



A computational approach for predicting off-target toxicity of antiviral ribonucleoside analogues to mitochondrial RNA polymerase

Received for publication, February 23, 2018, and in revised form, May 4, 2018. Published, Papers in Press, May 8, 2018, DOI 10.1074/jbc.RA118.002588

Holly Freedman^{‡1}, Philip Winter[§], Jack Tuszynski^{§¶}, D. Lorne Tyrrell[‡], and Michael Houghton^{‡2}

From the [‡]Li Ka Shing Applied Virology Institute, Department of Medical Microbiology and Immunology and the [¶]Department of Physics, University of Alberta, Edmonton, Alberta T6G 2E1, Canada and the [§]Department of Oncology, University of Alberta, Edmonton, Alberta T6G 2R7, Canada

Edited by Karin Musier-Forsyth

In the development of antiviral drugs that target viral RNA-dependent RNA polymerases, off-target toxicity caused by the inhibition of the human mitochondrial RNA polymerase (POLRMT) is a major liability. Therefore, it is essential that all new ribonucleoside analogue drugs be accurately screened for POLRMT inhibition. A computational tool that can accurately predict NTP binding to POLRMT could assist in evaluating any potential toxicity and in designing possible salvaging strategies. Using the available crystal structure of POLRMT bound to an RNA transcript, here we created a model of POLRMT with an NTP molecule bound in the active site. Furthermore, we implemented a computational screening procedure that determines the relative binding free energy of an NTP analogue to POLRMT by free energy perturbation (FEP), *i.e.* a simulation in which the natural NTP molecule is slowly transformed into the analogue and back. In each direction, the transformation was performed over 40 ns of simulation on our IBM Blue Gene Q supercomputer. This procedure was validated across a panel of drugs for which experimental dissociation constants were available, showing that NTP relative binding free energies could be predicted to within 0.97 kcal/mol of the experimental values on average. These results demonstrate for the first time that free-energy simulation can be a useful tool for predicting binding affinities of NTP analogues to a polymerase. We expect that our model, together with similar models of viral polymerases, will be very useful in the screening and future design of NTP inhibitors of viral polymerases that have no mitochondrial toxicity.

The mitochondria contain their own circular dsDNA genome that encodes 13 proteins involved in the electron transfer chain, in addition to 2 rRNAs and 22 tRNAs (1). This genome is replicated by mitochondrial DNA polymerase γ and transcribed by human mitochondrial DNA-directed RNA po-

lymerase, known as POLRMT.³ These polymerases are themselves encoded by nuclear DNA.

The inhibition of viral polymerases by nucleoside analogues has been the cornerstone of therapies against multiple pathogenic viruses, including HIV, hepatitis B virus, hepatitis C virus (HCV), and *Herpesvirus*. In particular, many viruses possessing RNA genomes, such as HCV, use RNA-dependent RNA polymerases for replication, which have been very successful targets for the development of antiviral drugs including ribonucleoside analogues. Ribonucleoside analogue drugs are converted to their active ribonucleoside triphosphate form by human kinases. After more than a decade of research aimed at identifying an anti-HCV ribonucleoside analogue, sofosbuvir was approved for medical use in the United States in 2013. Along the way, multiple attempts to develop ribonucleoside antivirals had to be discontinued because of toxicity, including NM283, R1626, PSI-938, and BMS-986094 (2–5). It had been known for years that deoxyribonucleoside analogues were associated with side effects caused by mitochondrial toxicity and off-target interactions with mitochondrial DNA polymerase in particular (6). However, it was not until 2012 that research performed by Arnold *et al.* (3) established that POLRMT is a major source of off-target toxicity for nucleoside analogue antivirals aimed at inhibiting microbial RNA polymerases. This work emphasized the need for setting up accurate screening tests for the interactions of ribonucleoside antivirals with POLRMT. One serious concern cited was that although nuclear RNA polymerase II has a proofreading mechanism through the activity of TFIIS, there does not seem to be any nucleotide excision mechanism in place for POLRMT (3). Pyrimidine nucleotide analogues pose an even greater risk for POLRMT inhibition than ATP derivatives, because CTP and UTP are present at 1–2 orders of magnitude lower concentrations in the mitochondria than ATP (1).

On the computational side, one previous predictive model of the mitochondrial toxicity of pharmaceutical drugs has focused on a cheminformatics approach (7). Here we describe the development of a computational model based upon the available crystal structure (8) that allows for the accurate prediction of relative binding affinities of NTPs to the POLRMT structure.

This work was supported by computing resources provided by the Southern Ontario Smart Computing Innovation Platform, Scinet, Compute Canada, and WestGrid. This work was also supported by funds from Alberta Innovates to the Li Ka Shing Applied Virology Institute and by a Canadian Excellence Research Chair grant (to M. H.). The authors declare that they have no conflicts of interest with the contents of this article.

¹ To whom correspondence may be addressed. E-mail: hfreedma@ualberta.ca.

² To whom correspondence may be addressed. E-mail: mhoughto@ualberta.ca.

³ The abbreviations used are: POLRMT, human mitochondrial RNA polymerase; FEP, free energy perturbation; NTE, N-terminal extension; MOE, molecular operating environment; CpHMD, constant pH molecular dynamics; RMSD, root-mean-square deviation; HCV, hepatitis C virus.

The detailed molecular structure of POLRMT as a closed-product elongation complex bound to template DNA and a small RNA transcript has been provided by X-ray crystallography, allowing for insights into NTP binding within the active site (Protein Data Bank code 4BOC) (8). The closed product is one of four stages of the nucleotide addition cycle of polymerases. In this cycle, an NTP first binds to the open post-translocated state of the polymerase, the complex undergoes a conformational change to the new pre-translocated insertion state, chemical reaction occurs leading to the closed product state, and finally translocation returns the system to the open product complex (9). To assess POLRMT inhibition, we had to create a model of the pre-translocated insertion state. We therefore adjusted the 4BOC structure to make it representative of the insertion state by using the crystal structure of the insertion state of the bacteriophage T7 RNA polymerase to which POLRMT is evolutionarily related (10, 11), to model the nucleoside triphosphate CTP and that of T7 DNA polymerase to model the active site metal ions (12). We also constructed models with each of ATP and GTP in the POLRMT nucleotide-binding site.

