environmental microbiology

Environmental Microbiology (2020) **22**(6), 1977–1985 doi:10.1111/1462-2920.15025

Hypothesis

Cytosine drives evolution of SARS-CoV-2

Antoine Danchin [1*] **and Philippe Marlière** [2]
[1]*Kodikos Labs, 24 rue Jean Baldassini, 69007 Lyon/ Institut Cochin, 75013 Paris, France.*
[2]*TESSSI, The European Syndicate of Synthetic Scientists and Industrialists, 81 rue Réaumur, 75002, Paris, France.*

Summary

In this article, we show, in the specific case of SARS-CoV-2, that the role of cytosine-based metabolites used as cell growth coordinators is central to understanding both innate antiviral immunity and the evolution of the virus.

An outbreak of atypical pneumonia, first noticed in Wuhan, Hubei province, China (Huang *et al*., 2020), developed into the COVID-19 pandemic (Jee, 2020) that displayed alarming similarities to symptoms caused by other β-coronaviruses (β-CoVs), SARS (Turinici and Danchin, 2007) and MERS (Bermingham *et al*., 2012) in particular. To contain the epidemic, efforts have been placed on diagnostic (Udugama *et al*., 2020), epidemiology (Park *et al*., 2020), vaccines (Zhang *et al*., 2020) and exploration of the use of old drugs to act as anti-coronaviral drugs (Du and Chen, 2020). Yet, all these endeavours underline an anthropocentric point of view that does not take properly into account the biology of the virus. We propose here that understanding at the molecular level how the virus multiplies and evolves in the metabolic context of its host may uniquely provide us with out-of-the-box solutions to fight the disease. To be sure, it is critical and urgent to be able to try and anticipate the future of the organism as it adapts to *Homo sapiens*.

The multiplication of β-CoVs requires unique interactions with the host cell. Upon internalization, the virus is stripped of its envelope and, mistaken for a messenger RNA, its positive sense RNA genome is immediately translated into an RNA-dependent RNA polymerase (replicase) and proteins that allow it to hijack relevant host functions. This involves an intricate series of events, beginning with replication of the virus into a complementary RNA template that serves to generate both new viral genomes and several individual transcripts of that template (Sawicki *et al*., 2007; Chen *et al*., 2020). As an immediate demand, the virus must access the pool of ribonucleoside triphosphates needed for the transcription of 50–100 copies of the replicated RNA strand produced at each multiplication cycle. This makes the viral sequence highly sensitive to the idiosyncrasies of nucleotide metabolism. We thus expect that the cell's nucleotide general makeup will shape virus evolution as it inevitably mutates when producing its large progeny. Coronaviruses have evolved a specific family of functions meant to overcome some of this limitation via a proofreading step coupled to the function of its RNA replicase (Sexton *et al*., 2016). Yet, mutations remain unavoidable and viruses, which generate a huge number of particles within a single patient, will progressively integrate the various types of selection pressure that each virus variant faces. Selection pressure via efficacy of transmission multiplied by number of replicates per cell and selection pressure via intracellular availability of essential precursors (nucleotides, lipids, amino acids, carbohydrates) create a variety of bottlenecks for viral evolution (Kutnjak *et al*., 2017; Arribas *et al*., 2018; Orton *et al*., 2020). A second key feature of β-CoVs is that they are enveloped. Furthermore, some proteins of the virion are glycosylated, which involves tapping into the cell's resources of UDP-sugars (Wellen and Thompson, 2012; Mayer *et al*., 2019). Besides this uracil metabolism-dependent protein-tagging feature, we focus here on the phospholipids of the membranes, which also derive from metabolites involving pyrimidines, specifically from CDP-containing liponucleotides (Kuo *et al*., 2016; Woods *et al*., 2016; Lee and Ridgway, 2020).

Briefly, we note that virus proliferation consists of reproduction of molecular sets whose chemical composition diverges grossly from the cell's average mRNA composition, already impacted by that of average nucleotide availability (Traut, 1994; Fig. 1). We highlight here the role of cytosine-based metabolites—and consequently that of guanine-based nucleotides—as critical coordinators of the

Received 12 April, 2020; revised XXXX; accepted 13 April, 2020.
*For correspondence. E-mail antoine.danchin@normalesup.org, Tel. +331 4441 2551, Fax +331 4441 2559

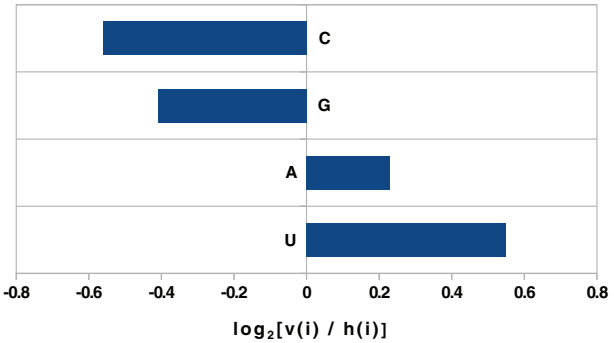

Fig. 1. Compositional biases human mRNA vs (+)SARS-CoV-2. The histogram of the compositional difference between *Homo sapiens* mRNAs and the SARS-CoV-2 genome shows a considerable counterselection for the presence of cytosine (and guanine) nucleotides. The figure plots the $\log_2$[v(i)/h(i)] ratio of the content in the virus over the content in human cytoplasmic mRNAs.

