THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCL

Letter

Theoretical Analysis Reveals the Cost and Benefit of Proofreading in Coronavirus Genome Replication

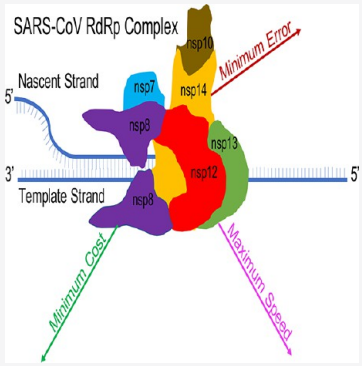
Joel D. Mallory, Xian F. Mallory, Anatoly B. Kolomeisky,* and Oleg A. Igoshin*

Cite This: *J. Phys. Chem. Lett.* 2021, 12, 2691−2698

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Severe acute respiratory syndrome coronaviruses have unusually large RNA genomes replicated by a multiprotein complex containing an RNA-dependent RNA polymerase (RdRp). Exonuclease activity enables the RdRp complex to remove wrongly incorporated bases via proofreading, a process not utilized by other RNA viruses. However, it is unclear why the RdRp complex needs proofreading and what the associated trade-offs are. Here we investigate the interplay among the accuracy, speed, and energetic cost of proofreading in the RdRp complex using a kinetic model and bioinformatics analysis. We find that proofreading nearly optimizes the rate of functional virus production. However, we find that further optimization would lead to a significant increase in the proofreading cost. Unexpected importance of the cost minimization is further supported by other global analyses. We speculate that cost optimization could help avoid cell defense responses. Thus, proofreading is essential for the production of functional viruses, but its rate is limited by energy costs.



The severe acute respiratory syndrome coronavirus (SARS-CoV) belongs to the *Nidovirales* order of viruses, which includes the family of coronaviruses (CoVs).[1] SARS-CoV has an unusually long RNA genome with ∼29.7 kb.[1−4] The SARS-CoV genome encodes 16 nonstructural proteins (nsps1−16) that all assemble into one huge replication−transcription complex.[1,5] This large multiprotein complex is comprised of an RNA-dependent RNA polymerase (nsp12 RdRp) and a collection of other cofactor enzymes that all cooperate with the RdRp during replication and transcription of the SARS-CoV genome.[1]

The SARS-CoV RdRp complex adds an incorrect nucleotide into the nascent RNA chain once in every one million nucleotides, and thus, has a very low error rate on the order of 10^{-6}.[1,4] In contrast, other RNA viruses such as the measles virus are more error-prone with an error rate on the order of 10^{-4}.[6] The strikingly low error rate of the SARS-CoV RdRp complex suggests that the virus most likely removes the wrongly incorporated nucleotides via a proofreading process. Another major clue that the SARS-CoV RdRp complex had proofreading ability was revealed when patients infected with SARS-CoV did not respond to treatments with ribavirin (Rbv), a drug that mimics naturally occurring nucleotides.[7] This lead to the hypothesis that the SARS-CoV RdRp complex could incorporate nucleotide analogues such as Rbv into the nascent RNA chain and then remove them with proofreading.[7−9] Kinetic and crystallographic experiments have established that the SARS-CoV RdRp complex does indeed possess proofreading activity.[10−12] It has been shown specifically through *in vitro* experiments that the nsp12 RdRp (in complex with cofactors nsp7 and nsp8) directly binds to nsp14.[11,13−16] The bifunctional nsp14 enzyme (in complex with its cofactor nsp10) consists of an exonuclease (ExoN) domain that confers 3′-5′ proofreading ability to the SARS-CoV RdRp complex, and an N7-methyltransferase (N7-MTase) domain that adds a 5′ methylguanosine cap to the viral mRNA transcripts produced by the nsp12 RdRp.[7,11−13,17,18]

Despite the extensive body of experimental evidence demonstrating that the SARS-CoV RdRp complex does have the proofreading ability due to the presence of the ExoN domain,[7,10−13,17,18] it is still not clear why the RdRp complex needs to detect and remove wrongly inserted nucleotides. This is especially surprising since the majority of other RNA viruses do not utilize proofreading. In addition, for biological systems with the proofreading ability, it has been shown that all of the so-called characteristic properties cannot be optimized at the same time because there are trade-offs that exist between them.[19−25] For example, it is possible that the RdRp complex operates at a slower speed during genome replication in order to achieve its low error rate on the order of 10^{-6}. Moreover, trade-offs may also arise between the speed or error rate and the energetic cost of proofreading (fraction of extra ATP molecules consumed during futile cycles per step).[23−25]

In this work, we address three important questions regarding the molecular mechanisms and the interplay between the characteristic properties of the SARS-CoV RdRp complex: (i)

Received: January 19, 2021
Accepted: March 3, 2021
Published: March 10, 2021

https://dx.doi.org/10.1021/acs.jpclett.1c00190
J. Phys. Chem. Lett. 2021, 12, 2691−2698

