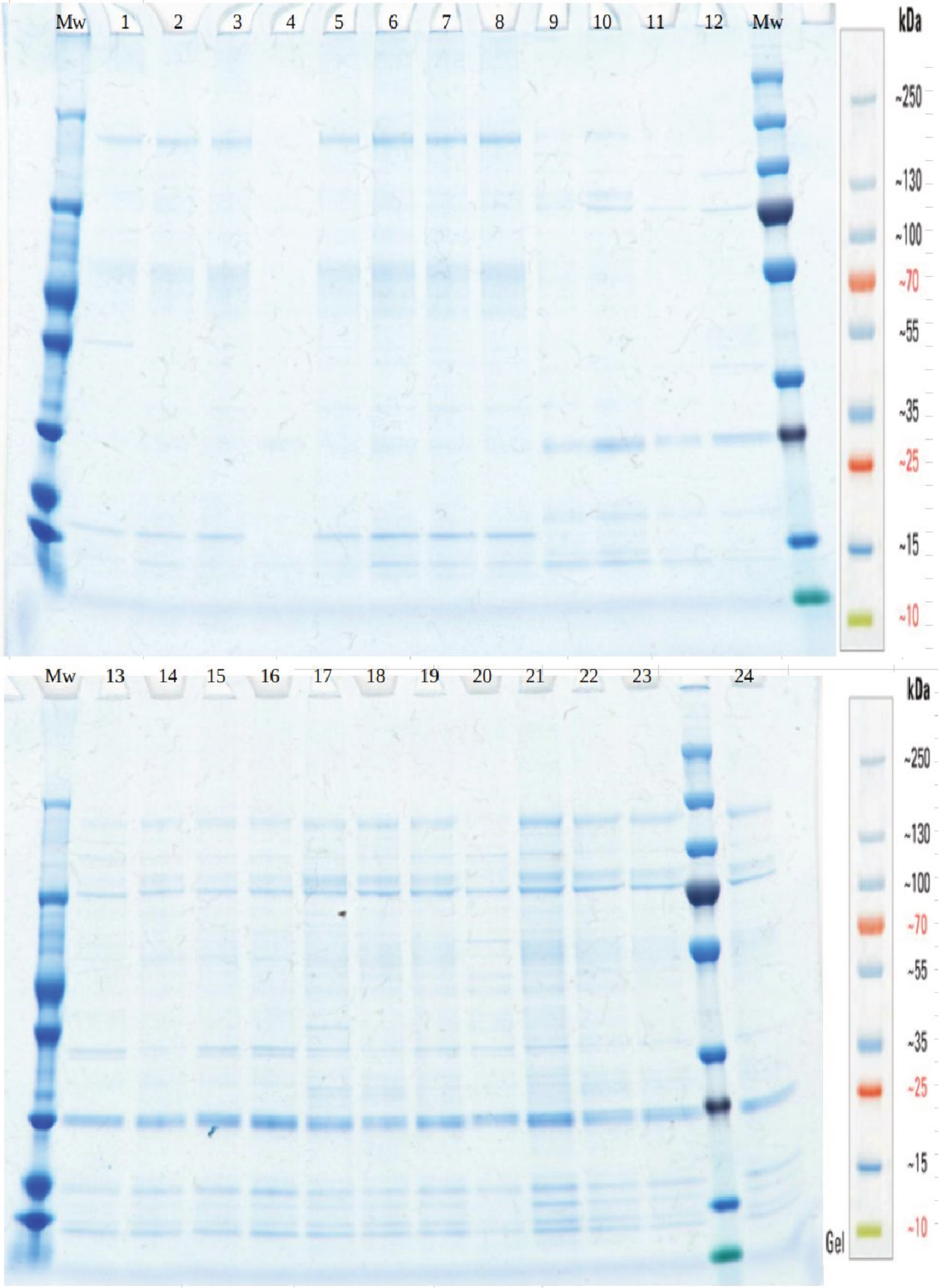
	Cellules				Plasmide			
N°	Nom Cell	LacI ?	Plasmide en plus ?	Résistance	Plasmide testé	Tag	Position tag	Résistance
37	BL21	-	-	-	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
38	C2523	-	-	-	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
39	C2530	-	-	-	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
40	Tuner	A vérifier	-	-	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
41	BL21-PGTF2	-	pGT-f2	CmR	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
42	C2523-RARE	-	pRARE2	CmR	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
43	C2523-RARElacI	oui	pRARE2/LacI	CmR	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
44	C2523-PGTF2	-	pGT-f2	CmR	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
45	C3037	oui	MiniF lacIq	CmmR	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
46	ROS2	-	pRARE2	CmR	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
47	ROS2(DE3)-pLysI	Peut être..	pLysI/RARE2	CmR	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR
48	ROS2(DE3)-pLysS	Peut être..	pLysS/RARE2	CmR	pJ404 :: NSP12TevH8 MERS	6His	Cter	AmpR

