

La cytosine, au cœur de l'immunité antivirale innée

Dans le cas spécifique du SARS-CoV-2, le rôle de métabolites à base de cytosine utilisés comme coordinateurs de la croissance cellulaire est central pour comprendre l'évolution du virus. En effet, la multiplication des virus à ARN monocaténaire de polarité positive nécessite des interactions coordonnées avec les fonctions de la cellule hôte. À cette fin, le génome du coronavirus (CoV) imite la structure d'un ARN messenger cellulaire, avec une coiffe méthylée en position 5' et une queue polyadénylée en position 3'. Il est rapidement traduit en une enzyme répliquant l'ARN (réplicase) et en protéines qui permettent au virus de détourner des fonctions spécifiques de l'hôte, utilisées lors de la traduction virale, de la réplication de l'ARN et de la construction de l'enveloppe du virus. Cela implique une série complexe d'événements, commençant par la réplication du virus en une matrice complémentaire de l'ARN viral, qui servira à générer de nouveaux génomes viraux et sera aussi transcrite puis traduite en plusieurs protéines virales [1]. Un ensemble critique de voies métaboliques est utilisé par le virus pour accéder à la réserve des ribonucléosides triphosphates nécessaires à la transcription des cinquante à cent copies du génome viral à chaque cycle de réplication. Cela rend la séquence virale très sensible aux spécificités métaboliques de la réserve des nucléotides, dont la composition va se refléter dans le génome viral au fur et à mesure qu'il évolue, en intégrant progressivement les différents types de pressions de sélection auxquelles le virus est confronté. Cette pression de sélection dépend de la disponibilité intracellulaire des précurseurs essentiels (nucléotides, lipides, acides aminés), ce qui crée de nombreux goulets d'étranglement pour l'évolution virale.

Outre l'ARN de son génome, une deuxième caractéristique clé du virus est qu'il s'agit d'un virus enveloppé. L'enveloppe est constituée de composants des membranes de la cellule hôte ainsi que de protéines spécifiques au virus. Comme l'ARN, les phospholipides de ces membranes dérivent de métabolites impliquant des nucléotides, en particulier des liponucléotides à base de CDP (cytidine diphosphate). Le caractère non homothétique de la croissance de l'hôte (voir ci-après) joue un rôle unique dans l'organisation du métabolisme des nucléotides.

Orchestrer la croissance non homothétique de la cellule

Avec le concept de croissance « non homothétique », utilisé en économie pour rendre compte de l'adéquation des ressources et de leur consommation dans de grands ensembles [2], nous soulignons ici le fait que le métabolisme, construit pour remplir le cytoplasme d'une cellule (donc à trois dimensions), crée des contraintes sur la croissance des membranes (deux dimensions) et du génome (une dimension). Regrouper ces trois contraintes tout en partageant un métabolisme commun n'est pas simple. Cependant, comme le nombre de modules élémentaires permettant la construction cellulaire est petit