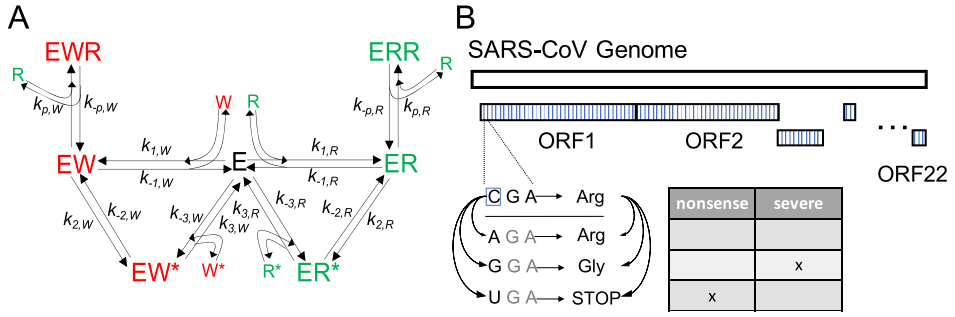


Figure 1. A. Kinetic mechanism for the SARS-CoV RdRp complex with proofreading. The RdRp complex can add either the right *R* (green) or wrong *W* (red) nucleotide to the growing chain during the polymerization step. Nucleotides may be removed during the futile resetting cycles, which consist of Pol-Exo sliding and hydrolysis steps. B. Bioinformatics pipeline for identifying potential nonsense and severe missense mutations. For each of the 22 ORFs of the SARS-CoV genome, we segment it into nonoverlapping codons. We then change each nucleotide base in the codons to one of the other three bases and check which of the substitutions may lead to a critical mutation by determining if there is a change in the amino acid. Nonsense mutations change the original codon to a stop codon. Severe missense mutations change the identity of the resulting amino acid from, e.g., arginine (Arg) to glycine (Gly).

Why does the virus need proofreading to make copies of its unusually large genome? (ii) What would be the functionally tolerable and optimal rate of proofreading? (iii) How does the RdRp complex prioritize among the error rate, speed, and proofreading cost? To answer these questions, we applied our recently developed methodology[21,23,24] to analytically compute the speed, accuracy, and cost and see how variations in the rate constants of underlying chemical transitions affect their trade-offs. To gain additional insights into the interplay between the error rate and accumulation of mutations, the fraction of nonfunctional virus copies from the SARS-CoV genome analysis was determined.[26,27]

To model the kinetic proofreading (KPR) mechanism[28,29] for the SARS-CoV RdRp complex, we used the scheme shown Figure 1A. The KPR mechanism consists of quasi-first-order transitions for the reversible elementary reactions with the rate constants (kinetic parameters) $k_{\pm i,R/W}$ as indicated. The RdRp complex replicates the SARS-CoV genome by adding RNA nucleotides to a nascent strand in the 5′-3′ direction using the complementary strand as a template.[1,5] The KPR mechanism gives the RdRp complex the ability to preferentially discriminate between cognate (right *R*) and noncognate (wrong *W*) nucleotides.[28,29] Following the insertion of one nucleotide, the RdRp complex can either incorporate another nucleotide into the growing chain or remove the newly incorporated nucleotide during proofreading in the 3′-5′ direction via the nsp14 ExoN domain.[7,10,10−12,18] The removal of a nucleotide from the nascent strand via the proofreading pathway consists of two steps, namely, polymerase-exonuclease (Pol-Exo) sliding and hydrolysis.[30] In the first step, the RNA nucleotide slides out of the nsp12 RdRp into the nsp14 ExoN domain. After the RNA nucleotide has entered the nsp14 ExoN domain, it is removed via hydrolysis during the second step. This KPR mechanism enables the RdRp complex to reset back to its original state prior to the addition of the new nucleotide. Thus, the proofreading activity of the RdRp complex leads to the futile hydrolysis of one nucleotide triphosphate (NTP) molecule that is not added to the growing RNA chain.[31]

The kinetic parameters $k_{\pm i,R}$ and the corresponding selectivities (discrimination factors) $f_{\pm i} = k_{\pm i,W}/k_{\pm i,R}$ for each elementary reaction of the KPR mechanism were determined from different kinetic experiments and theoretical studies on the SARS-CoV RdRp complex and on other viral enzymatic systems (see the Supporting Information Table S1). To this end, the polymerization rate constant $k_{1,R} = k_{p,R}$ comes from a recent kinetic experiment on the SARS-CoV RdRp complex.[32] However, due to the lack of kinetic experiments on the SARS-CoV RdRp complex and the rapidly evolving nature of this field, we had to extrapolate the other kinetic parameters from kinetic experiments and/or theoretical studies on the Poliovirus RdRp (3Dpol)[33] and the T7 DNA polymerase.[21,23,34,35] We used the kinetic parameters from these enzymes due to the proposed kinetic and structural similarities between these viral polymerases and the SARS-CoV RdRp complex.[5,10,33] Moreover, we constrained the KPR model by setting the error rate of the SARS-CoV RdRp complex to the experimentally measured one: $\eta_{nat} = 3.4 \times 10^{-6}$.[3,4,27] Then, this experimental error rate was used to determine the catalytic rate $k_{p,W}$ for incorporation of a mismatched nucleotide (i.e., for the $EW \rightarrow EWR$ step). This was done because the error rate is primarily controlled by the discrimination factor $f_p = k_{p,W}/k_{p,R}$ of the product formation (see the Supporting Information for more details).

The characteristic properties of the SARS-CoV RdRp complex, namely, error rate η, kinetic speed *V*, and proofreading cost *C*, are all defined in the so-called linear framework[36] (see the Supporting Information for the mathematical derivations from the forward master equations formalism). The forward master equations for probability distributions of the 5 states in the KPR scheme (Figure 1A) are solved exactly at steady state, and the solution can be used to obtain the stationary fluxes for each transition. Thus, it is straightforward to examine how perturbations of the kinetic parameters affect the interplay among the three characteristic properties. In this framework, the error rate is defined as the ratio of the stationary flux to create the wrong *W* product to the stationary flux to create the right *R* product, and the speed is defined as the stationary flux to create the right *R* product. The proofreading cost is defined as the ratio of the proofreading flux to the product formation flux in the steady state.[21,23,31,37] Thus, in the microscopic picture, the proofreading cost quantifies the fraction of extra NTP molecules hydrolyzed during the futile resetting cycles that do not result in product formation.[31] The proofreading cost is related to the energy dissipation normalized per flux of product formed as discussed in our previous works.[23−25,37] Thus, the proofreading cost can be interpreted as an extra energetic cost.