	Culture							
N°	Milieu utilisé	Double résistance pour le même antibio	Antibios Utilisé	Température avant induction	Temps avant induction	Do d'induction	Température après induction	Temps de culture
37	NZ	-	AmpR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
38	NZ	-	AmpR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
39	NZ	-	AmpR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
40	NZ	-	AmpR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
41	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
42	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
43	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
44	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
45	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
46	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
47	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h
48	NZ	-	AmpR & CmR	25°C	Environs 5h	Auto-inducteur	25°C	Environ 16h

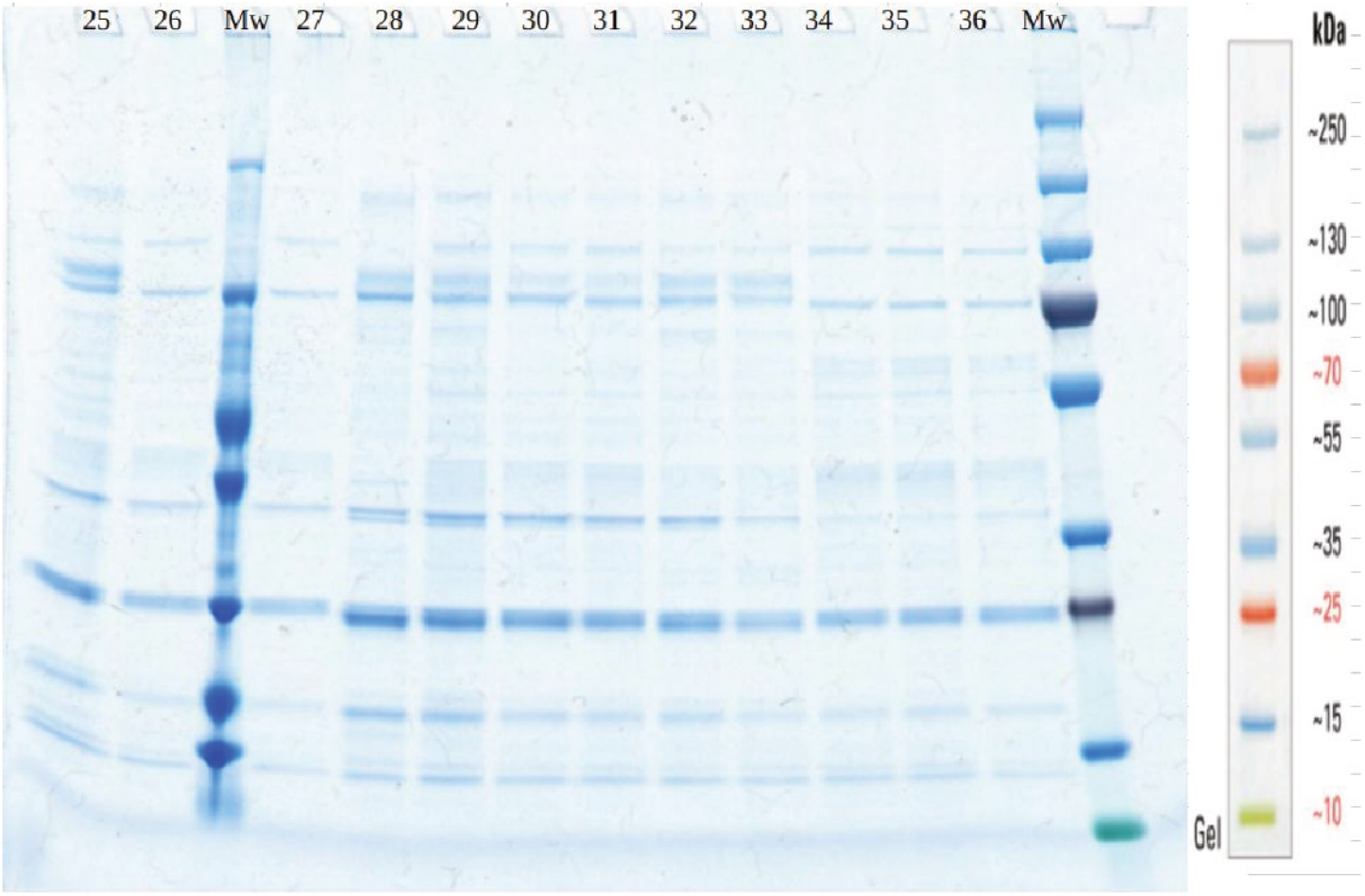
	Purification						Résultat Purif	
N°	Do Elution	Mg de proteine	Concentration en mg/ml	Quantité par L de culture théorique	Rendement	Précipitation de l'élution ?	Contamination sur gel	Choix final
37	0,709	0,939	0,375	15,645	+	-	--	1
38	0,705	0,931	0,372	15,519	+	-	---	0
39	0,71	0,941	0,376	15,676	+	-	---	0
40	0,546	0,692	0,253	10,534	+	-	---	0
41	0,608	0,749	0,299	12,478	+	-	--	1
42	0,761	1,036	0,415	17,275	+	-	--	1
43	0,589	0,713	0,285	11,882	+	-	--	1
44	0,865	0,856	0,342	14,265	+	-	--	1
45	1,059	1,597	0,639	26,618	++	-	-	1
46	0,6	0,734	0,293	12,227	+	-	---	-2
47	0,709	0,939	0,375	15,645	+	-	---	-2
48	0,654	0,835	0,334	13,920	+	-	---	-2