global cell metabolism. We also document its likely consequence on the evolution of the virus. By definition, metabolism encompasses all chemical reactions connecting a limited panel of molecules so as to convert nutrients into cells. As parasites, viruses—especially those with small genomes, such as RNA viruses—do not generally code for functions that result in the construction of metabolic pathways. As a consequence, the chemical constituents of viruses derive from precursors obtained from food through metabolic reaction steps, all but a few encoded in the host cell's genome. The question thus arises about the adequacy between the virus genome-encoded biocatalysts that perform all stages of its proliferation, on the one hand, and, on the other hand, the chemical composition of viruses as infectious vectors, whose total quantity is to be maximized for the virus to be successful. Here we (i) highlight the deviation of SARS-CoV-2 RNA chemical composition compared with that of its human host; (ii) formulate a hypothesis grounded on the canonical organization of cytosine metabolism as a way to coordinate non-homothetic growth of cells—i.e., the simultaneous growth of the cytoplasm (three dimensions), the membrane (two dimensions) and the genome (one dimension)—, and point out the emergence of the endogenous antinucleotide viperin as a cognate adaptive antiviral metabolite and (iii) predict evolutionary trends of CoV-2 for maximizing compositional fitness—which seem to show up in ongoing mutation survey of radiative evolution.

General features of the nucleotide build-up of SARS-CoV-2 in its cellular context

A striking feature of SARS-CoV-2 is that its genome appears to be depleted of cytosine nucleotides, pointing out an avoidance of cytosine as a most salient feature of its build-up. Because replication of the virus depends on a complementary RNA template, this deficiency is also reflected in a relative deficiency of guanine nucleotides. Indeed the virus is composed, on average, of 30.2% A, 19.9% G, 32.4% U and only 17.6% C (see also Fig. 1), probably reflecting the coupling between synthesis of viral particles and the host cell's metabolic capacity. Another noteworthy feature of the RNA sequence is that purines and pyrimidines are stoichiometrically balanced in the SARS-CoV-2(+) strand. This is interesting because the sequence does not follow the still enigmatic Chargaff's second parity rule, which would predict an equivalent amount of A and U as well as G and C (Forsdyke and Mortimer, 2000), and indicates that the virus is subjected to a metabolic balance equilibrating purines and pyrimidines. This is consistent with a study that used Flux Balance Analysis (FBA) that reviewed nucleotide stoichiometric availability (stoichiometric constraints as illustrated by Palsson and co-workers, for example, Schilling *et al*., 1999), and suggested guanylate kinase as a critical bottleneck in the build-up of the viral genome during infection (https://csbnc.informatik.uni-tuebingen.de/index.php/s/jd8rNcBJsmigFkz). When we compared the distribution of the four ribonucleotides in the human messenger RNAs with that in SARS-CoV-2, it appeared that cytosine was the rarest nucleotide of (+) sense virus genomes (exceptions are discussed in the Perspectives section), in a proportion considerably lower than in human mRNAs (not to mention tRNAs, rRNAs). Cytosine deficiency was matched by guanine deficiency (Fig. 1), as expected from the fact that the propagation of the virus results from a replication process incorporating complementary nucleotides. This process is amplified in the form of 50–100 copies, in a highly unsymmetrical operation, but this does not affect the pressure on the final guanine content.

Cytosine as an integrator for non-homothetic growth

To understand how viruses tap in the host's metabolic resources, allowing non-homothetic growth in a stable manner as cells multiply, it is critical to understand how the construction of the cell's building blocks is coordinated to allow matching the growth of its cytoplasm (three dimensions), with the growth of its envelope (two dimensions) and the growth of its genetic setup (one dimension). We propose here that the selection of a specific subset of intermediary metabolites took place in the course of evolution as a way to smooth out non-homothetic growth. Briefly, among a variety of alternatives, cytosine nucleotides ended up as the coordinating metabolites, tying up growth of the cell's genome and of its membrane to central metabolism (Fig. 2). To be sure, de novo synthesis of nucleotides allows direct production of all triphosphates, including cytidine triphosphate (CTP). Nevertheless, there might remain

some persisting negative trend against C, because CTP derives from uridine triphosphate (UTP) in a step that both requires adenosine triphosphate (ATP) and a nitrogen source, which makes availability of the molecule highly sensitive to energy and nitrogen availability. However, this type of negative pressure would equally apply to synthesis of ATP, the most abundant nucleotide in the cell (Zhang *et al*., 2008). This suggests that the organization of anabolic processes is not suitable per se to modulate the stoichiometry of specific nucleoside triphosphates (NTPs) according to a well-defined pattern. Then, what about their degradation and salvage?

At some point in all life cycles, metabolites are damaged then repaired or degraded then recycled, either as a whole or as parts. In the purine nucleotide salvage pathway, the key enzyme is adenine phosphoribosyltransferase (Wilson *et al*., 1986). In the same way, guanine is salvaged via hypoxanthine–guanine phosphoribosyltransferase (Balendiran *et al*., 1999). In pyrimidine salvage, uracil is recycled using uracil phosphoribosyltransferase (Li *et al*., 2007; blue arrow in Fig. 2). Yet, contrary to expectation, this does not go at all this way with cytosine. Surprisingly, no cytosine phosphoribosyltransferase has been identified in any living organism. For some reason, natural selection avoided retaining a direct route for cytosine salvage, despite that it does not seem difficult to evolve enzyme variants which would catalyse the reaction. The cytosine recycling process begins with converting cytosine-containing metabolites to cytidine then uridine, or cytosine to uracil, which is mainly scavenged directly by uracil phosphoribosyltransferase. Everything goes as if salvage of cytosine nucleotides had to go through deamination of cytosine-containing metabolites to uracil-containing molecules, followed by neosynthesis of UTP- and ATP-dependent amidation to CTP by CTP synthase (PyrG). This unique step is so essential that the *pyrG* gene is conserved in parasites (Aurrecoechea *et al*., 2009) and even in the smallest streamlined genome of an autonomous synthetic construct (Hutchison *et al*., 2016). What's more, the unique features of cytosine metabolism do not end up there, as CTP synthase displays a very unusual architecture. It forms filaments, named cytoophidia, in all organisms where its organization has been explored (Li *et al*., 2018). Cytoophidia create a strict compartmentalization of the corresponding activity (Liu, 2010; Sun and Liu, 2019). This organization is linear and therefore satisfies the bottom-up level constraints of non-homothetic growth (i.e., 1-D growth is even more constrained than 2-D growth, which itself is more constrained than 3-D growth), placing CTP uniquely at the crossroad of global metabolic controls.