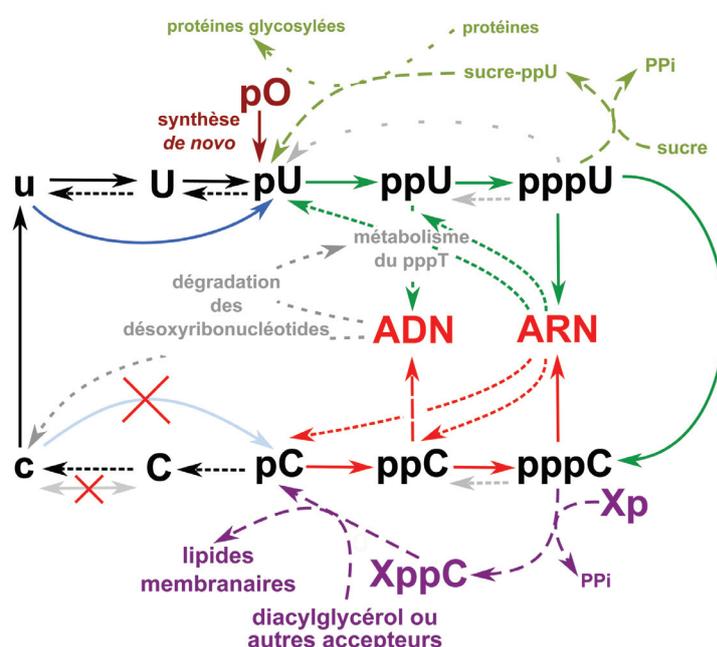


Figure 1 - Synthèse et recyclage des pyrimidines. En vert : la voie centrale de biosynthèse. La voie spécifique aux dérivés de la cytosine est en rouge. Les flèches en pointillé marquent le catabolisme. Les voies annexes sont en tirets : violet pour la synthèse des lipides et vert olive pour la glycosylation des protéines. En bleu est indiquée l'enzyme clé qui permet le recyclage des pyrimidines via l'uracile. La voie parallèle pour la cytosine n'a été découverte dans aucun organisme à ce jour.

(nucléotides, acides aminés, phospholipides et sucres), la sélection naturelle a fini par recruter une voie métabolique unique pour réussir cette prouesse. Nous montrons comment la cytosine et ses dérivés ont fini par devenir les métabolites de coordination, liant la croissance du génome et de la membrane au métabolisme central. Cette observation met l'accent sur des bizarreries du métabolisme des pyrimidines, ignorées jusqu'à présent (figure 1).

La synthèse *de novo* des nucléotides permet la production directe de tous les triphosphates, y compris de la cytidine triphosphate (CTP). Mais la molécule de CTP dérive de l'uridine triphosphate (UTP) en utilisant la CTP synthase (PyrG), dans une étape qui nécessite à la fois l'énergie de l'ATP et une source d'azote, ce qui rend la disponibilité de la molécule très sensible à la disponibilité en énergie et en azote métabolisable. Ce type de pression métabolique négative au cours de l'anabolisme s'appliquerait tout aussi bien à la synthèse d'ATP, par exemple, et ce n'est donc pas suffisant pour rendre la CTP originale. Cependant, les processus anaboliques ne décrivent qu'une partie du devenir des nucléotides : qu'en est-il de leur dégradation et de leur recyclage ? De fait, tous les métabolites s'usent avec le temps ; cela implique qu'ils sont dégradés ou recyclés. Cela est bien illustré dans le cas de trois des quatre nucléobases (adénine, guanine et uracile) par l'action d'une phosphoribosyltransférase spécifique de chaque base. On pourrait donc s'attendre à ce qu'il en soit de même avec la cytosine, permettant à la cellule de récupérer facilement cette

base à partir de son environnement. Pourtant, de manière surprenante, aucune cytosine phosphoribosyltransférase n'a encore jamais été identifiée dans un organisme vivant.

En revanche, l'expérience montre que le recyclage de la cytosine se fait à partir de dérivés de l'uracile (*figure 1*) et se complète grâce à la CTP synthase, enzyme présente même chez les parasites dépourvus de voie de biosynthèse des pyrimidines. Mieux, cette enzyme est codée par le gène *pyrG* conservé dans le plus petit génome d'une construction synthétique autonome (*Mycoplasma mycoides* Syn3.0 [3]). Soulignant le rôle clé de l'organisation spatiale dans la fonction de cette enzyme unique, nécessaire à la fois pour la synthèse *de novo* et le recyclage, son architecture est originale. Cette enzyme forme des filaments, appelés cytoophidies, dans tous les organismes où son organisation a été explorée. Les cytoophidies sont désormais considérées comme des organites sans membrane qui contrôlent la répartition spatiale d'un sous-ensemble du métabolisme intermédiaire au sein des cellules. Cette organisation est linéaire et satisfait donc les contraintes ascendantes de la croissance non homothétique – c'est-à-dire la croissance unidimensionnelle, encore plus contrainte que la croissance 2D par rapport à la croissance 3D. Cela est d'autant plus significatif que la CTP intervient dans un autre processus lié à la croissance cellulaire non homothétique : la synthèse des lipides membranaires. Les phospholipides qui forment les membranes dérivent de liponucléotides à base de cytosine – c'est une constatation universelle, qui n'a reçu aucune explication jusqu'à ce jour –, non seulement pour constituer la bicouche lipidique de la membrane, mais aussi pour contrôler sa forme via sa courbure. Tout cela place la CTP au carrefour des contrôles métaboliques. De quelle manière est-ce que cela affecte le développement du SARS-CoV-2 ?