https://dx.doi.org/10.1021/acs.jpclett.1c00190
J. Phys. Chem. Lett. 2021, 12, 2691−2698

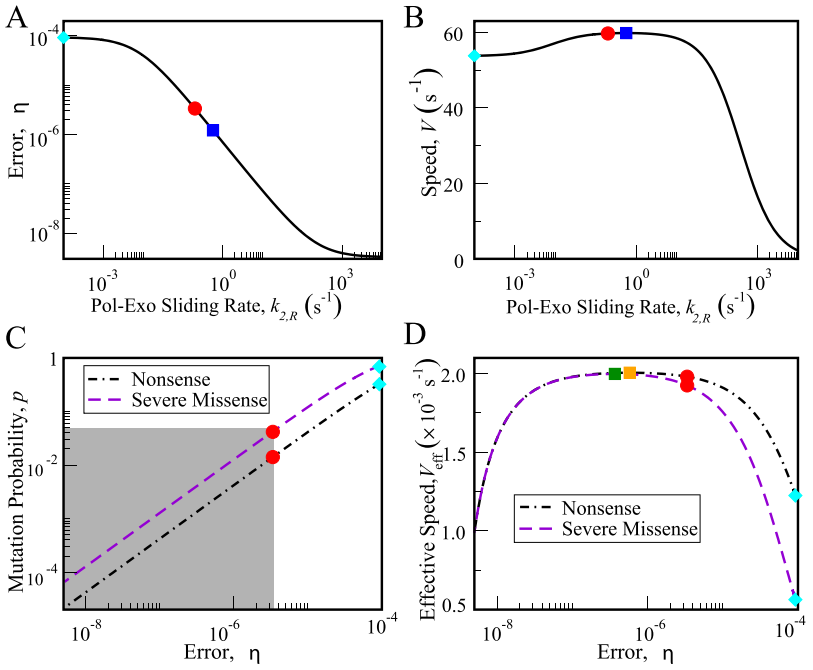


Figure 2. Interplay of the error rate and speed due to proofreading in the SARS-CoV RdRp complex. A. Dependence of the error rate η on the Pol-Exo sliding rate $k_{2,R}$. B. Dependence of the speed V on the Pol-Exo sliding rate $k_{2,R}$. C. Dependence of the mutation probability p on the error rate η for nonsense (dashed-dotted black curve) and severe missense (dashed violet curve) mutations. D. Dependence of the effective speed for the whole SARS-CoV genome V_{eff} on the error rate η for nonsense (dashed-dotted black curve) and severe missense mutations (dashed violet curve). The red dot denotes the native system, the blue square denotes the maximum speed, the orange and green squares denote the maximum effective speed, and the cyan diamonds denote the system at a very low Pol-Exo sliding rate, $k_{2,R} \sim 10^{-4}$ s^{-1}.

We first investigated why the SARS-CoV RdRp complex has the proofreading ability. To this end, we determined how the error rate and the speed change with the Pol-Exo sliding rate $k_{2,R}$. It is clear from Figure 2A that the error rate η monotonically decreases with increasing Pol-Exo sliding rate $k_{2,R}$. Thus, the SARS-CoV RdRp complex operates at a lower error rate and removes more wrongly incorporated nucleotides when the Pol-Exo sliding rate is higher. In contrast to the error rate, the speed V depends nonmonotonically on the Pol-Exo sliding rate $k_{2,R}$ (Figure 2B). For increasing the Pol-Exo sliding rate, the speed first increases and the RdRp complex gets faster until it reaches the maximum, but then the speed starts to decrease due to an increase in the probability of removing correctly incorporated nucleotides from the growing RNA chain. To the left of the maximum, the speed of the RdRp complex decreases as the lower Pol-Exo sliding rate will lead to the complex spending relatively more time in the *EW* state as is evident by the increased return of the probability P_{EW} (see the Supporting Information Figure S6). Notably, for the estimated parameters, the native system (red dot) is within ~0.2% of the maximum speed (blue square in Figure 2B).

Given that the Pol-Exo sliding rate constant does not optimize for error rate and further reduction of error rate is possible at a higher sliding rate (see Figure 2A), we asked whether the native error rate $\eta_{nat} = 3.4 \times 10^{-6}$ is a tolerable one for the virus given its large genome size. To answer this question, the interplay of the error rate η and the mutation probability p, which is defined as the fraction of virus copies that would have reduced fitness due to critical mutations, was examined. We argue that mutations that either insert a stop codon into a viral protein (nonsense) or that change one amino acid to a significantly different one (severe missense) can have such an effect (see the Supporting Information for exact definitions). The mutation probability p for these mutations was calculated by analyzing the effects of nucleotide substitutions in the SARS-CoV genome analysis (Figure 1B). Assuming that the errors in each step are independent, for each value of the error rate we compute the probability $p = 1 - \prod_i(1 - n_i\eta)$ that at least one such mutation will occur during replication of the whole genome. Here, n_i is the number of mutations on the nucleotide and the product is taken over all bases i that may lead to a virus copy with a nonsense or severe missense mutation. As expected, the decrease in the per-step error rate lowers the mutation probability (Figure 2C). We then selected approximate thresholds for the mutation probabilities of the nonsense and severe missense mutations based on the native error rate η_{nat} (red dot). Using this analysis, we determined thresholds of 2% for the probability of nonsense mutations and 5% for the probability of severe missense mutations to postulate a range of error rates that would be tolerable for the virus (gray shaded region in Figure 2C.) One can note that the error rates for the estimated rate of Pol-Exo sliding lie right near the boundary of this range. However, significant reduction or elimination of Pol-Exo sliding and, hence, proofreading would have much higher mutation probabilities as large as ~0.5 (cyan diamonds for $\eta \sim 10^{-4}$). Therefore, proofreading ensures that the SARS-CoV RdRp complex operates at a tolerable error rate that keeps the mutation probability and, thus, the fraction of nonfunctional virus copies small.