Correlated with 2-D growth, the synthesis of membrane lipids generates yet a further critical involvement of CTP-dependent metabolism for controlling non-homothetic cell

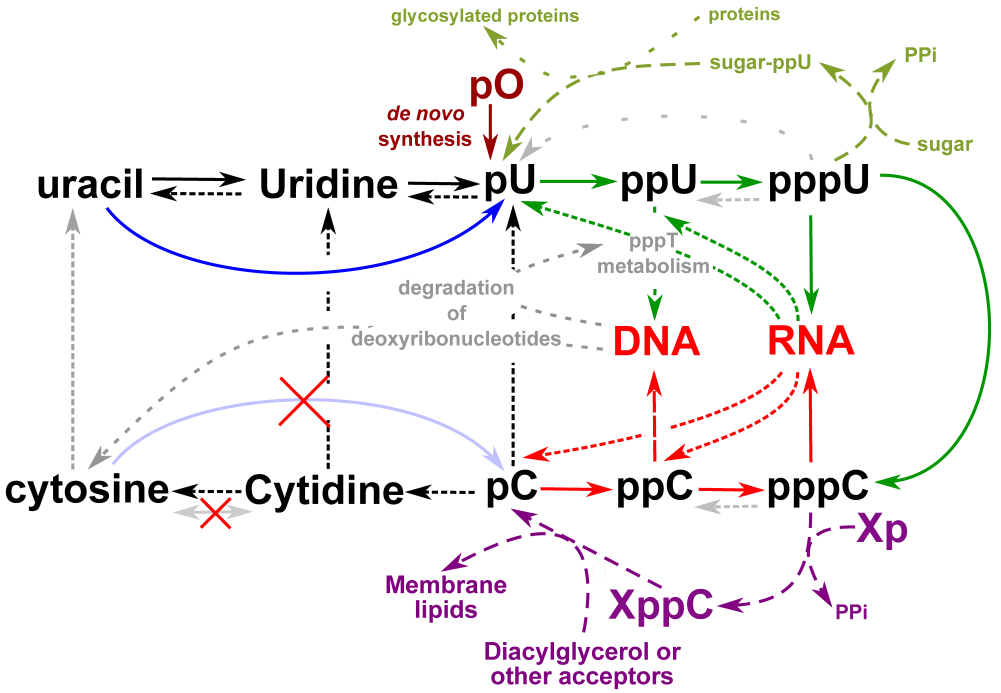

Fig. 2. Synthesis and salvage of pyrimidines.
In green, the central pyrimidine pathway for biosynthesis. The specific pathway for cytosine derivatives is in red. The dotted arrows mark catabolism. Additional routes are shown with dashed lines: purple for lipid synthesis and green for protein glycosylation. In blue we have shown the key enzyme which allows the recycling of pyrimidines via uracil. The parallel pathway for cytosine has not been discovered in any organism to date. [Correction statement added on 08 May 2020 after first online publication: Figure 2 has been corrected in this version.]

growth. As a matter of fact, the double layer of phospholipids forming most membranes derives from cytosine-based liponucleotides (the *raison d'être* of cytosine-specific nucleotides in this makeup has not been investigated previously, but control of non-homothetic growth makes it a brilliant choice). This involves a variety of pathways initiated by CDP-diglycerides or chemically related analogues (Chauhan *et al*., 2016; McMaster, 2018). While the membrane lipid composition differs in the three domains of life, the general organization of the cognate pathways is similar, using CTP-dependent enzymes not only to form the lipid bilayer of the membrane but also to control its shape via its curvature (Cornell and Ridgway, 2015). In a metabolic development important in the present context where innate immunity must play a central role (Di Conza and Ho, 2020), CTP-dependent phospholipid synthesis is also essential for the generation of the endoplasmic reticulum (ER; Lagace and Ridgway, 2013).

Finally, as now can be expected because of the importance of cytosine-based metabolites, the need for a specific organization of pyrimidine metabolism is implemented in further specific structures. As a case in point documented in many animal cells, the enzymes required for the onset of the de novo synthesis of pyrimidines, carbamoyl-phosphate synthetase (*C*PSase), aspartate transcarbamylase (*A*TCase, and dihydroorotase (*D*HOase) make a multifunctional structure, known as CAD. All three activities are supported by a single 243-kDa polypeptide that forms hexamers and higher oligomers (Del Caño-Ochoa and Ramón-Maiques, 2020). Remarkably, this multifunctional enzyme is sensitive to proteolysis by caspase during apoptosis, indicating that pyrimidine metabolism is involved in this critical process (Huang *et al*., 2002). Further in line with the metabolic pathways organization explored here in relation with viral infection, CAD is highly expressed in leukocytes, where it enables Toll-like receptor 8 expression in response to cytidine and single-stranded RNA (Furusho *et al*., 2019). It was also observed early on that an increase in cytoplasmic CTP accelerated the rate of phospholipids synthesis in poliomyelitis-infected cells (Choy *et al*., 1980). Yet another most significant feature of CAD, relevant to our hypotheses, is that its activity is modulated by a *dedicated viral protein* during enteroviral infection (Cheng *et al*., 2020). This should prompt further analysis of pyrimidine metabolism in relation with SARS-CoV-2 infection, in particular looking for virus-encoded functions involved in interference with cytosine metabolism, as we now document.