Les coronavirus sont sensibles au métabolisme de la cytosine

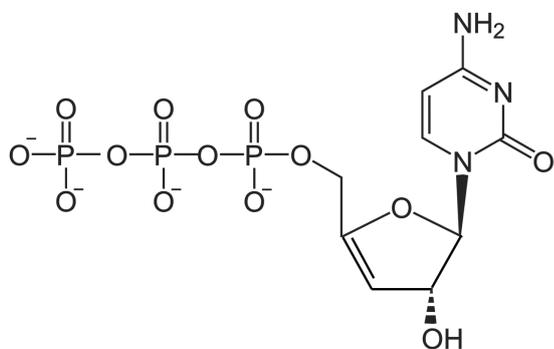
Une conséquence directe de cette organisation du métabolisme est que cela crée une force qui va tendre à diminuer la teneur en cytosine (C) des ARN, à moins que des processus opposés – et une pression de sélection conduisant à éliminer les organismes ayant une teneur en C trop faible, par exemple parce que cela créerait un insupportable biais dans la composition en acides aminés des protéines codées par ces génomes – aient le dessus au cours de l'évolution. Un détour nous conforte dans l'idée qu'il s'agit d'un phénomène crucial. Comme on peut s'y attendre en raison de l'importance des métabolites à base de cytosine, la nécessité d'une organisation spécifique du métabolisme des pyrimidines est implémentée dans d'autres structures que la CTP synthase. À titre d'exemple documenté dans de nombreuses cellules animales, les enzymes nécessaires au début de la synthèse *de novo* des pyrimidines, la carbamoyl-phosphate synthétase (CPSase), l'aspartate transcarbamylase (ATCase) et la dihydroorotase (DHOase) forment une structure multifonctionnelle, nommée CAD (à partir de la première lettre du nom des enzymes impliquées). Ces trois activités sont associées dans un seul polypeptide de 243 kDa. Cette protéine multi-enzymatique forme des hexamères et des oligomères plus compliqués [4]. Témoin de son importance, cette enzyme multifonctionnelle est sensible à la protéolyse par la caspase lors de l'apoptose, ce qui indique que le métabolisme de la pyrimidine est impliqué dans ce processus critique. De même, et cela est à mettre en relation

avec la réponse de l'hôte aux infections virales, CAD est fortement exprimée dans les leucocytes, où elle permet l'expression du récepteur 8 de type Toll en réponse à la teneur en cytidine et la présence d'ARN simple brin dans la cellule. Plus frappante encore, une autre caractéristique de la protéine CAD, pertinente pour notre propos, est que son activité est spécifiquement modulée par une protéine virale dédiée lors d'une infection par les entérovirus [5]. Cela nous incite à une analyse plus approfondie du métabolisme des pyrimidines en relation avec l'infection par le SARS-CoV-2, comme nous le voyons à présent.

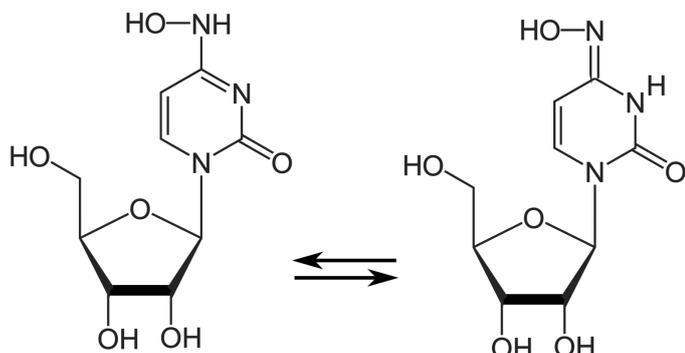
Les coronavirus et d'autres virus à ARN simple brin de sens positif produisent aussi des brins positifs avec un excès de 50 à 100 fois celui de leur matrice répliquée, à brin négatif. Cela signifie que la quantité globale des nucléotides qu'ils consomment dans les cellules n'est pas limitée par la deuxième règle de parité de Chargaff (qui se traduirait par une quantité d'adénine A égale à celle d'uracile U, ou de guanine G à celle de cytosine C). La « transcription » à partir de la matrice ARN à brin négatif du virus devrait ajouter un écart supplémentaire à cette règle de parité. Cependant, comme la multiplication du virus repose sur un processus de répllication, aussi asymétrique soit-il, toute pression sur la disponibilité d'une base donnée (ici C) devrait affecter son complément (G dans notre cas).