In our model development and validation, we take advantage of the experimentally determined dissociation constants for a panel of NTP analogues that were determined using a biochemical assay by Arnold *et al.* (3). The binding free-energy differences of these NTP analogues are determined relative to the binding free energies of the respective natural NTPs using the FEP methodology. FEP has the advantage over other less rigorous methods of being formally exact in the limit of complete phase-space sampling but requires lengthy simulations (13–17). Although using FEP to determine binding free energies is quite well-studied (16, 18), to the best of our knowledge this study is the first one to report the use of FEP to calculate binding free energies of NTP analogues to a polymerase.

Binding free energies of NTP analogues to polymerases are particularly hard to compute as averages over molecular dynamics simulations because of the high negative charge of $-4e$ on nucleoside triphosphates; that is, the response in the fluctuation of the environment to fluctuations in the positions of the charged phosphates and metal ions may be expected to occur on a slower time scale than for systems with less charge polarization, leading to slow convergence of computed binding free energies. To converge results, we have performed lengthy simulations on our IBM Blue Gene Q supercomputer (see below for details). We note that in FEP, only relative binding free energies are calculated, *i.e.* binding free-energy differences between different NTPs, instead of absolute binding free energies. This facilitates convergence because fluctuations in energy caused by fluctuations in charge density cancel out when the energetic difference is calculated for two NTPs. Relative binding free energies are determined using a thermodynamic cycle, *i.e.* from the difference between the free energy change for the transformation of the natural NTP to its analogue in aqueous solution and the corresponding free energy change within the solvated protein-bound environment (19).

The FEP approach presented here can be applied to the rational salvaging of drug candidates found to inhibit POLRMT and moreover can be extended to applications to viral polymerases

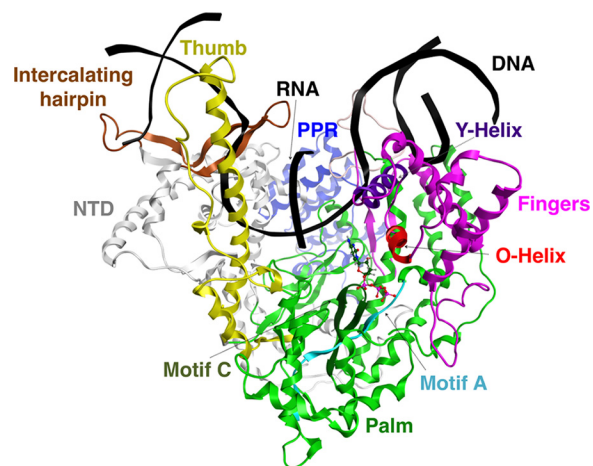


Figure 1. Molecular model of the nucleotide insertion state of POLRMT. Shown is a ribbon representation of our computational model of POLRMT in complex with DNA template and nontemplate strands, an RNA transcript, and an incoming CTP, with important domains and subdomains, which are discussed in the text, labeled. The upstream end of the template DNA is on the left-hand side of the figure, and the downstream end is on the right. Part of the nontemplate DNA forming the bubble region is missing from our model. PPR, pentatricopeptide repeat; NTD, N-terminal domain.

in general. When combined with experimental testing, it suggests an efficient strategic means for identifying nucleoside antivirals with high specificity and affinity for viral polymerases, leading to the design of safe and effective drugs.

Results

Molecular dynamics simulations of the NTP insertion state of POLRMT

We have performed molecular dynamics simulations of our model of the POLRMT NTP insertion state, which is based on the 4BOC (8) crystal structure of the POLRMT elongation complex. POLRMT is a 1230-amino acid protein consisting of three domains, namely the N-terminal extension (NTE) located at the very N terminus of the protein (residues 1–368), the N-terminal domain (residues 369–647) that follows the NTE in the primary sequence, and the C-terminal polymerase domain (residues 648–1230) (1). The polymerase domain has the classic structure resembling a cupped right hand as is seen in the RNA-dependent polymerases of viruses and bacteriophages and in DNA polymerases (Fig. 1). It has 41% sequence similarity and 25% sequence identity to T7 RNA polymerase (20). The pentatricopeptide repeat, making up the C terminus of the NTE, the N-terminal domain, and the C-terminal domain all are present in the 4BOC crystal structure and in our model.

The NTP-binding pocket is located on top of the center of the palm and is formed mainly by structural motifs A and C, and the helical motif B, also known as the O-helix, belonging to the fingers region (Fig. 2A). Conserved residues within the active site contribute to the formation of the catalytic complex. Several residues are involved in coordinating the two active site Mg^{2+} metal ions, A and B (Fig. 2A). In our model (Fig. 2A), the side chains of two conserved Asp residues, Asp⁹²² and Asp¹¹⁵¹, respectively, belonging to motifs A and C help to coordinate metal A, along with the 3'OH O atom of the NTP, an α -phosphoryl O atom of the NTP, the main-chain carbonyl O atom of Gly⁹²³, and a water molecule. Metal B is coordinated by β - and

Toxicity of ribonucleosides to mitochondrial RNA polymerase

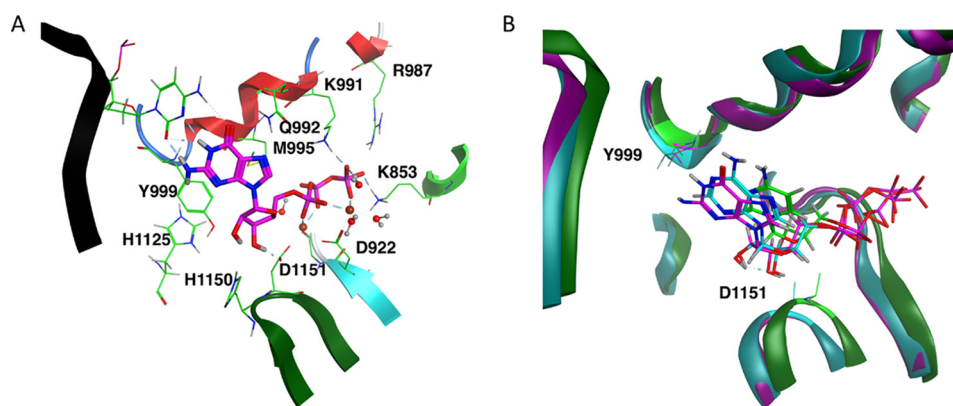


Figure 2. Active site of our computational model of the POLRMT complex. *A*, active site of the GTP-bound complex, with conserved catalytic residues, and residues and water molecules important in coordinating active site Mg^{2+} metal ions displayed in *line* representation and GTP in *stick* mode. Metal ion A is to the left of metal ion B, and both are shown as *spheres*. *B*, comparison between the active sites of the GTP-, ATP-, and CTP-bound POLRMT complexes averaged over simulation trajectories, shown in *magenta*, *teal*, and *green*, respectively. For each system, the NTP molecule is shown in *stick* mode, and α and β carbons of the catalytic residues Tyr⁹⁹⁹ and Asp¹¹⁵¹ are shown in *line* mode.