Back to coronaviruses and the cytosine-sensitivity of their construction, virulence and evolution

The requirement of cytosine nucleotides for the virus' genome and envelope synthesis is inevitably connected to the striking limitation in synthesis of cytosine-based metabolites. The most straightforward consequence of this metabolic design is that there is a force that will keep driving the cytosine content of RNAs to lower values, unless opposing processes (and selection pressure leading to discard organisms with too low cytosine content, for example, because this would create intolerable biases in the amino acid composition of the proteins coded by these genomes) had the upper hand during evolution. Coronaviruses, and other positive-sense single-stranded RNA viruses, produce plus strands at a 50- to 100-fold excess of their minus-strand replicated template. A further consequence of the replication process, however unsymmetrical, is that any pressure on a given base availability—here C—would affect its complement—G in our case—as noted above (Fig. 1). The mutations ordained to occur as the virus evolves will therefore reflect both the physicochemical forces acting during replication (typically triggered by cytosine deamination and reactions involving reactive species, resulting in formation of 8-oxoguanine, for example) and the metabolic setup of the host. As discussed in the previous sections, we expect a general selection pressure operating on CTP and tending, in the long run, to decrease the C content of the RNA virus. This is certainly qualified, however, by direct selection pressure on the functions that drive virus replication and propagation and operate on the corresponding codons, hence proteins. This is particularly important for the proline residue, encoded by CCN codons and essential in the folding of key viral protein domains (Li *et al*., 2014). The presence of a four codon insertion in the spike protein of SARS-CoV-2 is a case in point (Li *et al*., 2020). The C at the second codon position is also required to allow introduction of threonine or alanine residues in the viral proteins, while the first position is required for histidine and glutamine coding. In this context, it seems relevant to note that, in the SARS-CoV-1 in 2003, one of the critical changes with respect to innocuous counterparts was a leucine to alanine change at the junction between domains S1 and S2 of the spike protein and that this required a U to C change (Song *et al*., 2005).

Now comes an extraordinary feature that accounts for the way animals control RNA virus diseases via innate immunity. It happens that the unique role of cytosine as a coordinator of global metabolism has been exploited by natural selection to endow hosts with antiviral processes based on interference with the metabolic involvement of this nucleobase. At least three responses have evolved in animals to further prevent virus multiplication, building up an efficient innate antiviral immunity based on cytosine metabolism. Animals, man in particular but this extends even to oysters (Green *et al*., 2014), have recruited an AdoMet-dependent biosynthesis pathway to construct a mimic of CTP, 3′-deoxy-3′,4′-didehydro-CTP (ddhCTP, Fig. 3), using an enzyme named viperin, for *v*irus inhibitory *p*rotein, ER-associated, interferon-inducible (Duschene and Broderick,

2010; Gizzi *et al*., 2018), that responds to interferon gamma (Chin and Cresswell, 2001; Zhang *et al*., 2007). The exact antiviral role of viperin appears to involve a variety of targets, as indeed expected from interference with cytosine metabolism. As hinted at the end of the previous section, an obvious target is CTP-dependent lipid metabolism (Seo and Cresswell, 2013) but it also appears that it can, directly or via depletion of the cellular nucleotide pools (Ebrahimi *et al*., 2020), also interfere with viral transcription or replication, acting as a replication chain terminator (Fang *et al*., 2016; Dukhovny *et al*., 2018; Wei *et al*., 2018).

The second antiviral process involving cytosine that puts viral infection in check is methylation of the cytosine of CpG dinucleotides in the viral sequence. In *Drosophila*, it has been shown that antiviral innate immunity uses methylase Dnmt2 as an efficient factor to inactivate the *Drosophila* positive-sense RNA C virus (Durdevic *et al*., 2013). This methylase, often annotated as a DNA methyltransferase, is known to modify a variety of small RNA molecules on the cytosine of CpG dinucleotides present in specific contexts. Its major role is in the modification of tRNA molecules next to the anticodon (Thiagarajan *et al*., 2011; Dev *et al*., 2017; Bohnsack *et al*., 2019). Whether this extends to the viral RNA genome or to its transcripts, or whether it uses modified tRNAs as allosteric regulators, is not known. However there is *in vitro* evidence that Dnmt2 can efficiently methylate DNA when DNA fragments are presented as covalent DNA–RNA hybrids in the structural context of tRNA genes (Kaiser *et al*., 2017). This might also affect RNA–RNA hybrids during SARS-CoV-2 replication, but this has not been investigated. It may be relevant that the loops in stem-loop sequences SL-2 and SL-3 of the viral 5′ region that are necessary for replication of the virus contain CpG motifs at the stem-loop junction (Chen and Olsthoorn, 2010). As a third role of cytosine-related functions, another unique involvement of CpG sequences rests in recognition of the viral sequence by the zinc-finger antiviral protein (ZAP; Zhu *et al*., 2020), which binds specifically to CpG viral sequences, recruiting multiple RNA degradation machines to degrade target viral RNA (Luo *et al*., 2020). As a consequence of this antiviral response, CpG dinucleotides are suppressed in the genomes of many RNA viruses. However, the context where CpG sequences have a role in this process is critical, as the antiviral response is not correlated to their abundance in any straightforward manner (Ficarelli *et al*., 2020). The targeting of cytosine functions thus seems to be key to innate immunity toward RNA viral infection.