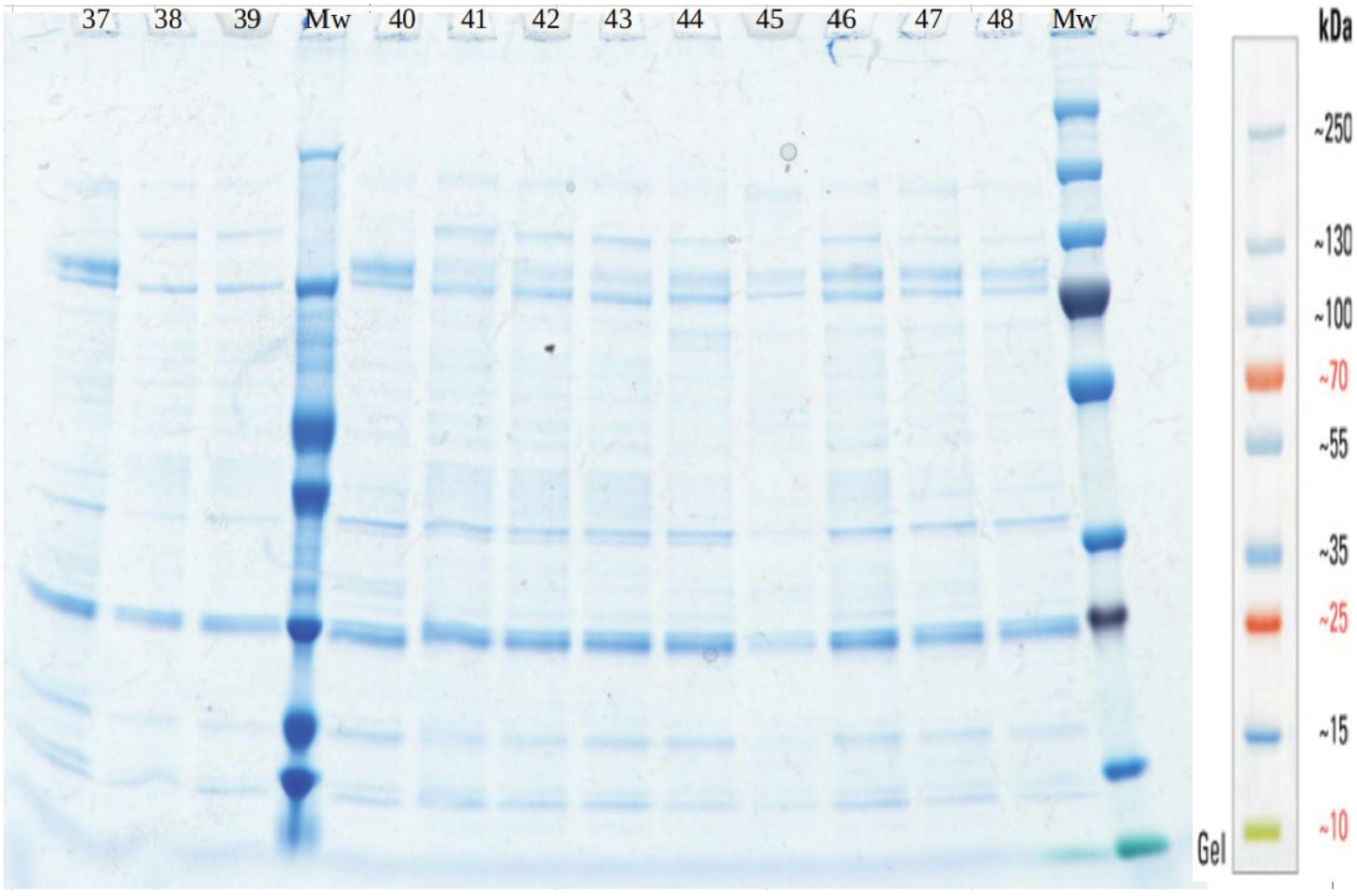
Annexe C1 : Tableaux des Essais d'expression à haut débit (De 37 à 48).
Voir Gels SDS-PAGES correspondant en annexes C2.

C2 : Gel SDS-PAGE 4-20% des échantillons élution effectuées en haut débit, après purifications IMAC-Ni



Annexe C2 : Gel SDS-PAGE 4-20% des échantillons élution effectuées en haut débit, après purifications IMAC-Ni.
Voir tableau des purifications correspondante en Annexes C1.