Par ailleurs, les mutations observées à mesure que le virus évolue refléteront aussi un peu les forces physico-chimiques agissant pendant la répllication, mais elles seront toujours dominées par la disponibilité des métabolites due à l'organisation générale du métabolisme. Nous l'avons vu, nous nous attendons à une pression de sélection générale opérant sur la CTP et tendant, à long terme, à diminuer la teneur en cytosine du virus à ARN. Cette tendance sera toutefois mise en échec par toutes les formes de pression de sélection agissant sur les fonctions qui dirigent la répllication et la propagation du virus et opèrent sur les protéines correspondantes, donc sur les codons. Cette pression est particulièrement importante pour le résidu proline, essentiel dans le repliement des domaines clés des protéines virales, parce que cette amine secondaire cyclique essentielle aux protéines est codée par les codons CCN. La cytosine en deuxième position des codons est aussi essentielle pour permettre l'introduction de résidus thréonine ou alanine dans les protéines virales, tandis que la première position est essentielle pour le codage de l'histidine et de la glutamine. Dans ce contexte, il est notable que l'un des changements critiques dans le coronavirus du SRAS en 2003 (SARS-CoV-1) par rapport à ses homologues inoffensifs était un changement d'une leucine en alanine à la jonction entre les domaines S1 et S2 de la protéine de pointe et que cela nécessitait un changement de U vers C [6].

Le rôle unique de la cytosine en tant que coordinateur du métabolisme cellulaire a été exploité par la sélection naturelle pour doter les hôtes de processus d'immunité innée basés sur l'interférence avec cette nucléobase. Au moins deux processus fondés sur le métabolisme de la cytosine sont apparus chez les animaux pour empêcher davantage la multiplication des virus, créant une immunité antivirale innée efficace. Les animaux – l'homme en particulier, mais cela s'étend même à l'huître – ont recruté une voie de biosynthèse dépendante de la S-Adénosyl-méthionine pour construire un analogue toxique de la CTP, la 3'-désoxy-3',4'-didéhydro-CTP (ddhCTP, *figure 2*), à l'aide d'une enzyme appelée vipérine (pour « virus inhibitory protein, endoplasmic reticulum-associated,



3'-désoxy-3',4'-didéhydro-CTP (ddhCTP)



N4-hydroxycytidine : tautomérisation

Figure 2 - Formule de la 3'-désoxy-3',4'-didéhydro-CTP et de la N4-hydroxycytidine. La ddhCTP, antimétabolite naturel de la défense antivirale innée, est produite par la vipérine. Des analogues de la N4-hydroxycytidine sont souvent choisis comme candidats antiviraux, mais si leurs propriétés chimiques les conduisent à produire de la N4-hydroxycytidine, la molécule sera immédiatement phosphorylée et conduira à un mutagène puissant. Le plus grand soin doit donc être apporté au choix des analogues de nucléosides utilisés comme antiviraux.

interferon-inducible»), qui répond à l'interféron gamma. Le rôle antiviral de la vipérine semble impliquer une variété de cibles, comme on peut s'y attendre en raison de son interférence avec le métabolisme de la cytosine. Comme nous l'avons vu, une cible probable est le métabolisme des lipides dépendant de la CTP, mais il semble que la ddhCTP puisse aussi interférer avec la transcription ou la réplication virale, agissant comme un terminateur de réplication.