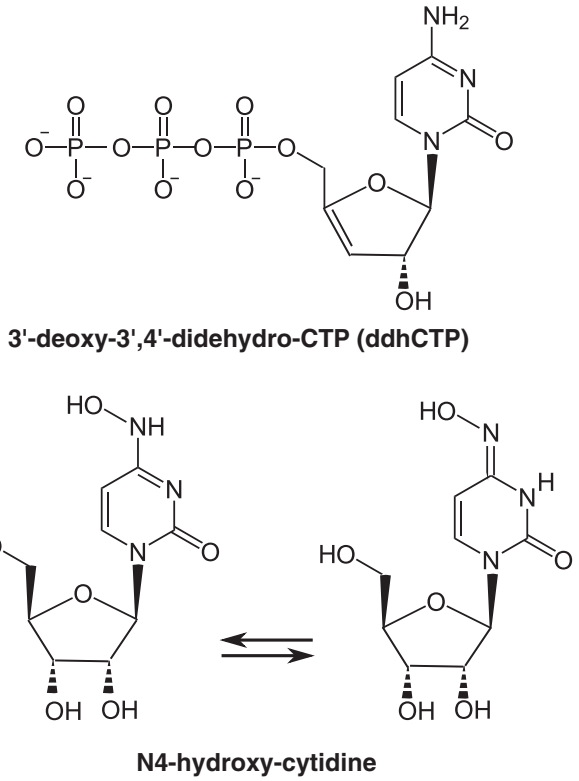

Fig. 3. Viperin-derived ddhCTP and N4-hydroxy-cytidine.
ddhCTP is the natural antimetabolite of the innate antiviral defence metabolite produced by viperin. N4-hydroxy-cytidine analogues are often chosen as antiviral candidates, but if their chemical properties lead them to produce N4-hydroxy-cytidine, the molecule will be immediately phosphorylated and lead to a potent mutagen. Great care must therefore be taken when choosing the nucleoside analogues used as antivirals.

Perspectives

Here, we reviewed the role of a unique role of cytosine-related metabolism as the master coordinator of non-homothetic cell growth. Viruses tap into the cell's resources, and this original setup of intermediary metabolism creates an intracellular chemical pressure that must constrain the evolution of the genome sequence of RNA viruses, in particular SARS-CoV-2. Because the RNA(+) virus sequence is replicated into a (−) template, any pressure on the content in cytosine will be indirectly passed on its guanine content, as indeed observed. Interestingly, a role of guanylate kinase as a possible target for antiviral drugs has been identified by FBA (Schilling *et al*., 1999) as a unique step whose suppression would disable SARS-CoV-2 biosynthesis without impending host cell biosyntheses (https://csbnc.informatik.uni-tuebingen.de/index.php/s/jd8rNcBJsmigFkz).

Our working hypothesis is that the availability of CTP (and hence of cytosine-based precursors) is a dominant driving force in the way RNA viruses evolve a new progeny. The progressive build-up, during evolution, of antiviral immunity steps that are based on the presence of cytosine in viral genomes supports this view considerably. We note however that, while in the long term this

trend toward a loss of C residues should lead to attenuation, it may have been used as a natural mechanism for the virus to escape innate immunity, at least during the first steps of evolution of the pathogens. Indeed, the very fact that the genomes lose progressively their cytosines, in particular in CpG dinucleotides, allow them to escape the ZAP and Dnmt2 innate immunity, while less C will also give less room for ddhCTP to interfere with replication. However the latter effect is probably less of a concern because ddhCTP will still interfere with the construction of the viral envelope. In the long term, however, a progressive loss of C in the genome will deplete the proteins it encodes in alanine, histidine, glutamine, proline and threonine, thus narrowing considerably the evolutionary landscape of the pathogen.

Based on a loose analogy with computer viruses, which hack operating systems by tapping into "system resources," yet another feature critical to anticipate the ultimate evolution of RNA viruses is as follows. Natural selection is based on the existence of functions that are a posteriori found to benefit the future of any species. When a biochemical function is implemented—here the control of nucleotide metabolism via cytosine nucleotides synthesis and salvage—its nullification is open to positive selection. This implies that we should look for RNA viruses that have overcome this limitation and look for guanine–cytosine (GC)-rich RNA viruses. Some exist, as the rubella viruses, which have genomes in the 70% GC range (Zhu *et al*., 2016). There is no straightforward indication that they code for systems that interfere with CTP metabolism. However, despite the limited length of their genome (approximately one third of that of coronaviruses), the 3′ half of their genome codes for unknown functions associated to proteins of the capsid—hence immediately available upon infection—that have been shown to interact with the host cell's machinery in regions certainly connected to metabolism (Zheng and Kielian, 2013). An obvious experimental program derived from the present hypothesis would be, therefore, to set up experiments to explore this possibility. This would considerably improve our knowledge of the way viruses develop.

We are well aware of the limitations of the present study, due to the small number of samples and fairly short time span of the present epidemic. However we feel that, in view of the urgent situation we are facing with the COVID-19 pandemic, it is important to communicate our observations while relating them to previously unrecognized pressure that must have considerable importance in the evolution of viruses and in the design or new treatments.

A cautionary note: Primum non nocere

The emergence of ddhCTP as an antiviral selected by evolution as an innate immunity weapon shows us a path to follow. We should however take great caution in the development of nucleoside analogues as therapeutic agents. The effectiveness of such molecules, sometimes called antinucleosides, may come from the fact that they inhibit directly the activity of a viral enzyme, for example by blocking the activity of replicases. A completely different scenario arises when an antinucleoside is converted by metabolic enzymes of the host into a substrate of viral enzymes, such as a chemically altered analogue of the NTPs, CTP or ATP, for example. The incorporation by the replicase of the antinucleoside into the RNA of the virus will stop its proliferation, as anticipated. This could be by interrupting the elongation of the replicated RNA chain, we then speak of a "chain terminator." As discussed above, this is one of the antiviral mechanisms observed for the viperin product. Alternatively, this could result from incorporating ambivalent nucleobases into the viral genome, allowing pairing with more than one canonical base. Such ambivalent antinucleosides literally lead to blurring the genetic message of the virus while it replicates, as we have learned from the use of ribavirin, which requires its activation into a triphosphate to sabotage the proliferation of riboviruses (Tanaka *et al*., 2019). In the same way, the mechanism of action of favipiravir is understood to combine the two effects of pairing ambivalence and chain termination once it is converted into its nucleoside triphosphate counterpart (Furuta *et al*., 2017).