γ -phosphoryl O atoms of the NTP, the Asp⁹²² side chain, and three water molecules. The NTP is bound to the polymerase principally by the electrostatic interaction between the highly charged triphosphate group and positively charged residues in the active site, namely Lys⁸⁵³, Arg⁹⁸⁷, and Lys⁹⁹¹ (8). As seen in Fig. 2A, the NTP is also stabilized by a sulfur- π interaction with residue M995 belonging to the O-helix in our model.

A bound NTP is thought to be distinguished from a deoxyribonucleoside triphosphate because of the hydrogen bond formed by the NTP 2'-OH to residue Tyr⁹⁹⁹ (8) (Fig. 2). In Fig. 3, we plot the distance between the hydroxyl H atom of the Tyr⁹⁹⁹ side chain and the 2'-hydroxyl O atom of the bound ribonucleotide as a function of time for the second half of the simulation of the CTP-bound complex (*A*) and for the ATP- and GTP-bound complexes (*B* and *C*, respectively). Based on the plots, whereas the hydrogen bond is largely maintained throughout all three simulations, the hydrogen bond distance is significantly larger on average for the CTP-bound complex than for the two purine-nucleotide-bound complexes. This hydrogen bond is particularly stable for the ATP-bound complex.

The nucleotidyl transfer mechanism of POLRMT includes two proton transfer reactions, and conserved residues help to catalyze this mechanism (21–23). The first proton transfer occurs during the nucleophilic attack of the RNA transcript's 3' hydroxyl O atom on the α -phosphate of the incoming NTP to create a new phosphodiester bond. During this step, a proton is transferred from the 3'-hydroxyl O atom to a catalytic base, thought to be Asp¹¹⁵¹ (8) (Fig. 2). In Fig. 3, we plotted the distance between the acidic O atom of the Asp¹¹⁵¹ side chain and the 3'-hydroxyl H atom H3T of the bound nucleotide as a function of time for each of the CTP-, ATP-, and GTP-bound complexes; note that this distance is on average much closer for both the purine-nucleotide-bound complexes than for the CTP-bound complex. Metal A stabilizes this proton transfer by lowering the pKa of the 3'-hydroxyl group. In the course of the nucleotide addition reaction, the pyrophosphate leaving group also becomes protonated in the second proton-transfer reaction, with a residue from the O-helix acting as the proton donor (1), and metal B helping to stabilize this acid-catalysis step.

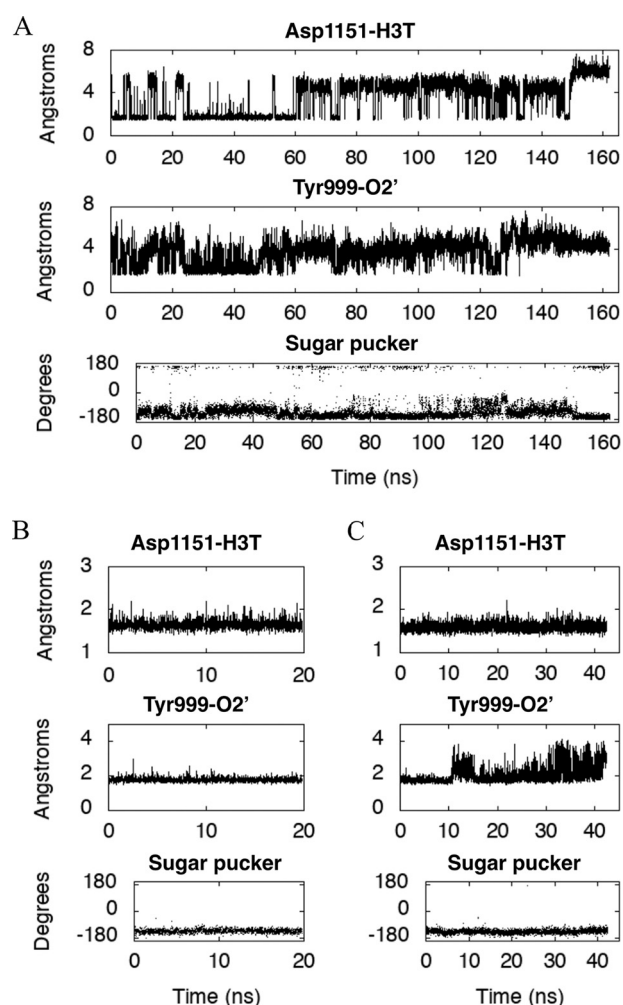


Figure 3. Geometric parameters characterizing the structure of our POLRMT model. Shown are the distances (Å), plotted for our model of the POLRMT complex as a function of simulation time (ns) between the acidic O atom of the Asp¹¹⁵¹ side chain and the 3'-hydroxyl H atom H3T of the bound NTP (*top panels*) and between the hydroxyl H atom of the Tyr⁹⁹⁹ side chain and the 2'-hydroxyl O atom O_{2'} of the bound NTP (*middle panels*). In the *bottom panels*, the sugar pucker of the ribose belonging to the bound NTP is plotted as a function of simulation time (ns). The bound NTP is CTP in *A*, ATP in *B*, and GTP in *C*.