To understand the effect of proofreading on the overall replication rate of the SARS-CoV RdRp complex, we also studied the interplay of the error rate with the effective speed $V_{eff} = V(1 - p)/N$, which is defined as the rate of production of functional virus copies (without the above-mentioned deleterious mutations) over the entire SARS-CoV genome

https://dx.doi.org/10.1021/acs.jpclett.1c00190
J. Phys. Chem. Lett. 2021, 12, 2691−2698

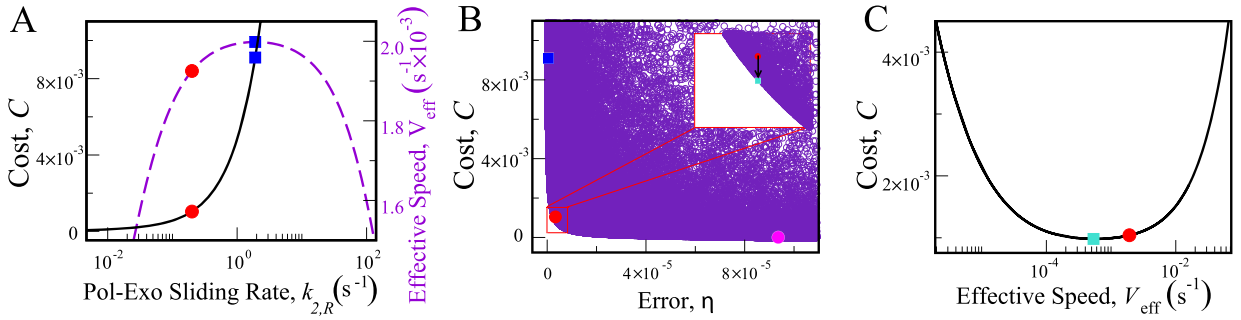


Figure 3. Interplay between the proofreading cost C, effective speed V_{eff}, and error rate η due to variation of different kinetic parameters. A. Dependence of the cost and effective speed on the Pol-Exo sliding rate $k_{2,R}$. B. Scatter plot from global parameter sampling showing the trade-off between the cost and error rate near the cost−error bound. Note the inset in the region of the native system (red dot). C. Dependence of the cost on the effective speed from variation of the polymerization rate $k_{1,R} = k_{p,R}$ and the Pol-Exo sliding rate $k_{2,R}$ at a fixed error rate $\eta_{\text{nat}} = 3.4 \times 10^{-6}$. The blue square denotes the maximum effective speed, the magenta dot denotes the error rate η_0 at $C = 0$ without proofreading, and the turquoise square denotes the theoretical minimum cost for the native error rate on the cost−error bound.

with $N = 29725$ bases. Depending on which mutations are postulated to be nonfunctional, different mutation probabilities p lead to different values of V_{eff}. It is expected that this quantity should have a nonmonotonic dependence on the error rate as an increase of speed due to a decrease in Pol-Exo sliding rate competes with the reduction of the fraction of nonmutated viruses produced. The result shown in Figure 2D demonstrates that there is indeed a nonmonotonic dependence of the effective speed on the error rate. The native experimentally observed value of the error rate at $\eta_{\text{nat}} = 3.4 \times 10^{-6}$ nearly maximizes the effective speed (within ∼1% of the maximum shown by the orange square for nonsense mutations and within ∼4% of the maximum shown by the green square for severe missense mutations). Elimination of the proofreading or a significant increase in the proofreading would lead to a decrease in the effective speed. Therefore, due to the large size of the SARS-CoV genome, the error rate must be small enough to keep the effective speed of the RdRp complex close to its optimal value.

Interestingly, the SARS-CoV RdRp complex does not completely optimize its effective speed, and, in fact, the native error rate lies on the non-trade-off branch of the plot in Figure 2D where an increase in the Pol-Exo sliding rate will significantly decrease the error and slightly improve the effective speed. While we could not exclude the possibility that this effect is due to uncertainty in the estimation of the kinetic parameters, we nevertheless asked if other characteristic properties of the system can explain why the SARS-CoV RdRp complex does not completely maximize its effective speed. To this end, the interplay between the effective speed V_{eff} and the proofreading cost C was analyzed. The value of the Pol-Exo sliding rate $k_{2,R}$ was varied, and the corresponding properties are presented as a parametric plot in Figure 3A. The results indicate that a further increase in effective speed V_{eff} and the corresponding decrease in the error rate would significantly increase the cost. For instance, for the severe missense mutations (blue square), the cost would increase almost by ∼1 order of magnitude relative to the cost of the native system (red dot). This nearly 1 order of magnitude increase in the cost is much larger than the corresponding increase in the effective speed of only ∼4%. Thus, it is possible that the SARS-CoV RdRp complex cannot increase its effective speed to the maximum V_{eff} because the proofreading cost C would increase dramatically.

To test the robustness of our conclusions with respect to the estimated uncertainties in the kinetic parameters of the SARS-CoV RdRp complex, we performed a sensitivity analysis. To this end, we randomly perturbed the values of the kinetic parameters $k_{\pm i,R}$ from their estimated values stated in the Table S1. Perturbed values of the right nucleotide parameters were sampled from uniform distributions with the respective ranges of $[0.5k_{\pm i,R}, 2k_{\pm i,R}]$ and $[3k_{\pm i,R}, 6k_{\pm i,R}]$ (200 from each) while keeping all of the discrimination factors f_i fixed. The first range was chosen because we would like to examine the robustness of our conclusions for smaller variations (i.e., perturbations) of the kinetic parameters. We chose the second range to include the range of the polymerization rate constants at the physiological temperature (i.e., 600−700 s^{-1}) proposed by Shannon et al.[32] For each parameter set, we analyzed the effect of the perturbations in the kinetic parameters on the error rate η, speed V, and proofreading cost C of the SARS-CoV RdRp complex. First, we tested whether or not the native error rate would still be far from the minimum error rate but much smaller than the error rate with no proofreading. Second, we tested if the speed would still be nearly optimized. Last, we tested if the proofreading cost at the optimal speed would still be much larger than the native cost. The results of the analysis (see the Supporting Information, Figures S4 and S5, for more details) show that our conclusions about the error rate, speed, and perhaps to a lesser extent the cost are robust in both ranges.