Yet, once input in the viral genome, these analogues will be readily converted to nucleoside monophosphates and diphosphates which will be processed by the host cell's metabolism. The diphosphate analogues will be converted, via the enzyme ribonucleoside diphosphate reductase, into deoxyribonucleoside diphosphates, then triphosphates, and get into the cell's genome (Fig. 2). This creates a situation where agents of these antiviral families will lead to alterations in the human DNA, i.e., accumulate somatic or germinal mutations leading to cancer, foetal malformations or hereditary diseases, all conditions that cannot be observed in the short term. Alas, the standard course of clinical trials does not easily involve long-term observations, especially in times of emergency. As a consequence, we strongly recommend that mutagenesis tests be carefully performed during the validation procedures required for identification of antivirals directed against SARS-CoV-2. For example, the bacteriological literature is replete with data demonstrating the highly mutagenic nature of N4-hydroxy-cytidine (Sledziewska and Janion, 1980; Fig. 3, as well as homologous structures N4-methoxy-cytidine, N4-amino-cytidine, etc.) via their conversion into its deoxynucleoside triphosphate counterpart. Health authorities should therefore be aware of the long-term risk of compounds capable of delivering these mutagenic nucleosides, such as EIDD-2801 (Sheahan *et al*. 2020), regardless of their antiviral efficacy.

Acknowledgements

This work benefited from discussions with members of the Stanislas Noria seminar and from critical comments of Pierre-Yves Bourguignon, Félix Rey and Ken Timmis. AD is CSO of the company Kodikos Labs, and PhM is president of TESSSI and director of TheraXen SA.

Note added in proof

An article supporting the role of CpG dinucleotides in the evolution of SARS-CoV-2 appeared after the present article was submitted for publication: Xuhua Xia Extreme genomic CpG deficiency in SARS-CoV-2 and evasion of host antiviral defense. *Molecular Biology and Evolution*, msaa094, https://doi.org/10.1093/molbev/msaa094

References

Arribas, M., Aguirre, J., Manrubia, S., and Lázaro, E. (2018) Differences in adaptive dynamics determine the success of virus variants that propagate together. *Virus Evol* **4**: vex043.

Aurrecoechea, C., Brestelli, J., Brunk, B.P., Carlton, J.M., Dommer, J., Fischer, S., *et al*. (2009) GiardiaDB and TrichDB: integrated genomic resources for the eukaryotic protist pathogens *Giardia lamblia* and *Trichomonas vaginalis*. *Nucleic Acids Res* **37**: D526–D530.

Balendiran, G.K., Molina, J.A., Xu, Y., Torres-Martinez, J., Stevens, R., Focia, P.J., *et al*. (1999) Ternary complex structure of human HGPRTase, pRpp, Mg^{2+}, and the inhibitor HPP reveals the involvement of the flexible loop in substrate binding. *Protein Sci* **8**: 1023–1031.

Bermingham, A., Chand, M.A., Brown, C.S., Aarons, E., Tong, C., Langrish, C., *et al*. (2012) Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill* **17**: 20290.

Bohnsack, K.E., Höbartner, C., and Bohnsack, M.T. (2019) Eukaryotic 5-methylcytosine (m^5C) RNA methyltransferases: mechanisms, cellular functions, and links to disease. *Genes (Basel)* **10**: E102.

Chauhan, N., Farine, L., Pandey, K., Menon, A.K., and Bütikofer, P. (2016) Lipid topogenesis—35 years on. *Biochim Biophys Acta* **1861**: 757–766.

Chen, S.-C., and Olsthoorn, R.C.L. (2010) Group-specific structural features of the 5′-proximal sequences of coronavirus genomic RNAs. *Virology* **401**: 29–41.

Chen, Y., Liu, Q., and Guo, D. (2020) Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* **92**: 418–423.

Cheng, M.-L., Chien, K.-Y., Lai, C.-H., Li, G.-J., Lin, J.-F., and Ho, H.-Y. (2020) Metabolic reprogramming of host cells in response to enteroviral infection. *Cell* **9**: E473.

Chin, K.C., and Cresswell, P. (2001) Viperin (cig5), an IFN-inducible antiviral protein directly induced by human cytomegalovirus. *Proc Natl Acad Sci U S A* **98**: 15125–15130.

Choy, P.C., Paddon, H.B., and Vance, D.E. (1980) An increase in cytoplasmic CTP accelerates the reaction catalyzed by CTP:phosphocholine cytidylyltransferase in poliovirus-infected HeLa cells. *J Biol Chem* **255**: 1070–1073.

Cornell, R.B., and Ridgway, N.D. (2015) CTP:phosphocholine cytidylyltransferase: function, regulation, and structure of an amphitropic enzyme required for membrane biogenesis. *Prog Lipid Res* **59**: 147–171.

Del Caño-Ochoa, F., and Ramón-Maiques, S. (2020) 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*. https://doi.org/10.1080/15257770.2019.1706743.

Dev, R.R., Ganji, R., Singh, S.P., Mahalingam, S., Banerjee, S., and Khosla, S. (2017) Cytosine methylation by DNMT2 facilitates stability and survival of HIV-1 RNA in the host cell during infection. *Biochem J* **474**: 2009–2026.

Di Conza, G., and Ho, P.-C. (2020) ER stress responses: an emerging modulator for innate immunity. *Cell* **9**: E695.