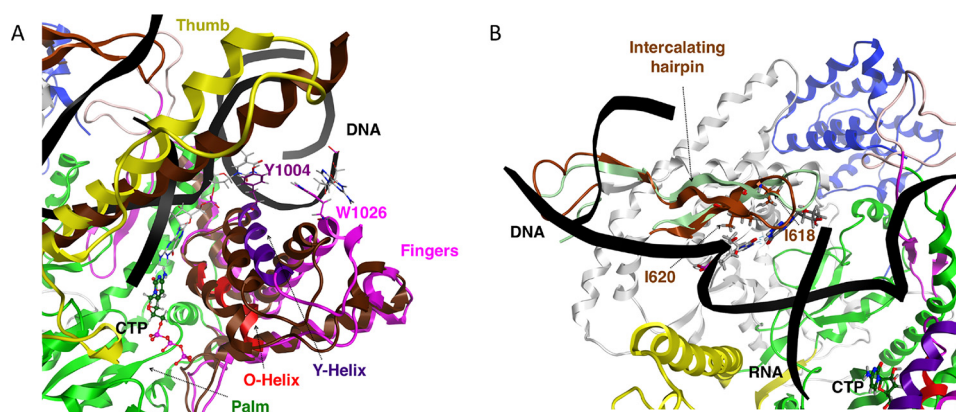


Figure 4. Nucleic acid strand separation in our model of the CTP-bound POLRMT complex. A, the separation of downstream DNA strands by conserved residues in the fingers subdomain. The positions of the fingers and thumb subdomains are compared with their corresponding positions in the 4BOC starting structure, shown in dark brown. Line representation is used to show Tyr¹⁰⁰⁴ and Trp¹⁰²⁶, and the DNA template and nontemplate bases that interact with Tyr¹⁰⁰⁴ and Trp¹⁰²⁶, respectively; CTP is shown in ball-and-stick mode. B, the separation of the upstream DNA template from the RNA transcript by conserved residues belonging to the intercalating hairpin. The position of the intercalating hairpin is compared with its corresponding position in the 4BOC starting structure, shown in light green. Line representation is used to show Ile⁶¹⁸ and Ile⁶²⁰, and the RNA transcript and DNA template bases that interact with Ile⁶¹⁸ and Ile⁶²⁰, respectively; CTP is shown in ball-and-stick mode.

In Fig. 3, we also plot the ribose pucker, or pseudorotation angle, as a function of time for the NTP molecule belonging to each of the three nucleotide-bound complexes. In the case of the ATP- and GTP-bound complexes, the pseudorotation angle generally lies in the interval from -144 to -108° , known as the C4'-endo configuration. For CTP-bound POLRMT, the pseudorotation angle is generally either in the C4'-endo configuration or in the C3'-exo configuration (-180 to -144°) or more rarely in the C2'-endo configuration (144 to 180°) or in the O4'-exo configuration (-108 to -72°) (24).

As mentioned above, during nucleotide addition, the POLRMT catalytic complex cycles from an open state to a closed state and back. It has been suggested that this cycling is driven by the interaction between the O-helix and the incoming NTP (1). As the NTP enters, the complex is in an open state and later forms a catalytically competent closed or clenched state in which the fingers, including the Y-helix, reposition, and the entrance to the active site is temporarily blocked (1). This closed structure, seen in the crystallized structure of the closed product state (8), is known as a clenched conformation. Tyr¹⁰⁰⁴, belonging to the Y-helix, and Trp¹⁰²⁶, belonging to another helix of the fingers domain, base-stack with the base at the +1 position in the template DNA and the base at the 0 position in the nontemplate DNA, respectively. These interactions between the fingers domain and the DNA maintain the separation of the two strands of the downstream DNA (8). In Fig. 4A, we show a comparison between the positions of the fingers in our model of the CTP-bound POLRMT complex and their corresponding positions in the aligned 4BOC structure of the closed product complex. As might be expected, the fingers take on a somewhat more open configuration in our model of the nucleotide insertion state compared with in the closed-product elongation complex.

We have also computed the tilt angle as a function of time for the bundle of helices in the fingers domain including the O- and Y-helices and the helices composed of residues 954–974, 1024–1044, and 1044–1064, respectively, for the CTP-bound POLRMT complex (Fig. 5). It may be seen from the figure that

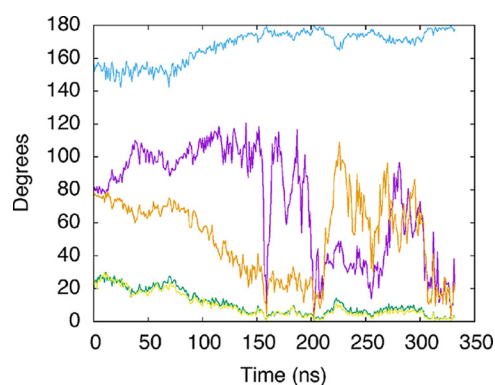


Figure 5. Tilt of axes corresponding to a bundle of helices in the POLRMT finger subdomain. Tilt, with respect to the average axis, is plotted as a function of time for the CTP-bound POLRMT complex. These helices include the helix composed of residues 954–974 (purple line), the O-helix (residues 986–997, green line), the Y-helix (residues 1002–1016, blue line), and the helices composed of residues 1024–1044 (orange line) and 1044–1064 (yellow line).

all of the helices reorient within the bundle. The O and Y helices and the 1044–1064 helix eventually settle into new stable conformations, whereas the helices composed of residues 954–974 and 1024–1044, respectively, continue to undergo large fluctuations in helical tilt during the second half of the simulation.

The thumb domain (residues 705–790) contains a long α -helix that interacts with and stabilizes the DNA template during transcript elongation (1, 8, 25). In Fig. 4A, we compare its position in our model of the insertion state to its position in the crystal structure of the closed product state. Like the fingers domain, it also appears more open in our model of the insertion state compared with in the closed product complex.

The intercalating hairpin of POLRMT (residues 591–624) separates the upstream DNA template strand from the RNA transcript in the elongation complex. The separation of upstream DNA from the transcript allows for the displacement of the transcript and its exit through a channel toward the back of the complex oriented as shown in Figs. 1 and 4. This interaction involves residues Ile⁶²⁰ and Ile⁶¹⁸ in the 4BOC crystal structure (8), and these interactions are retained in our model (Fig. 4B). In our model in Fig. 4B, the intercalating hairpin,

Table 1

Comparison of calculated relative binding free energies of NTP and dNTP analogues to experimental values

Compound	$\Delta\Delta G_{\text{MOE}}$	$\Delta\Delta G_{\text{FEP}}^a$	$\Delta\Delta G_{\text{FEP}}^b$	$-\Delta(\text{p}K_a)_{\text{experimental}}^c$	$\text{Error}_{\text{FEP}}^b$
	kcal/mol	kcal/mol	kcal/mol	kcal/mol	kcal/mol
3'-dATP	0.1	-1.6	-0.5	-0.5	0.0
7-Deaza-ATP	-7.6	1.3	-0.2	0.8	-2.0
2'-Deoxy-2'-fluoro-CTP	-4.4	-0.4	0.2	1.1	-0.9
2'-C-Methyl-CTP	-5.8	0.6	3.2	1.3	1.9
6-Methylpurine-TP	-19.3	0.3	1.2	1.6	-0.4
3-Deaza-ATP	-1.3	2.5	0.1	1.7	-1.6
2'-dATP	1.0	0.0	-0.2	1.8	-2.0
2'-C-Methyl-ATP	1.6	0.5	1.0	2.0	-1.0
4'-Methyl-CTP	-10.4	3.7	1.1	2.0	-0.9
2'-C-Methyl-GTP	7.9	1.1	2.8	2.5	0.3
Ribavirin-TP	11.9	2.5	3.7	3.0	0.7

^a Calculated under the condition of H1125 having a neutral charge state.