The results discussed so far investigated the interplay among the error rate (mutation probability), speed, and proofreading cost of the SARS-CoV RdRp complex as we varied the rate of Pol-Exo sliding. This analysis shows that the value of this rate constant estimated from the experimentally determined error rate nearly optimizes the speed while keeping the error rate tolerable. Somewhat surprisingly, in contrast to other studied enzymatic systems,[21,23,24,37] our analysis points to the importance of minimizing the proofreading cost rather than maximization of the speed. To understand how these results would hold when other kinetic parameters are varied, we next performed a more global analysis of the trade-offs.

First, following the approach introduced in Yu et al.,[24] we asked if the observed proofreading cost is at or near its global minimum value for a given error rate. To this end, we generated the scatter plot in Figure 3B from variation of all the rate constants with global parameter sampling under thermodynamic constraints, i.e., fixed discrimination factors

https://dx.doi.org/10.1021/acs.jpclett.1c00190
J. Phys. Chem. Lett. 2021, 12, 2691−2698

(selectivities) $f_i = k_{i,W}/k_{i,R}$ for each elementary reaction and fixed chemical potential differences $\Delta\mu_{cycle}$ around each biochemical cycle (see the Supporting Information for details). The interplay between the proofreading cost C and the error rate η shows a trade-off between the two properties with the cost−error bound that gives the theoretical minimum cost for a given error rate. Along this boundary, a decrease in the error rate leads to an unavoidable increase in the proofreading cost. On one hand, without the proofreading the cost might vanish (magenta dot for $C = 0$ from a corresponding Michaelis−Menten system with no futile cycles where $k_{2,R} = k_{3,R} = 0$), but this leads to an ∼30 fold increase in the error. Thus, these results further confirm that the dissipation is unavoidable for the SARS-CoV RdRp complex to achieve the tolerable error rate. On the other hand, further minimizing the error would lead to a significant increase in the cost of ∼10-fold (blue square) as we saw with the single parameter variation in Figure 3A. Consequently, due to the global trade-off between the cost and the error rate, the SARS-CoV RdRp complex must keep its proofreading cost small enough to maintain a tolerable error rate. Importantly, as one can see from the inset of Figure 3B, the native system lies very close to the trade-off bound (within ∼6%). Thus, the RdRp complex nearly minimizes the proofreading cost for the observed error rate, and our analysis further confirms the importance of considering the energy dissipation for understanding the functionality of the SARS-CoV RdRp complex.

To understand the underlying reason why the cost of the native system does not lie exactly on the cost−error bound, the interplay between the cost C and the effective speed V_{eff} for a fixed error rate was analyzed. To determine if changes in other kinetic parameters can lead to further cost minimization for the same error, we varied the polymerization rate $k_{1,R} = k_{p,R}$ and the Pol-Exo sliding rate $k_{2,R}$ so that the error rate was fixed at its native value and computed the proofreading cost C and effective speed V_{eff}. The result in Figure 3C shows the nonmonotonic curve for the cost versus the effective speed. The native SARS-CoV RdRp complex (red dot) sits on a non-trade-off branch of the cost−speed curve; i.e., a decrease in the costs will result in a decrease in the speed. The SARS-CoV RdRp complex could further decrease the cost (red dot) by almost ∼6% to reach the minimum cost (turquoise square). However, the RdRp complex is prevented from completely optimizing the cost because the effective speed decreases by ∼4-fold. Therefore, the SARS-CoV RdRp complex adopts an economic strategy to keep the proofreading cost small (i.e., close to the theoretical minimum cost) while maintaining a high speed to replicate its genome.

The results of our study on the interplay of the characteristic properties from single and global parameter variation appear to indicate that the proofreading cost is highly prioritized by the SARS-CoV RdRp complex. To gain further insight into the importance of the proofreading cost relative to the speed and the error rate as criteria for optimization, we followed the protocol developed in ref 23 and analyzed how far each characteristic property is from its optimal value that respectively minimizes the cost C, maximizes the speed V, and minimizes the error rate η. To this end, each kinetic parameter was varied over 6 orders of magnitude around the "native" value of k_0 ($[10^{-3}k_0, 10^3k_0]$), and the following function was considered,

$$d_{k,i} \equiv -\log_{10}\left(\frac{|f(k_0) - f(k_{max/min})|}{f(k_{max}) - f(k_{min})}\right); \quad f(k_i) = \eta, V, C \tag{1}$$

Here, k_i denotes the rate constant being varied with index i, and $k_{max/min}$ is the rate constant at which the cost, speed, or error rate $f(k_{max/min})$ has an optimal value (maximum or minimum). The larger this value of $d_{k,i}$, the closer the native value is to the optimal value (see the Supporting Information for more details). The results shown in Figure S12 indicate that at least three of the kinetic parameters $k_{1,R} = k_{p,R}$, $k_{-1,R}$, and $k_{2,R}$ are optimized for the proofreading cost. The other two parameters, $k_{-2,R}$ and $k_{3,R}$, are optimized for the maximal speed with the error close in second. Given the difference in trends for different steps, we opted against averaging these quantities as we previously did in ref 23. For comparison, the analysis was repeated also for T7 DNA polymerase that uses the same kinetic scheme. The results are qualitatively similar with one exception, $k_{2,R}$, where the speed is optimized instead of the proofreading cost.