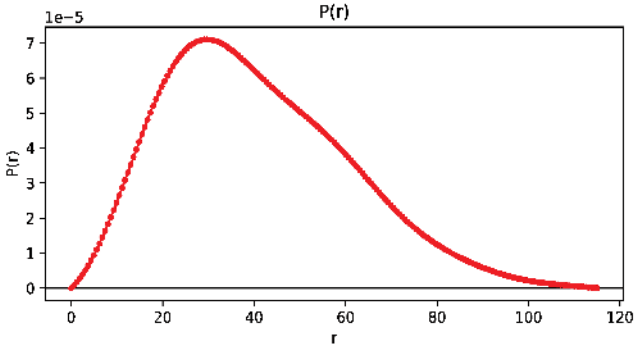


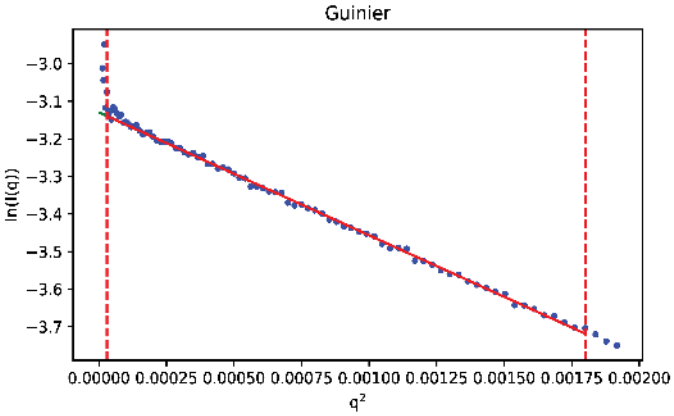


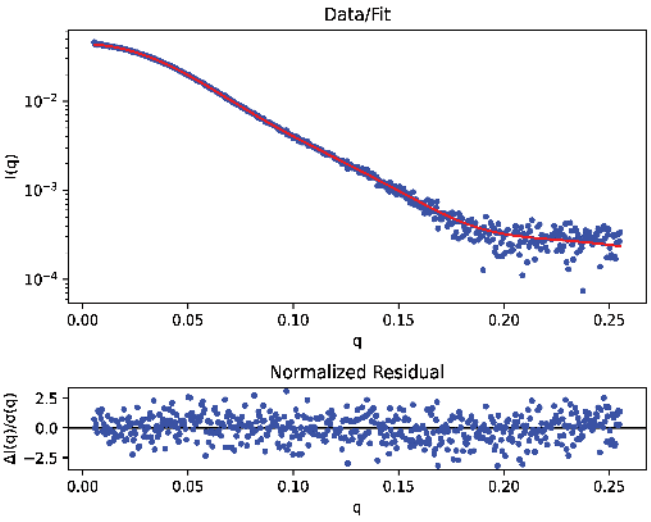
Annexe C2 : Gel SDS-PAGE 4-20% des échantillons élution effectuées en haut débit, après purifications IMAC-Ni
Voir tableau des purifications correspondante en Annexes C1.