^b Calculated under the condition of H1125 having a +1 charge state.

^c Experimental results of Arnold *et al.* (3), except in the case of 2'-C-methyl-ATP; $-\Delta(\text{p}K_a)_{\text{experimental}}$ is taken from the average of the two values reported in Ref. 55.

colored in *brown*, has moved closer to the template DNA strand by comparison to its position in 4BOC, shown in *light green*.

Binding affinities from FEP

Our model contains two active site histidine residues, one of which, His¹¹⁵⁰, is near the ribose moiety of the bound NTP molecule, and the other of which, His¹¹²⁵, is separated from the NTP by residue Tyr⁹⁹⁹. Because the charge states of these two histidine residues may impact the strength of nucleotide binding in our calculations, we first sought to determine these charge states by performing explicitly solvated constant pH molecular dynamics simulations. Simulations were performed both for the ATP-bound POLRMT complex and for the apo POLRMT complex. For both simulations, His¹¹²⁵ was found to have a charge of +1, and His¹¹⁵⁰ was found to be neutral for almost the entire trajectories. On the other hand, when we performed implicitly solvated constant pH molecular dynamics simulations at pH 7.8 of the ATP-bound complex, we found that His¹¹²⁵ was neutral for 89% of the trajectory, whereas His¹¹⁵⁰ was again found to be neutral throughout almost the entire simulation. Because implicitly solvated simulations are expected to be less accurate than explicitly solvated simulations, it is likely that His¹¹²⁵ has a charged state of +1.

Starting with our model of the CTP-bound POLRMT complex obtained by clustering our 333-ns-long MD trajectory as described under “Experimental procedures” or with the final structure from our simulation of either the ATP or the GTP bound complex, we ran lengthy FEP simulations in which the corresponding natural nucleotide was slowly converted into each of a set of nucleotide analogues previously studied experimentally by Arnold *et al.* (3). FEP simulations were then run in the opposite direction, and the results of forward and backward simulation were averaged. The results are given in Table 1. In Fig. 6A we show the correlation between our computational results and the experimental results corresponding to the natural logarithms of the K_D values determined by Arnold *et al.* (3). For each nucleotide analogue, we calculated the experimental value of $\Delta\Delta G$ as the difference between the $\ln(K_D)$ value determined for that nucleotide analogue and the value of the same expression for the corresponding natural nucleotide. The average unsigned error is 0.97 kcal/mol, with a maximum unsigned error of 2.0 kcal/mol. The R^2 value calculated for the linear fit, which is much less relevant than the error for the narrow data

range considered here, is 0.43 ($r = 0.66$). Part of the error is due not only to experimental noise, which has been estimated to typically lie in the range of 0.3–0.5 kcal/mol (18, 26), but also to differences in the sequence of the RNA/DNA hybrid used in the experiments compared with the sequence of the one present in the crystal structure that was used in our computations. These results demonstrate that FEP can be applied to determine relative binding free energies of NTP analogues to a polymerase with a significant degree of accuracy.

Because there was some doubt concerning the protonation state of residue His¹¹²⁵, in Fig. 6B we show the same correlation, this time calculated under the condition of His¹¹²⁵ being neutral. Here the average unsigned error (1.14 kcal/mol) is slightly higher than for His¹¹²⁵ protonated, and the R^2 value (0.44) is approximately the same, supporting the protonated state of His¹¹²⁵.

Comparison to other binding affinity calculations

We also performed single-point free-energy calculations of the NTP-binding affinities to POLRMT in the molecular operating environment (MOE) program for comparison. An implicit solvent representation was used in this approach, namely the reaction field method. The range of computed values falls far outside of the experimental range, and the square of the correlation coefficient R^2 between the resulting affinities and the experimental values is only 0.13 (Fig. 6 and Table 1). Thus the FEP results are much more accurate than the quick single-point free-energy predictions. Calculations performed in MOE could, however, be useful in a preliminary screen for nucleotide analogues binding to a polymerase, prior to FEP screening, by narrowing the search space.

In a previous study on a related topic, Florian and co-workers (27) applied FEP to compute the difference between the binding free energy of dCTP and dTTP to a DNA duplex. In that study it was concluded that at least 20 ns of simulation time was required for sufficient sampling (~80 ns was used for each compound in the present study). In addition to the availability of adequate computational resources, another difficulty that has hindered efforts at performing free energy perturbation calculations is the intricacy of the setup, because programs such as NAMD do not offer automated scripts for creating the required dual topology files. In the present work, we have taken advantage of state-of-the-art computing power and in-house scripts

Toxicity of ribonucleosides to mitochondrial RNA polymerase

fect sampling. Thus the close correspondence between our results and experimental values is attributable to the availability of not only low-resolution crystal structures but also highly accurate force fields and superb high-performance computing resources.

The application of FEP to model interactions between NTP analogues and the catalytic site of mitochondrial RNA polymerase will be of use, along with mitochondrial RNA polymerase catalytic assays, to identify potential inhibitors and also in the rational design and salvaging of inhibitory drugs. Moreover, applying this technology to viral-encoded and other pathogen-encoded polymerases and other key replicative enzymes will aid in the rational design of specific small molecule inhibitors, as well as being of value in the rational design of specific inhibitors of various host cell and cancer targets.

Experimental procedures

System setup and equilibration

Our model of POLRMT is based upon the 4BOC crystal structure of the protein in complex with a 28-mer dsDNA including a bubble region of 9 bases and a 14-mer RNA transcript (8). The crystal structure has a resolution of 2.65 Å and shows the entire DNA template except for the 5' base and all of the nontemplate DNA except for the 5' base and the bubble region. It includes a 9-bp DNA/RNA hybrid. Because a few loops were missing from the crystal structure of the protein, we used homology modeling in MOE (37) to complete this structure, for which the DNA and RNA were used as an environment for the induced fit. To model the incoming CTP and Mg^{2+} ions, the position of the base of the 3'-terminal RNA nucleotide was preserved, whereas the ribose and triphosphate atoms were modeled based on the 1QLN structure of the insertion state of RNA polymerase from bacteriophage T7 (36), simultaneously superimposed on 4BOC, and on the 1T7P structure of the homologous T7 DNA polymerase (12), because this structure includes active site Mg^{2+} ions.