Taken all together, our results show that the proofreading cost is one of the most important features of the SARS-CoV RdRp complex, perhaps along with the speed. In comparison with other biological systems (e.g., T7 DNA polymerase, *E. coli* ribosome, and isoleucine tRNA synthetase),[21,23,24,35] the relative importance of cost versus speed may be even larger for the SARS-CoV RdRp complex. This finding raises a question on the physiological meaning of the cost optimization from the evolutionary perspective. To better understand the prioritization of the proofreading cost in the SARS-CoV RdRp complex, we propose the following biochemical arguments to explain it. By keeping the proofreading cost small, the SARS-CoV RdRp complex limits the cytoplasmic accumulation of monophosphate nucleotides that are produced as a result of the proofreading steps during the futile cycles. It is well established that accumulation of these molecules, e.g., AMP molecules created as byproducts of ATP hydrolysis, can trigger innate immune defense mechanisms or cell autophagy.[38] In particular, activation of AMP-activated protein kinase (AMPK) in response to an increased cellular AMP/ATP ratio has been linked to cellular defenses against pathogens.[39−41] Notably, the ability of SARS-CoV-2 to interfere with AMPK and thereby reduce autophagy has been reported.[42] Thus, keeping the proofreading cost minimal can be critical for the success of the SARS-CoV RdRp complex to replicate the viral genome in the infected host cell without alerting the cell defense mechanisms. At the same time, it is possible that other molecular mechanisms might force the SARS-CoV RdRp complex to optimize the proofreading cost. It will be important to test this idea in more detail using both experimental and theoretical approaches.

In summary, our analysis of the interplay among error rate, speed, and proofreading cost for the SARS-CoV RdRp complex has revealed some important and unexpected results on the functionality of this complicated multiprotein structure. We have shown that the SARS-CoV RdRp complex operates at a tolerable error rate that keeps the fraction of nonfunctional virus copies low. Furthermore, the tolerable error rate allows the SARS-CoV RdRp complex to nearly optimize its effective speed when certain types of mutations (i.e., nonsense and severe missense) are considered. Consequently, proofreading is necessary for the SARS-CoV RdRp complex to keep the rate of production of functional virus copies nearly maximal, but the

https://dx.doi.org/10.1021/acs.jpclett.1c00190
J. Phys. Chem. Lett. 2021, 12, 2691−2698

rate would decrease if the Pol-Exo sliding rate is too high. Furthermore, the SARS-CoV RdRp complex cannot reach the maximum effective speed for the production of functional virus copies because the proofreading cost would increase by about 1 order of magnitude. Moreover, the error rate for the SARS-CoV RdRp complex cannot be further decreased for the same reason. Given that the proofreading cost prevents the SARS-CoV RdRp complex from further reducing its error rate and completely optimizing its effective speed, it would be easy to assume that this quantity is completely minimized. However, the SARS-CoV RdRp complex is prevented from completely minimizing the proofreading cost to the theoretical minimum cost due to a slow down in the effective speed. Finally, our quantitative study on ranking the proofreading cost, speed, and error rate of the SARS-CoV RdRp complex places the proofreading cost as one of the most important characteristic properties along with the speed.

Our results in this work have uncovered some intriguing features on the trade-offs and optimization of the characteristic properties of the SARS-CoV RdRp complex. Specifically, we now understand that the proofreading ability conferred by the nsp14 ExoN domain[43] is necessary for SARS-CoV to replicate its unusually large genome at nearly optimal speed while, at the same time, keeping the fraction of nonfunctional virus copies low. This finding directly contrasts with other RNA viruses that do not have the proofreading ability but instead rely heavily on high mutation rates and diversity of their small genomes to continue infecting their hosts.[1,9] Furthermore, the proofreading cost itself is nearly optimized for the RdRp complex, and the importance of the cost may enable SARS-CoV to avoid triggering defense pathways in the host cell.[38] However, our current knowledge on the functionality of SARS-CoV and its associated proofreading mechanism is still far from complete. We note that another advantage of proofreading in the SARS-CoV RdRp complex is that the error-correction ability renders the virus resistant to medical treatments with certain types of drugs (i.e., nucleotide analogues that mimic naturally occurring nucleotides) such as Ribavirin.[7,9] Therefore, it is important to encourage the development and testing of viable drugs against SARS-CoV such as nucleotide analogues like Remdesivir and others that can strategically target both the nsp12 RdRp and the nsp14 ExoN domains of the SARS-CoV RdRp complex.[9,44−47]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.1c00190.

Derivation and description of the SARS-CoV RdRp complex KPR model and the underlying mathematical formalism; definitions of the error rate η, speed V, and proofreading cost C; qualitative and quantitative sensitivity analysis on the robustness of our conclusions for the error rate η, speed V, and proofreading cost C due to variation of all the kinetic parameters $k_{i,R}$; interplay of the stationary probabilities for each biochemical state in the KPR scheme with the Pol-Exo sliding rate $k_{2,R}$; complete description and mathematical details of our bioinformatics analysis; effective speed V_{eff} as a function of the error rate η using the complex formalism and the simple power law for nonsense mutations and severe missense mutations; speed V and effective speed V_{eff} for two different definitions of the speed; interplay of the effective speed V_{eff} with the proofreading cost C; histograms of the effective speed V_{eff} distributions at a fixed error rate η_{nat} from the variation of both the polymerization $k_{1,R} = k_{p,R}$ and the Pol-Exo sliding $k_{2,R}$ rate constants; and mathematical and graphical description of the quantitative methods for ranking the importance of the characteristic properties for each kinetic parameter $k_{i,R}$ individually (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Oleg A. Igoshin − *Department of Bioengineering, Center for Theoretical Biological Physics, Department of Chemistry, and Department of Biosciences, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-1449-4772; Email: igoshin@rice.edu

Anatoly B. Kolomeisky − *Department of Chemistry, Department of Chemical and Biomolecular Engineering, Department of Physics and Astronomy, and Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0001-5677-6690; Email: tolya@rice.edu

Authors

Joel D. Mallory − *Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005, United States;* orcid.org/0000-0002-0251-5724

Xian F. Mallory − *Department of Computer Science, Florida State University, Tallahassee, Florida 32306, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.1c00190

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by Center for Theoretical Biological Physics National Science Foundation (NSF) Grant PHY-2019745 as well as Welch Foundation Grants C-1995 (to O.A.I.) and C-1559 (to A.B.K.). A.B.K. also acknowledges NSF Grants CHE-1953453 and MCB-1941106.