C3 : Compléments des données de SEC-SAXS

Panel NSP14



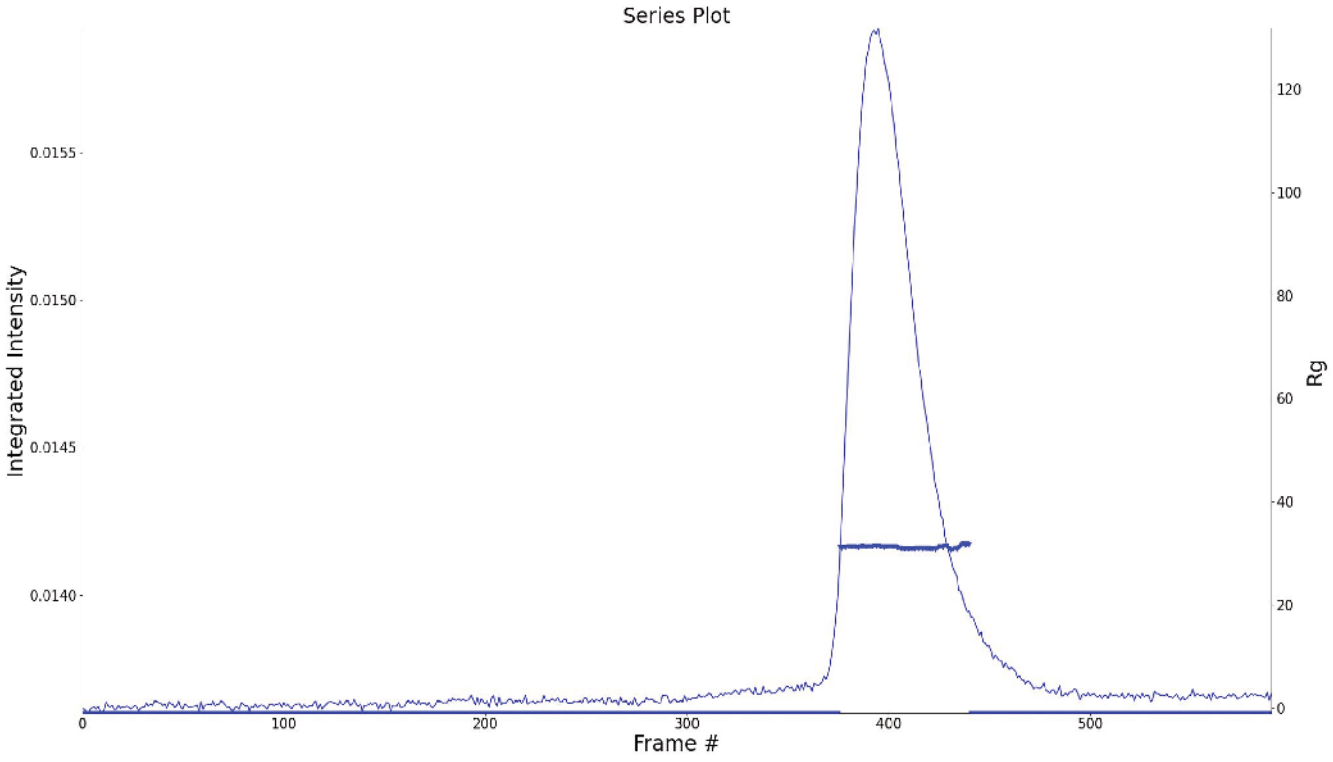




Guinier Fit	Rg	+/-	I(0)	+/-
	31,33	0,07	0,04	5,88e-5

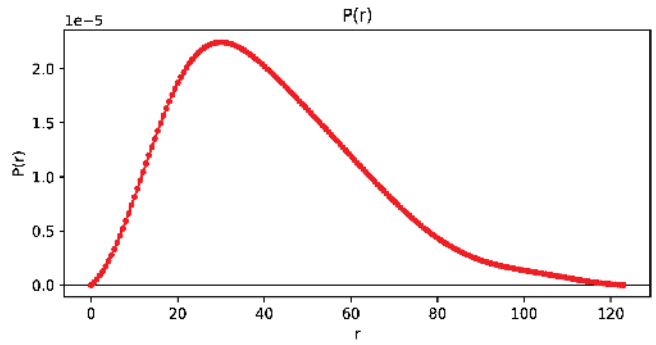
Molecular Weight	Vc	Bayes	Vp	S&S
	55,0	58,1	64,5	66,0

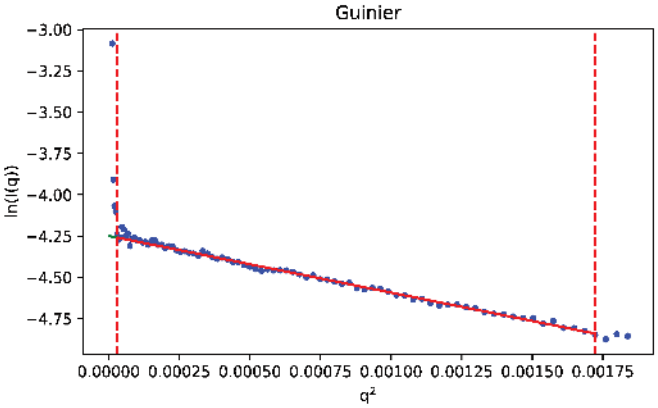
GNOM	Dmax	Rg	+/-	I(0)	+/-
	115	32,13	0,08	0,04	6,7e-5

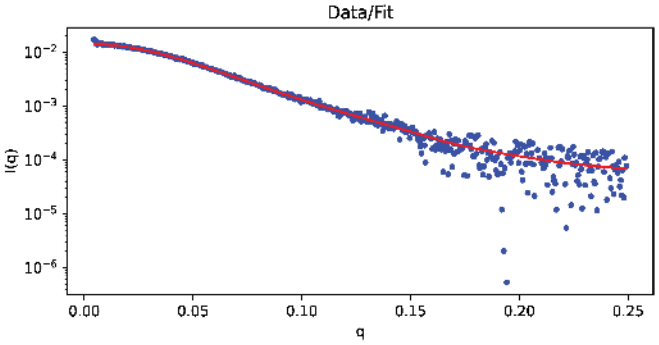


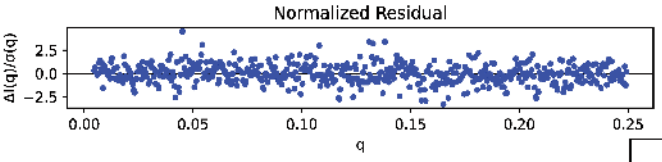
Annexe C3 : Compléments des données de SEC-SAXS pour la protéines nsp14 de SARS-CoV-2.