Residue protonation states were assigned at a pH of 7.8 by the MOE program (37), corresponding to mitochondrial pH (although mitochondrial pH has been shown to fluctuate, depending upon metabolic state, from 7.5 to 8.2 (38)). The protonation states of the active site histidine residues 1125 and 1150 were evaluated using constant pH molecular dynamics (CphMD) (39, 40). Simulations were performed for both the ATP-bound POLRMT complex and for the apo-complex in explicit solvent and for the apo-complex in implicit solvent. During explicitly solvated CphMD simulations (40), protonation state changes were periodically attempted every 100 steps with a time step of 2 fs, *i.e.* every 0.002 ps, and production simulations (following heating and equilibration) were conducted for 800 ps (ATP-bound) or 500 ps (apo state). During implicitly solvated CphMD simulations (39), protonation state changes were attempted at every time step (2 fs), and production simulations were conducted for 2 ns.

Simulation setup was performed in the leap module of the AMBER 14 program package (41). We used parameters for CTP kindly provided by Francois-Yves Dupradeau through the RED database (42). The protein and nucleic acid were parame-

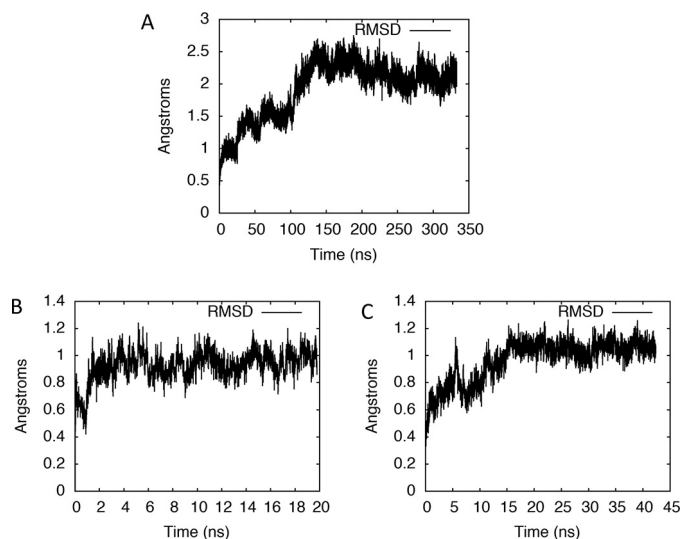


Figure 7. RMSD plots showing convergence and stability of simulations. Shown is the RMSD, relative to the starting structure, of the active site of our computational model of the POLRMT complex bound by CTP, A, as a function of simulation time. RMSD plots for simulations initiated after replacing the CTP in the active site by ATP or GTP at the end of the MD simulation of the CTP-bound complex are shown in B and C, respectively.

terized by the AMBER ff12SB force field (43–46). A cubic box of pre-equilibrated explicit TIP3P water was added extending to a distance of 15 Å from the protein. The total charge on the system was neutralized by the addition of Na^+ ions, and then an additional 137 Na^+ ions and the same number of Cl^- ions were added, bringing the ionic concentration to 150 mM. Molecular dynamics simulations were performed with periodic boundary conditions using the pmemd.MPI module of AMBER 14 (41). The water surrounding the molecular system was first relaxed, and then the entire system was relaxed. Next, restraining all solute atoms by a force constant of 25 kcal/mol, the system was heated to a final temperature of 300 K at a constant volume over a period of 50 ps using a Langevin thermostat. Restraints were then gradually relaxed to zero over 100 ps with a time step of 2 fs at constant pressure, and then a 2 ns equilibration was performed also at constant pressure.

We then ran a total of 333 ns of MD simulations in NAMD (30). The plot of active site root-mean-square deviation (RMSD) versus time became relatively flat approximately halfway through the simulation (Fig. 7A), and thus the last 135 ns of the trajectory was used in cluster analysis. Trajectory snapshots were clustered with respect to positions of active site residues into 18 sets (based on Davies Bouldin index (DBI) values and SSR/SST values (the ratio of the sum of squares regression to the total sum of squares)). A representative of the largest cluster was chosen as our starting model for binding free-energy simulations. This model of the CTP-bound POLRMT complex was also used for our depictions in Figs. 1, 2A, and 4.

Starting with our model of the CTP-bound POLRMT complex, the CTP residue and its base-pairing dG residue in the template DNA strand were manually mutated to GTP-dC and ATP-dT pairs, respectively. We then performed MD simulations of the GTP- and ATP-bound POLRMT complexes, respectively. Because the active site RMSD stabilized very fast for the ATP-bound complex, this MD simulation was run for

only 20 ns (Fig. 7B), whereas the simulation of the GTP-bound complex, which took longer to equilibrate based on the RMSD plot, was carried out for 42 ns (Fig. 7C). The final structures from our simulations of the ATP or the GTP-bound complexes, respectively, were used as starting models for free-energy simulations.

Free-energy simulations

Free-energy simulations were performed using the free energy perturbation method implemented in NAMD (30). Ligands were parameterized with the MATCH program (31), whereas protein and nucleic acid residues were parameterized using CHARMM22 (32) and CHARMM27 (33, 34), respectively. Dual topology files and coordinates were then created in NAMD. For each system, two sets of input files were created: one for simulation of the ligand in aqueous 0.150 M NaCl solution and another for simulation of the same ligand within the solvated protein-bound environment also containing 0.150 M NaCl, so that binding free energies could then be determined using a thermodynamic cycle. The computational details were as follows.

Systems were first minimized for 2000 steps and then gradually heated to 300 K (to match experimental conditions) using a Langevin thermostat over 60 ps with harmonic restraints. The restraints were gradually released over a period of 250 ps, and afterward systems were equilibrated for 1 ns, using a 0.5-fs time step. The coupling parameter, λ , which describes the fractional potential energy character of the second molecule associated with each window, was then scaled from 0 to 1 by increments of 0.02, *i.e.* over 50 simulation windows. In each window, equilibration was performed over 200,000 steps followed by 200,000 steps of data collection, with a time step of 2 fs. We made use of a “soft core” potential, *i.e.* van der Waals and electrostatic potentials were scaled separately to prevent atom clashes, which might occur if atoms appear or disappear in a highly charged state. The particle mesh Ewald method was used to treat long-range electrostatics, and SHAKE (47) was used to constrain bonds involving hydrogen atoms. The system energy was sampled every 10 steps. After having performed 40-ns simulations in the forward direction from the corresponding natural NTP to the nucleotide analogue, simulations were also run in the backward direction for 40 ns for each system, and the two results were averaged. For each alchemical transformation, the free-energy difference $\Delta\Delta G$ is found as (13–15, 30),

$$\Delta\Delta G = -k_B T \sum_{i=1}^N \ln \langle \exp[-(H(\mathbf{x}, \mathbf{p}, \lambda_{i+1}) - H(\mathbf{x}, \mathbf{p}, \lambda_i))/k_B T] \rangle$$

(Eq. 1)

where H is the system Hamiltonian; N is the number of windows, in our case 50; k_B is the Boltzmann constant; T is absolute temperature; and the angle brackets denote averaging over all sampled phase space points (\mathbf{x}, \mathbf{p}) .