■ REFERENCES

(1) Denison, M. R.; Graham, R. L.; Donaldson, E. F.; Eckerle, L. D.; Baric, R. S. Coronaviruses: An RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.* **2011**, *8*, 270−279.

(2) Stadler, K.; Masignani, V.; Eickmann, M.; Becker, S.; Abrignani, S.; Klenk, H.-D.; Rappuoli, R. SARS—beginning to understand a new virus. *Nat. Rev. Microbiol.* **2003**, *1*, 209−218.

(3) Zhu, Q.; Yu, M.; Fan, B.; Chang, G.; Si, B.; Peng, W.; Jiang, T.; Liu, B.; Deng, Y.; Liu, H.; et al. A Complete Sequence and Comparative Analysis of a SARS-Associated Virus (Isolate BJ01). *Chin. Sci. Bull.* **2003**, *48*, 941−948.

(4) Yeh, S.-H.; Wang, H.-Y.; Tsai, C.-Y.; Kao, C.-L.; Yang, J.-Y.; Liu, H.-W.; Su, I.-J.; Tsai, S.-F.; Chen, D.-S.; Chen, P.-J.; et al. Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: Molecular epidemiology and genome evolution. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2542−2547.

(5) Smith, E. C.; Denison, M. R. Coronaviruses as DNA wannabes: A new model for the regulation of RNA virus replication fidelity. *PLoS Pathog.* **2013**, *9*, No. e1003760.

(6) Drake, J. W.; Holland, J. J. Mutation rates among RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 13910−13913.