Panel NSP14-10





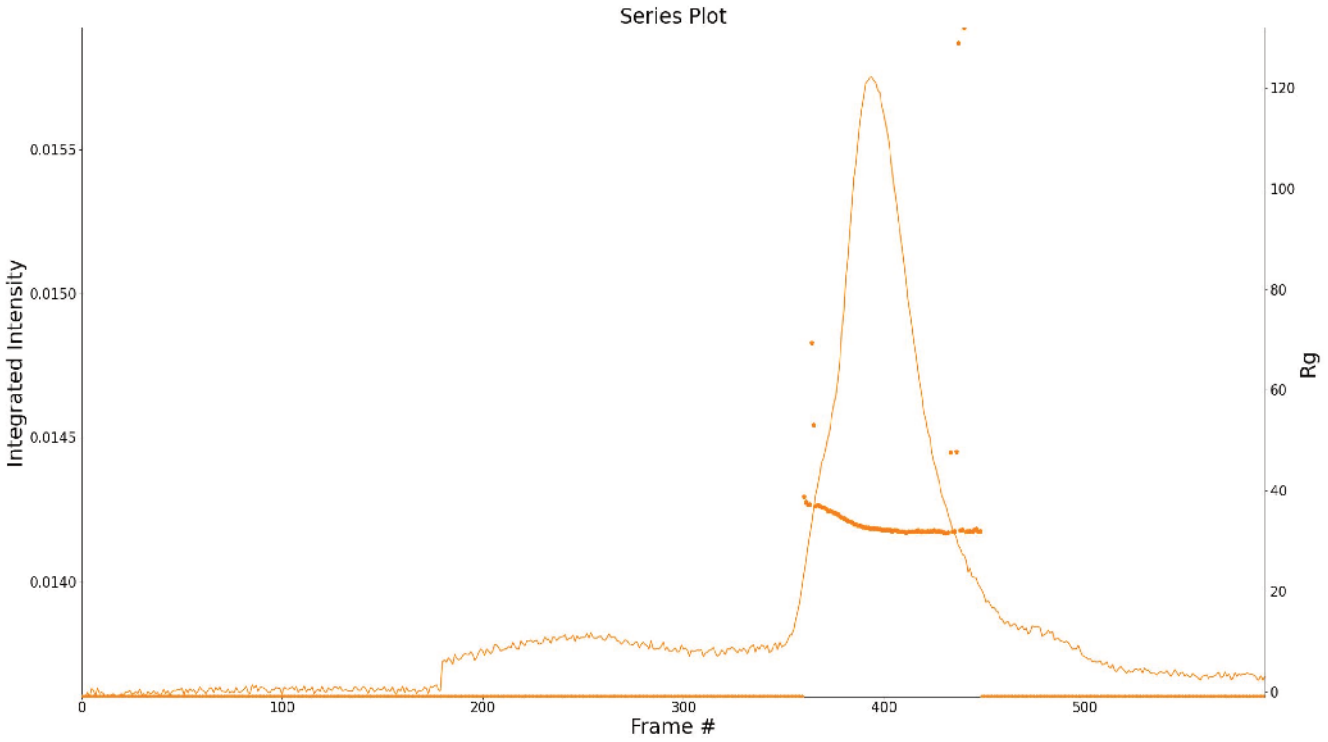




Guinier Fit	Rg	+/-	I(0)	+/-
	32,11	0,14	0,01	4,04e-5

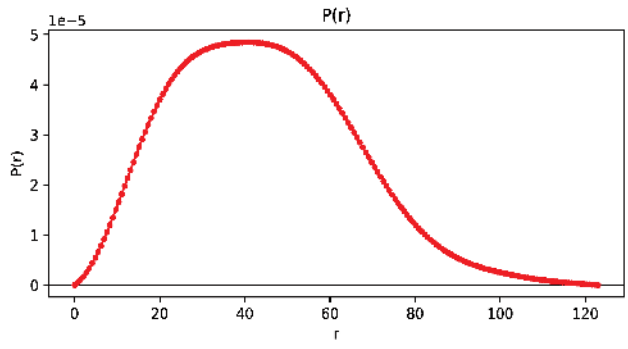
Molecular Weight	Vc	Bayes	Vp	S&S
	55,5	59,5	66,9	65,6

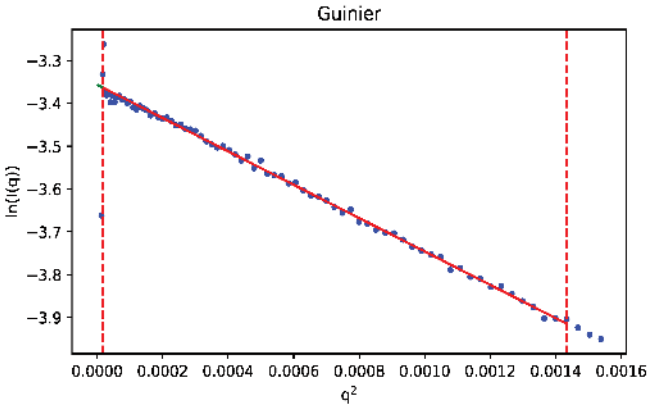
GNOM	Dmax	Rg	+/-	I(0)	+/-
	123,0	33,17	0,2	0,01	4,99e-5

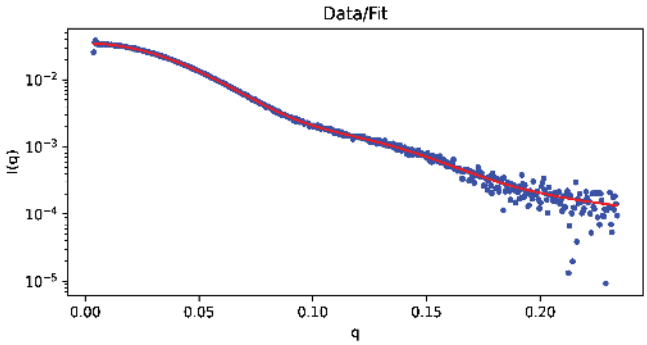


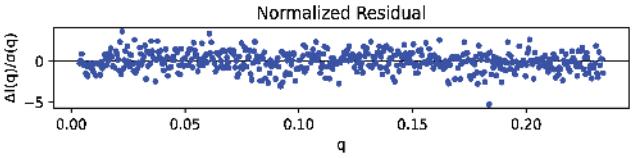
Annexe C3 : Compléments des données de SEC-SAXS pour le complexe nsp14/nsp10 de SARS-CoV-2.