Un deuxième processus antiviral impliquant la cytosine qui met l'infection virale en échec est la méthylation de la cytosine des dinucléotides CpG dans la séquence du virus. Chez la drosophile, il a été démontré que l'immunité innée antivirale utilise la méthylase Dnmt2 comme mécanisme efficace pour inactiver le virus C, à ARN positif. Cette enzyme pourrait aussi être impliquée dans la modification des hybrides ARN-ARN pendant la réplication du SARS-CoV-2, mais cela n'a pas été étudié. Enfin, un autre rôle unique des séquences CpG repose sur la reconnaissance de la séquence virale par la protéine à doigt de zinc antivirale (ZAP), qui se lie spécifiquement aux séquences CpG du virus, recrutant plusieurs systèmes de dégradation de l'ARN, et clive le virus en fragments inactifs [7]. Une conséquence de cette réponse antivirale est que beaucoup de dinucléotides CpG sont absents dans les génomes de nombreux virus à ARN. Cependant, le contexte

dans lequel les séquences CpG ont un rôle dans ce processus est critique, car la réponse antivirale n'est pas corrélée à leur abondance de manière simple.

Un plaidoyer pour des antiviraux

En résumé, le fait que les cellules aient sélectionné le métabolisme de la cytosine pour assurer la coordination de leur croissance non homothétique (cytoplasme, membranes et chromosomes) a créé une forte contrainte sur la disponibilité en cytosine pour la construction des virus à ARN. Cela se reflète directement dans la composition des coronavirus, et en particulier de SARS-CoV-2. En effet, ce virus est composé en moyenne de 30,2 % en adénine, 19,9 % en guanine, 32,4 % en uracile et seulement de 17,6 % en cytosine. On remarquera que dans un premier temps, la perte de cytosine dans le génome du virus lui a probablement permis de mieux échapper à la réponse antivirale de l'hôte. Mais on doit espérer que la perte continue de cytosine (et par conséquent de guanine) atténuera progressivement la virulence et la propagation du virus. Nous pouvons aussi comprendre qu'avoir découvert cette contrainte devrait permettre le développement de médicaments antiviraux efficaces. Il faut toutefois considérer les analogues de nucléosides avec la plus grande prudence ; ils sont souvent sérieusement mutagènes (et cancérigènes), comme l'est la N4-hydroxycytidine [8].

Cet article est le résumé d'une revue soumise pour publication.

- [1] Chen Y., Liu Q., Guo D., Emerging coronaviruses: genome structure, replication, and pathogenesis, *J. Med. Virol.*, **2020**, *92*, p. 418, doi: 10.1002/jmv.25681
- [2] Alonso-Carrera J., de Miguel C., Manzano B., Economic growth and environmental degradation when preferences are non-homothetic, *Environ. Resource Econ.*, **2019**, *74*, p. 1011, doi: 10.1007/s10640-019-00357-4
- [3] Hutchison C.A. et al., Design and synthesis of a minimal bacterial genome, *Science*, **2016**, *351*, aad6253, doi: 10.1126/science.aad6253
- [4] Del Caño-Ochoa F., Ramón-Maiques S., The multienzymatic protein CAD leading the de novo biosynthesis of pyrimidines localizes exclusively in the cytoplasm and does not translocate to the nucleus, *Nucleosides Nucleotides Nucleic Acids*, **2020**, doi: 10.1080/15257770.2019.1706743
- [5] Cheng M.-L. et al., Metabolic reprogramming of host cells in response to enteroviral infection, *Cells*, **2020**, *9*, 473, doi: 10.3390/cells9020473
- [6] Song H.-D. et al., Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human, *Proc. Natl Acad. Sci. USA*, **2005**, *102*, p. 2430, doi: 10.1073/pnas.0409608102
- [7] Luo X. et al., Molecular mechanism of RNA recognition by zinc-finger antiviral protein, *Cell Rep.*, **2020**, *30*, p. 46, doi: 10.1016/j.celrep.2019.11.116
- [8] Sledziewska E., Janion C., Mutagenic specificity of N4-hydroxycytidine, *Mutat. Res.*, **1980**, *70*, p. 11, doi: 10.1016/0027-5107(80)90053-6

Antoine DANCHIN^{1*} et **Philippe MARLIÈRE**², biologistes.

¹ Kodikos Labs, Lyon/Institut Cochin, Paris.

* antoine.danchin@normalesup.org

² TESSI (The European Syndicate of Synthetic Scientists and Industrialists), Paris.