(7) 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. Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E162−E171.
(8) Pruijssers, A. J.; Denison, M. R. Nucleoside analogues for the treatment of coronavirus infections. *Curr. Opin. Virol.* **2019**, *35*, 57−62.
(9) Robson, F.; Khan, K. S.; Le, T. K.; Paris, C.; Demirbag, S.; Barfuss, P.; Rocchi, P.; Ng, W.-L. Coronavirus RNA proofreading: Molecular basis and therapeutic targeting. *Mol. Cell* **2020**, *79*, 710−727.
(10) Bouvet, M.; Imbert, I.; Subissi, L.; Gluais, L.; Canard, B.; Decroly, E. RNA 3′-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 9372−9377.
(11) Subissi, L.; Posthuma, C. C.; Collet, A.; Zevenhoven-Dobbe, J. C.; Gorbalenya, A. E.; Decroly, E.; Snijder, E. J.; Canard, B.; Imbert, I. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E3900−E3909.
(12) Ma, Y.; Wu, L.; Shaw, N.; Gao, Y.; Wang, J.; Sun, Y.; Lou, Z.; Yan, L.; Zhang, R.; Rao, Z. Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 9436−9441.
(13) Minskaia, E.; Hertzig, T.; Gorbalenya, A. E.; Campanacci, V.; Cambillau, C.; Canard, B.; Ziebuhr, J. Discovery of an RNA virus 3 - 5 exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 5108−5113.
(14) Te Velthuis, A. J.; Arnold, J. J.; Cameron, C. E.; van den Worm, S. H.; Snijder, E. J. The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res.* **2010**, *38*, 203−214.
(15) Kirchdoerfer, R. N.; Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **2019**, *10*, 2342.
(16) Chen, J.; Malone, B.; Llewellyn, E.; Grasso, M.; Shelton, P. M.; Olinares, P. D. B.; Maruthi, K.; Eng, E. T.; Vatandaslar, H.; Chait, B. T. Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* **2020**, *182*, 1560−1573.e13.
(17) Sevajol, M.; Subissi, L.; Decroly, E.; Canard, B.; Imbert, I. Insights into RNA synthesis, capping, and proofreading mechanisms of SARS-coronavirus. *Virus Res.* **2014**, *194*, 90−99.
(18) Becares, M.; Pascual-Iglesias, A.; Nogales, A.; Sola, I.; Enjuanes, L.; Zuñiga, S. Mutagenesis of coronavirus nsp14 reveals its potential role in modulation of the innate immune response. *J. Virol.* **2016**, *90*, 5399−5414.
(19) Bennett, C. H. Dissipation-error tradeoff in proofreading. *BioSystems* **1979**, *11*, 85−91.
(20) Johansson, M.; Zhang, J.; Ehrenberg, M. Genetic code translation displays a linear trade-off between efficiency and accuracy of tRNA selection. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 131−136.
(21) Banerjee, K.; Kolomeisky, A. B.; Igoshin, O. A. Elucidating interplay of speed and accuracy in biological error correction. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 5183−5188.
(22) Wong, F.; Amir, A.; Gunawardena, J. Energy-speed-accuracy relation in complex networks for biological discrimination. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2018**, *98*, 012420.
(23) Mallory, J. D.; Kolomeisky, A. B.; Igoshin, O. A. Trade-Offs between Error, Speed, Noise, and Energy Dissipation in Biological Processes with Proofreading. *J. Phys. Chem. B* **2019**, *123*, 4718−4725.
(24) Yu, Q.; Mallory, J. D.; Kolomeisky, A. B.; Ling, J.; Igoshin, O. A. Trade-Offs between Speed, Accuracy, and Dissipation in tRNAIle Aminoacylation. *J. Phys. Chem. Lett.* **2020**, *11*, 4001−4007.
(25) Mallory, J. D.; Igoshin, O. A.; Kolomeisky, A. B. Do We Understand the Mechanisms Used by Biological Systems to Correct Their Errors? *J. Phys. Chem. B* **2020**, *124*, 9289−9296.
(26) Henikoff, S.; Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 10915−10919.
(27) Xu, J.; Hu, J.; Wang, J.; Han, Y.; Hu, Y.; Wen, J.; Li, Y.; Ji, J.; Ye, J.; Zhang, Z.; et al. Genome Organization of the SARS-CoV. *Genomics, Proteomics Bioinf.* **2003**, *1*, 226−235.
(28) Hopfield, J. J. Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U. S. A.* **1974**, *71*, 4135−4139.
(29) Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **1975**, *57*, 587−595.
(30) Kunkel, T. A. DNA replication fidelity. *J. Biol. Chem.* **2004**, *279*, 16895−16898.
(31) Savageau, M. A.; Lapointe, D. S. Optimization of kinetic proofreading: a general method for derivation of the constraint relations and an exploration of a specific case. *J. Theor. Biol.* **1981**, *93*, 157−177.
(32) Shannon, A.; Selisko, B.; Huchting, J.; Touret, F.; Piorkowski, G.; Fattorini, V.; Ferron, F.; Decroly, E.; Meier, C.; Coutard, B.; et al. Rapid incorporation of Favipiravir by the fast and permissive viral RNA polymerase complex results in SARS-CoV-2 lethal mutagenesis. *Nat. Commun.* **2020**, *11*, 4682.
(33) Arnold, J. J.; Cameron, C. E. Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg^{2+}. *Biochemistry* **2004**, *43*, 5126−5137.
(34) Wong, I.; Patel, S. S.; Johnson, K. A. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry* **1991**, *30*, 526−537.
(35) Banerjee, K.; Kolomeisky, A. B.; Igoshin, O. A. Accuracy of Substrate Selection by Enzymes is Controlled by Kinetic Discrimination. *J. Phys. Chem. Lett.* **2017**, *8*, 1552−1556.
(36) Gunawardena, J. A linear framework for time-scale separation in nonlinear biochemical systems. *PLoS One* **2012**, *7*, No. e36321.
(37) Mallory, J. D.; Kolomeisky, A. B.; Igoshin, O. A. Kinetic control of stationary flux ratios for a wide range of biochemical processes. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 8884−8889.
(38) Silwal, P.; Kim, J. K.; Yuk, J.-M.; Jo, E.-K. AMP-activated protein kinase and host defense against infection. *Int. J. Mol. Sci.* **2018**, *19*, 3495.
(39) Mankouri, J.; Harris, M. Viruses and the fuel sensor: the emerging link between AMPK and virus replication. *Rev. Med. Virol.* **2011**, *21*, 205−212.
(40) Xie, N.; Yuan, K.; Zhou, L.; Wang, K.; Chen, H.-N.; Lei, Y.; Lan, J.; Pu, Q.; Gao, W.; Zhang, L.; et al. PRKAA/AMPK restricts HBV replication through promotion of autophagic degradation. *Autophagy* **2016**, *12*, 1507−1520.
(41) Prantner, D.; Perkins, D. J.; Vogel, S. N. AMP-activated kinase (AMPK) promotes innate immunity and antiviral defense through modulation of stimulator of interferon genes (STING) signaling. *J. Biol. Chem.* **2017**, *292*, 292−304.
(42) Gassen, N. C.; Papies, J.; Bajaj, T.; Dethloff, F.; Emanuel, J.; Weckmann, K.; Heinz, D. E.; Heinemann, N.; Lennarz, M.; Richter, A. et al. Analysis of SARS-CoV-2-controlled autophagy reveals spermidine, MK-2206, and niclosamide as putative antiviral therapeutics. *BioRxiv*, 2020.
(43) Ogando, N.; Ferron, F.; Decroly, E.; Canard, B.; Posthuma, C.; Snijder, E. The curious case of the nidovirus exoribonuclease: Its role in RNA synthesis and replication fidelity. *Front. Microbiol.* **2019**, *10*, 1813.
(44) Barnard, D. L.; Hubbard, V. D.; Burton, J.; Smee, D. F.; Morrey, J. D.; Otto, M. J.; Sidwell, R. W. Inhibition of severe acute respiratory syndrome-associated coronavirus (SARSCoV) by calpain inhibitors and β-D-N4-hydroxycytidine. *Antivir. Chem. Chemother.* **2004**, *15*, 15−22.
(45) Furuta, Y.; Komeno, T.; Nakamura, T. Favipiravir (T-705), a broad spectrum inhibitor of viral RNA polymerase. *Proc. Jpn. Acad., Ser. B* **2017**, *93*, 449−463.

https://dx.doi.org/10.1021/acs.jpclett.1c00190
J. Phys. Chem. Lett. 2021, 12, 2691−2698

(46) Gordon, C. J.; Tchesnokov, E. P.; Woolner, E.; Perry, J. K.; Feng, J. Y.; Porter, D. P.; Götte, M. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J. Biol. Chem.* **2020**, *295*, 6785−6797.
(47) de Wit, E.; Feldmann, F.; Cronin, J.; Jordan, R.; Okumura, A.; Thomas, T.; Scott, D.; Cihlar, T.; Feldmann, H. Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 6771−6776.

https://dx.doi.org/10.1021/acs.jpclett.1c00190
J. Phys. Chem. Lett. 2021, 12, 2691−2698