Single-point free-energy calculations

For single-point free-energy calculations, we used the MOE program (37) with the reaction field solvation model and the AMBER 10:EHT force field. In this force field, parameters of

proteins and nucleic acids are taken from the AMBER *ff10* force field (48), and those for small molecules are parameterized using 2D extended Hueckel theory (49). Energy minimizations were performed on nucleotide analogues placed in the NTP binding site of POLRMT prior to calculations of binding affinities.

Trajectory analysis

RMSD values of the active site relative to the initial configuration were calculated using the AMBER analysis tool *cpttraj* (50) for the 333-ns MD simulation of the CTP-bound POLRMT complex, the 20-ns MD simulation of the ATP-bound POLRMT complex, and the 42-ns MD simulation of the GTP-bound POLRMT complex. The active site was defined as all residues, not including the bound NTP, within 6 Å of the bound NTP. Selected distances, pucker angles, and average structure files were also calculated using *cpttraj* (50) along the entire trajectories of the ATP-bound and GTP-bound POLRMT complexes and for the second half of the simulation of the CTP-bound complex. Average structure files are used in our depictions in Fig. 2B.

The Gromacs (51–53) analysis tool *gmx_bundle* was employed in analyzing the bundle of helices belonging to the POLRMT finger subdomain, to calculate their axial tilts with respect to the average axis. Prior to applying this tool, the 333-ns NAMD trajectory of the CTP-bound catalytic complex was converted to Gromacs format with the *acpype-0.1.0* program (54).

Author contributions—H. F. and M. H. conceptualization; H. F. and P. W. software; H. F. formal analysis; H. F. investigation; H. F. visualization; H. F. methodology; H. F. writing-original draft; H. F., P. W., J. T., D. L. T., and M. H. writing-review and editing; J. T., D. L. T., and M. H. supervision; D. L. T. and M. H. funding acquisition.

Acknowledgment—We thank Wendy Magee for helpful discussion.

References

- Arnold, J. J., Smidansky, E. D., Moustafa, I. M., and Cameron, C. E. (2012) Human mitochondrial RNA polymerase: structure-function, mechanism and inhibition. *Biochim. Biophys. Acta* **1819**, 948–960 [CrossRef Medline](#)
- Jin, Z., Kinkade, A., Behera, I., Chaudhuri, S., Tucker, K., Dyatkina, N., Rajwanshi, V. K., Wang, G., Jekle, A., Smith, D. B., Beigelman, L., Symons, J. A., and Deval, J. (2017) Structure-activity relationship analysis of mitochondrial toxicity caused by antiviral ribonucleoside analogs. *Antiviral Res.* **143**, 151–161 [CrossRef Medline](#)
- Arnold, J. J., Sharma, S. D., Feng, J. Y., Ray, A. S., Smidansky, E. D., Kireeva, M. L., Cho, A., Perry, J., Vela, J. E., Park, Y., Xu, Y., Tian, Y., Babusis, D., Barauskas, O., Peterson, B. R., *et al.* (2012) Sensitivity of mitochondrial transcription and resistance of RNA polymerase II dependent nuclear transcription to antiviral ribonucleosides. *PLoS Pathog.* **8**, e1003030 [CrossRef Medline](#)
- Feng, J. Y., Xu, Y., Barauskas, O., Perry, J. K., Ahmadyar, S., Stepan, G., Yu, H., Babusis, D., Park, Y., McCutcheon, K., Perron, M., Schultz, B. E., Sankowicz, R., and Ray, A. S. (2016) Role of mitochondrial RNA polymerase in the toxicity of nucleotide inhibitors of hepatitis C virus. *Antimicrob. Agents Chemother.* **60**, 806–817 [CrossRef Medline](#)
- Feng, J. Y., Tay, C. H., and Ray, A. S. (2017) Role of mitochondrial toxicity in BMS-986094-induced toxicity. *Toxicol. Sci.* **155**, 2 [CrossRef Medline](#)