Du, Y.-X., and Chen, X.-P. (2020) Favipiravir: pharmacokinetics and concerns about clinical trials for 2019-nCoV infection. *Clin Pharmacol Ther* (in press). https://doi.org/10.1002/cpt.1844.

Dukhovny, A., Shlomai, A., and Sklan, E.H. (2018) The antiviral protein Viperin suppresses T7 promoter dependent RNA synthesis-possible implications for its antiviral activity. *Sci Rep* **8**: 8100.

Durdevic, Z., Hanna, K., Gold, B., Pollex, T., Cherry, S., Lyko, F., and Schaefer, M. (2013) Efficient RNA virus control in drosophila requires the RNA methyltransferase Dnmt2. *EMBO Rep* **14**: 269–275.

Duschene, K.S., and Broderick, J.B. (2010) The antiviral protein viperin is a radical SAM enzyme. *FEBS Lett* **584**: 1263–1267.

Ebrahimi, K.H., Howie, D., Rowbotham, J., McCullagh, J., Armstrong, F., and James, W.S. (2020) Viperin, through its radical-SAM activity, depletes cellular nucleotide pools and interferes with mitochondrial metabolism to inhibit viral replication. *FEBS Lett* (in press). https://doi.org/10.1002/1873-3468.13761.

Fang, J., Wang, H., Bai, J., Zhang, Q., Li, Y., Liu, F., and Jiang, P. (2016) Monkey viperin restricts porcine reproductive and respiratory syndrome virus replication. *PLoS One* **11**: e0156513.

Ficarelli, M., Antzin-Anduetza, I., Hugh-White, R., Firth, A.E., Sertkaya, H., Wilson, H., *et al*. (2020) CpG dinucleotides inhibit HIV-1 replication through zinc finger antiviral protein (ZAP)-dependent and -independent mechanisms. *J Virol* **94**: e01337-19.

Forsdyke, D.R., and Mortimer, J.R. (2000) Chargaff's legacy. *Gene* **261**: 127–137.

Furusho, K., Shibata, T., Sato, R., Fukui, R., Motoi, Y., Zhang, Y., *et al*. (2019) Cytidine deaminase enables toll-like receptor 8 activation by cytidine or its analogs. *Int Immunol* **31**: 167–173.

Furuta, Y., Komeno, T., and Nakamura, T. (2017) Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proc Jpn Acad Ser B Phys Biol Sci* **93**: 449–463.

Gizzi, A.S., Grove, T.L., Arnold, J.J., Jose, J., Jangra, R.K., Garforth, S.J., *et al*. (2018) A naturally occurring antiviral ribonucleotide encoded by the human genome. *Nature* **558**: 610–614.

Green, T.J., Benkendorff, K., Robinson, N., Raftos, D., and Speck, P. (2014) Anti-viral gene induction is absent upon secondary challenge with double-stranded RNA in the Pacific oyster, *Crassostrea gigas*. *Fish Shellfish Immunol* **39**: 492–497.

Huang, M., Kozlowski, P., Collins, M., Wang, Y., Haystead, T.A., and Graves, L.M. (2002) Caspase-dependent cleavage of carbamoyl phosphate synthetase II during apoptosis. *Mol Pharmacol* **61**: 569–577.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., *et al*. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**: 497–506.

Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., *et al*. (2016) Design and synthesis of a minimal bacterial genome. *Science* **351**: aad6253.

Jee, Y. (2020) WHO international health regulations emergency committee for the COVID-19 outbreak. *Epidemiol Health* **42**: e2020013.

Kaiser, S., Jurkowski, T.P., Kellner, S., Schneider, D., Jeltsch, A., and Helm, M. (2017) The RNA methyltransferase Dnmt2 methylates DNA in the structural context of a tRNA. *RNA Biol* **14**: 1241–1251.

Kuo, L., Koetzner, C.A., and Masters, P.S. (2016) A key role for the carboxy-terminal tail of the murine coronavirus nucleocapsid protein in coordination of genome packaging. *Virology* **494**: 100–107.

Kutnjak, D., Elena, S.F., and Ravnikar, M. (2017) Time-sampled population sequencing reveals the interplay of selection and genetic drift in experimental evolution of potato virus Y. *J Virol* **91**: e00690-17.

Lagace, T.A., and Ridgway, N.D. (2013) The role of phospholipids in the biological activity and structure of the endoplasmic reticulum. *Biochim Biophys Acta* **1833**: 2499–2510.

Lee, J., and Ridgway, N.D. (2020) Substrate channeling in the glycerol-3-phosphate pathway regulates the synthesis, storage and secretion of glycerolipids. *Biochim Biophys Acta Mol Cell Biol Lipids* **1865**: 158438.

Li, J., Huang, S., Chen, J., Yang, Z., Fei, X., Zheng, M., *et al*. (2007) Identification and characterization of human uracil phosphoribosyltransferase (UPRTase). *J Hum Genet* **52**: 415–422.

Li, Y., Surya, W., Claudine, S., and Torres, J. (2014) Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *J Biol Chem* **289**: 12535–12549.

Li, H., Ye, F., Ren, J.-Y., Wang, P.-Y., Du, L.-L., and Liu, J.-L. (2018) Active transport of cytoophidia in *Schizosaccharomyces pombe*. *FASEB J* **32**: 5891–5898.

Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B.T., and Chaillon, A. (2020) Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *J Med Virol* (in press). https://doi.org/10.1002/jmv.25731

Liu, J.-L. (2010) Intracellular compartmentation of CTP synthase in drosophila. *J Genet Genomics* **37**: 281–296.

Luo, X., Wang, X., Gao, Y., Zhu, J., Liu, S., Gao, G., and Gao, P. (2020) Molecular mechanism of RNA recognition by zinc-finger antiviral protein. *Cell Rep* **30**: 46–52 e4.