Panel NSP14-vhh





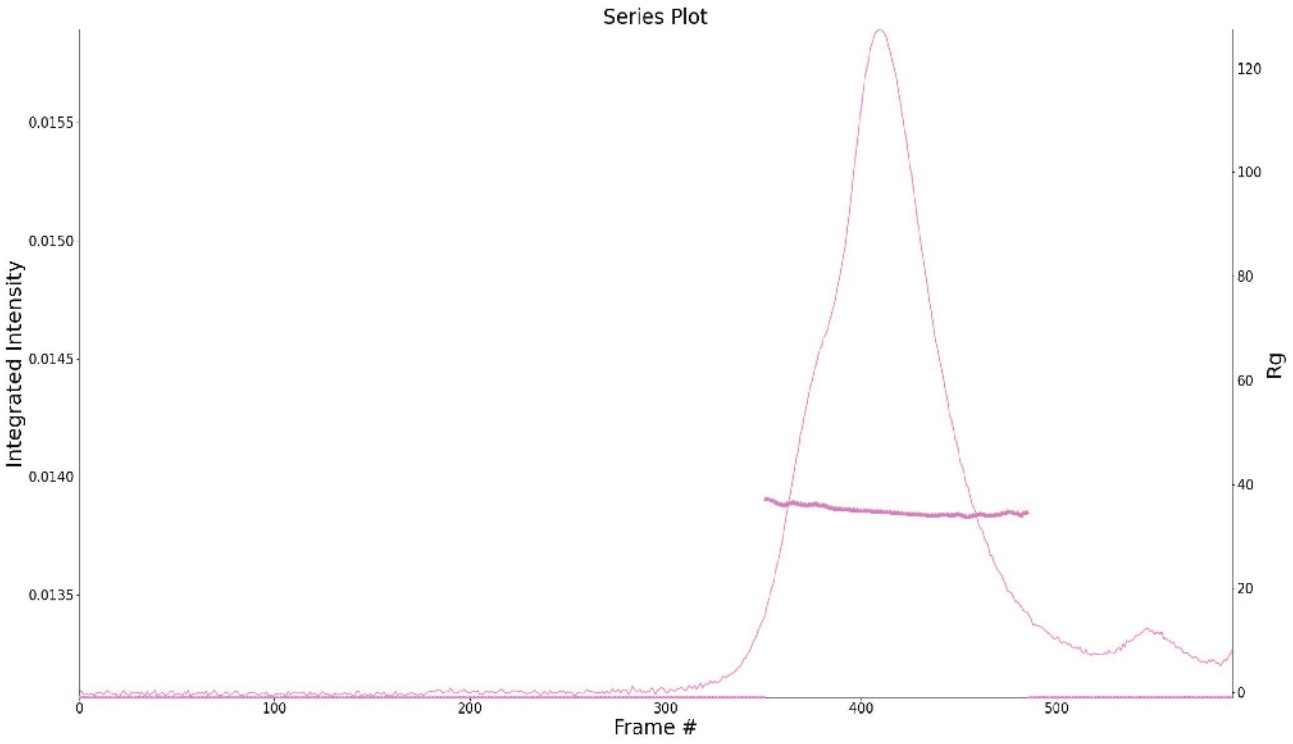




Guinier Fit	Rg	+/-	I(0)	+/-
	34,19	0,07	0,03	4,59e-5

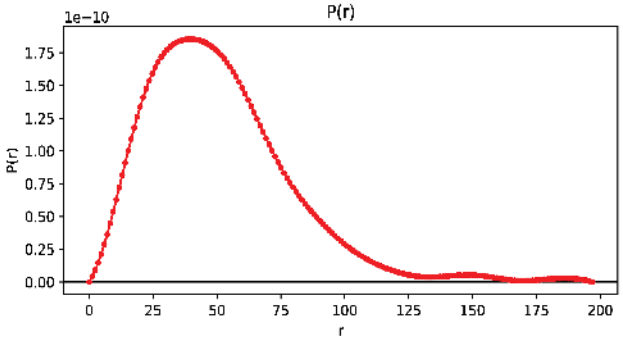
Molecular Weight	Vc	Bayes	Vp	S&S
	76,4	76,3	86,3	76,1

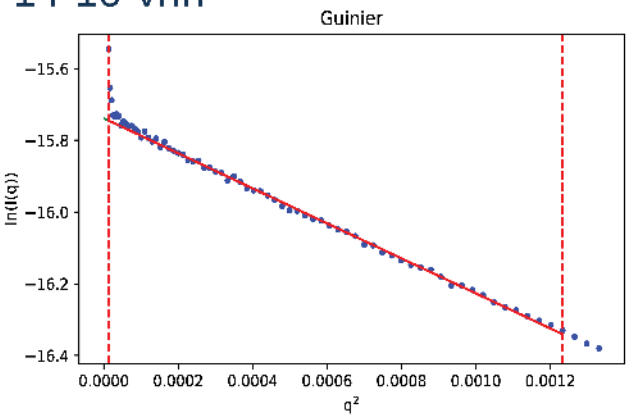
GNOM	Dmax	Rg	+/-	I(0)	+/-
	123,0	34,66	0,08	0,03	5,34e-5