Mayer, K.A., Stöckl, J., Zlabinger, G.J., and Gualdoni, G.A. (2019) Hijacking the supplies: metabolism as a novel facet of virus-host interaction. *Front Immunol* **10**: 1533.

McMaster, C.R. (2018) From yeast to humans - roles of the Kennedy pathway for phosphatidylcholine synthesis. *FEBS Lett* **592**: 1256–1272.

Orton, R.J., Wright, C.F., King, D.P., and Haydon, D.T. (2020) Estimating viral bottleneck sizes for FMDV transmission within and between hosts and implications for the rate of viral evolution. *Interface Focus* **10**: 20190066.

Park, M., Cook, A.R., Lim, J.T., Sun, Y., and Dickens, B.L. (2020) A systematic review of COVID-19 epidemiology based on current evidence. *J Clin Med* **9**: E967.

Sawicki, S.G., Sawicki, D.L., and Siddell, S.G. (2007) A contemporary view of coronavirus transcription. *J Virol* **81**: 20–29.

Schilling, C.H., Edwards, J.S., and Palsson, B.O. (1999) Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol Prog* **15**: 288–295.

Seo, J.-Y., and Cresswell, P. (2013) Viperin regulates cellular lipid metabolism during human cytomegalovirus infection. *PLoS Pathog* **9**: e1003497.

Sexton, N.R., Smith, E.C., Blanc, H., Vignuzzi, M., Peersen, O.B., and Denison, M.R. (2016) Homology-based identification of a mutation in the coronavirus RNA-dependent RNA polymerase that confers resistance to multiple mutagens. *J Virol* **90**: 7415–7428.

Sheahan, T.P., Sims, A.C., Zhou, S., Graham, R.L., Pruijssers, A.J., Agostini, M.L., *et al.* (2020) An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Sci Transl Med* 06 Apr 2020: eabb5883.

Sledziewska, E., and Janion, C. (1980) Mutagenic specificity of N4-hydroxycytidine. *Mutat Res* **70**: 11–16.

Song, H.-D., Tu, C.-C., Zhang, G.-W., Wang, S.-Y., Zheng, K., Lei, L.-C., *et al*. (2005) Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci U S A* **102**: 2430–2435.

Sun, Z., and Liu, J.-L. (2019) mTOR-S6K1 pathway mediates cytoophidium assembly. *J Genet Genomics* **46**: 65–74.

Tanaka, Y., Inoue, A., Mizunuma, T., Matsumura, H., Yokomori, H., Komiyama, T., and Otori, K. (2019) Tolerability of erythrocyte ribavirin triphosphate concentrations depends on the ITPA genotype. *Ther Drug Monit* **41**: 497–502.

Thiagarajan, D., Dev, R.R., and Khosla, S. (2011) The DNA methyltranferase Dnmt2 participates in RNA processing during cellular stress. *Epigenetics* **6**: 103–113.

Traut, T.W. (1994) Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem* **140**: 1–22.

Turinici, G., and Danchin, A. (2007) The SARS case study: an alarm clock? In *Encyclopedia of Infectious Diseases*, Tibayrenc, M. (ed). Hoboken, NJ: John Wiley & Sons, Inc., pp. 151–162.

Udugama, B., Kadhiresan, P., Kozlowski, H.N., Malekjahani, A., Osborne, M., Li, V.Y.C., *et al*. (2020) Diagnosing COVID-19: the disease and tools for detection. *ACS Nano* (in press). https://doi.org/10.1021/acsnano.0c02624

Wei, C., Zheng, C., Sun, J., Luo, D., Tang, Y., Zhang, Y., *et al*. (2018) Viperin inhibits enterovirus A71 replication by interacting with viral 2C protein. *Viruses* **11**: E13.

Wellen, K.E., and Thompson, C.B. (2012) A two-way street: reciprocal regulation of metabolism and signalling. *Nat Rev Mol Cell Biol* **13**: 270–276.

Wilson, J.M., O'Toole, T.E., Argos, P., Shewach, D.S., Daddona, P.E., and Kelley, W.N. (1986) Human adenine

phosphoribosyltransferase. Complete amino acid sequence of the erythrocyte enzyme. *J Biol Chem* **261**: 13677–13683.

Woods, P.S., Doolittle, L.M., Rosas, L.E., Joseph, L.M., Calomeni, E.P., and Davis, I.C. (2016) Lethal H1N1 influenza a virus infection alters the murine alveolar type II cell surfactant lipidome. *Am J Physiol Lung Cell Mol Physiol* **311**: L1160–L1169.

Zhang, Y., Burke, C.W., Ryman, K.D., and Klimstra, W.B. (2007) Identification and characterization of interferon-induced proteins that inhibit alphavirus replication. *J Virol* **81**: 11246–11255.

Zhang, Y., Morar, M., and Ealick, S.E. (2008) Structural biology of the purine biosynthetic pathway. *Cell Mol Life Sci* **65**: 3699–3724.

Zhang, J., Zeng, H., Gu, J., Li, H., Zheng, L., and Zou, Q. (2020) Progress and prospects on vaccine development against SARS-CoV-2. *Vaccines (Basel)* **8**: E153.

Zheng, Y., and Kielian, M. (2013) Imaging of the alphavirus capsid protein during virus replication. *J Virol* **87**: 9579–9589.

Zhu, Z., Chen, M., Abernathy, E., Icenogle, J., Zhou, S., Wang, C., *et al*. (2016) Analysis of complete genomes of the rubella virus genotypes 1E and 2B which circulated in China, 2000–2013. *Sci Rep* **6**: 39025.

Zhu, M., Zhou, J., Liang, Y., Nair, V., Yao, Y., and Cheng, Z. (2020) CCCH-type zinc finger antiviral protein mediates antiviral immune response by activating T cells. *J Leukoc Biol* **107**: 299–307.