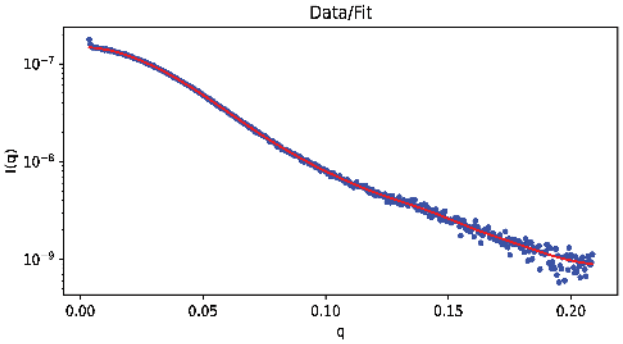


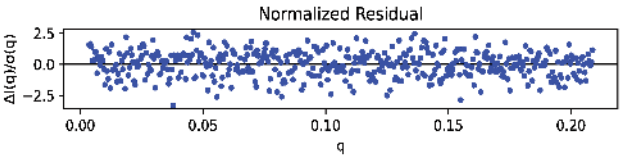
Annexe C3 : Compléments des données de SEC-SAXS pour le complexe nsp14/nsp10 de SARS-CoV-2.

Panel NSP14-10-vhh





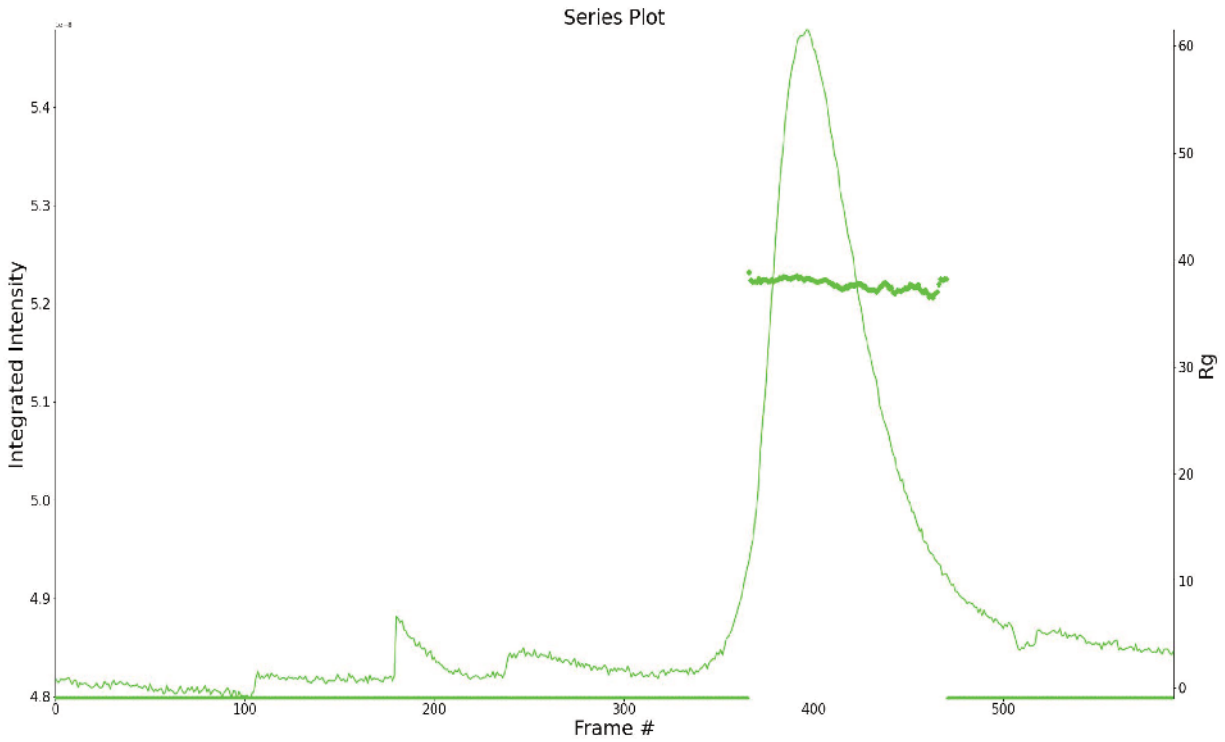




	Rg	+/-	I(0)	+/-
Guinier Fit	38,28	0,08	1,47e-7	1.99e-10

	Vc	Bayes	Vp	S&S
Molecular Weight	80,6	78,5	98,2	92,6

	Dmax	Rg	+/-	I(0)	+/-
GNOM	197,0	41,03	0,38	1.48e-7	3.94e-10



Annexe C3 : Compléments des données de SEC-SAXS pour le complexe nsp14/nsp10 de SARS-CoV-2.