Toxicity of ribonucleosides to mitochondrial RNA polymerase

- Cihlar, T., and Ray, A. S. (2010) Nucleoside and nucleotide HIV reverse transcriptase inhibitors: 25 years after zidovudine. *Antiviral Res.* **85**, 39–58 [CrossRef Medline](#)
- Nelms, M. D., Mellor, C. L., Cronin, M. T., Madden, J. C., and Enoch, S. J. (2015) Development of an *in silico* profiler for mitochondrial toxicity. *Chem. Res. Toxicol.* **28**, 1891–1902 [CrossRef Medline](#)
- Schwinghammer, K., Cheung, A. C., Morozov, Y. I., Agaronyan, K., Temiakov, D., and Cramer, P. (2013) Structure of human mitochondrial RNA polymerase elongation complex. *Nat. Struct. Mol. Biol.* **20**, 1298–1303 [CrossRef Medline](#)
- Ng, K. K., Arnold, J. J., and Cameron, C. E. (2008) Structure-function relationships among RNA-dependent RNA polymerases. *Curr. Top. Microbiol. Immunol.* **320**, 137–156 [Medline](#)
- Masters, B. S., Stohl, L. L., and Clayton, D. A. (1987) Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* **51**, 89–99 [CrossRef Medline](#)
- Cermakian, N., Ikeda, T. M., Miramontes, P., Lang, B. F., Gray, M. W., and Cedergren, R. (1997) On the evolution of the single-subunit RNA polymerases. *J. Mol. Evol.* **45**, 671–681 [CrossRef Medline](#)
- Doublé, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251–258 [CrossRef Medline](#)
- Zwanzig, R. W. (1954) High-temperature equation of state by a perturbation method: I. nonpolar gases. *J. Chem. Phys.* **22**, 1420–1426 [CrossRef](#)
- Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997) The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **72**, 1047–1069 [CrossRef Medline](#)
- Beveridge, D. L., and DiCapua, F. M. (1989) Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Chem.* **18**, 431–492 [CrossRef Medline](#)
- Deng, Y., and Roux, B. (2009) Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B.* **113**, 2234–2246 [CrossRef Medline](#)
- Karplus, M., Gelin, B. R., and McCammon, J. A. (1980) Internal dynamics of proteins. Short time and long time motions of aromatic sidechains in PTL. *Biophys. J.* **32**, 603–618 [CrossRef Medline](#)
- Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lypyan, D., Robinson, S., Dahlgren, M. K., Greenwood, J., Romero, D. L., Masse, C., Knight, J. L., Steinbrecher, T., Beumung, T., *et al.* (2015) Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **137**, 2695–2703 [CrossRef Medline](#)
- Kollman, P. A. (1993) Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* **93**, 2395–2417 [CrossRef](#)
- Choi, K. H. (2012) Viral polymerases. *Adv. Exp. Med. Biol.* **726**, 267–304 [CrossRef Medline](#)
- Castro, C., Smidansky, E. D., Arnold, J. J., Maksimchuk, K. R., Moustafa, I., Uchida, A., Götte, M., Konigsberg, W., and Cameron, C. E. (2009) Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat. Struct. Mol. Biol.* **16**, 212–218 [CrossRef Medline](#)
- Shen, H., Sun, H., and Li, G. (2012) What is the role of motif D in the nucleotide incorporation catalyzed by the RNA-dependent RNA polymerase from poliovirus? *PLoS Comput. Biol.* **8**, e1002851 [CrossRef Medline](#)
- Cameron, C. E., Moustafa, I. M., and Arnold, J. J. (2009) Dynamics: the missing link between structure and function of the viral RNA-dependent RNA polymerase? *Curr. Opin. Struct. Biol.* **19**, 768–774 [CrossRef Medline](#)
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*, p. 19, Springer, New York
- Velazquez, G., Sousa, R., and Brieba, L. G. (2015) The thumb subdomain of yeast mitochondrial RNA polymerase is involved in processivity, transcript fidelity and mitochondrial transcription factor binding. *RNA Biol.* **12**, 514–524 [CrossRef Medline](#)
- Brown, S. P., Muchmore, S. W., and Hajduk, P. J. (2009) Healthy skepticism: assessing realistic model performance. *Drug Discov. Today.* **14**, 420–427 [CrossRef Medline](#)
- Bren, U., Martínek, V., and Florian, J. (2006) Free energy simulations of uncatalyzed DNA replication fidelity: structure and stability of T.G and dTTP.G terminal DNA mismatches flanked by a single dangling nucleotide. *J. Phys. Chem. B.* **110**, 10557–10566 [CrossRef Medline](#)
- Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., and Shaw, D. E. (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In: *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, Tampa, FL, November 11–17, 2006*, Article 84, Association for Computing Machinery, New York
- Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theory Comput.* **6**, 1509–1519 [CrossRef Medline](#)
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 [CrossRef Medline](#)
- Yesselman, J. D., Price, D. J., Knight, J. L., and Brooks, C. L., 3rd (2012) MATCH: an atom-typing toolset for molecular mechanics force fields. *J. Comput. Chem.* **33**, 189–202 [CrossRef Medline](#)
- MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., *et al.* (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* **102**, 3586–3616 [CrossRef Medline](#)
- Foloppe, N., and MacKerell, J. A. (2000) All-atom empirical force field for nucleic acids: 1. parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **21**, 86–104 [CrossRef](#)
- MacKerell, J., and Banavali, N. (2000) All-atom empirical force field for nucleic acids: 2) application to molecular dynamics simulations of DNA and RNA in solution. *J. Comp. Chem.* **21**, 105–120 [CrossRef](#)
- Cheatham, G. M., Jeruzalmi, D., and Steitz, T. A. (1999) Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399**, 80–83 [CrossRef Medline](#)
- Cheatham, G. M., and Steitz, T. A. (1999) Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309 [CrossRef Medline](#)
- Chemical Computing Group, Inc. (2013) Molecular operating environment (MOE), **2013.08**
- Santo-Domingo, J., and Demaurex, N. (2012) Perspectives on: SGP symposium on mitochondrial physiology and medicine: the renaissance of mitochondrial pH. *J. Gen. Physiol.* **139**, 415–423 [CrossRef Medline](#)
- Mongan, J., Case, D. A., and McCammon, J. A. (2004) Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **25**, 2038–2048 [CrossRef Medline](#)
- Swails, J. M., York, D. M., and Roitberg, A. E. (2014) Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: implementation, testing, and validation. *J. Chem. Theory Comput.* **10**, 1341–1352 [CrossRef Medline](#)
- Case, D. A., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., Cheatham Iii, T., Darden, T., Duke, R., and Gohlke, H. (2014) AMBER 14
- Vanquelef, E., Simon, S., Marquand, G., Garcia, E., Klimerek, G., Delepine, J. C., Cieplak, P., and Dupradeau, F. Y. (2011) R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res.* **39**, W511–W517 [CrossRef Medline](#)
- Wang, J. M., Cieplak, P., and Kollman, P. A. (2000) How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074 [CrossRef](#)
- Cheatham, T. E., 3rd, Cieplak, P., and Kollman, P. A. (1999) A modified version of the Cornell *et al.* force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **16**, 845–862 [CrossRef Medline](#)
- Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T. E., 3rd, Laughton, C. A., and Orozco, M. (2007) Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Bio-phys. J.* **92**, 3817–3829 [CrossRef Medline](#)

46. Zgarbová, M., Otyepka, M., Sponer, J., Mládek, A., Banás, P., Cheatham, T. E., 3rd, and Jurecka, P. (2011) Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 [CrossRef Medline](#)
47. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. (1977) Numerical integration of Cartesian equations of motion of a system with constraints: molecular-dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341 [CrossRef](#)
48. Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, K. M., Crowley, M., Walker, R. C., Zhang, W., Merz, K. M., Wang, B., Hayik, S., Roitberg, A., Seabra, G., *et al.* (2008) AMBER 10
49. Gerber, P. R., and Müller, K. (1995) MAB, a generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput. Aided Mol. Des.* **9**, 251–268 [CrossRef Medline](#)
50. Roe, D. R., and Cheatham, T. E., 3rd. (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 [CrossRef Medline](#)
51. Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory. Comput.* **4**, 435–447 [CrossRef Medline](#)
52. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 [CrossRef Medline](#)
53. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 [CrossRef Medline](#)
54. Sousa da Silva, A. W., and Vranken, W. F. (2012) ACPYPE: Ante-Chamber PYthon Parser interface. *BMC Res. Notes* **5**, 367 [CrossRef Medline](#)
55. Smidansky, E. D., Arnold, J. J., Reynolds, S. L., and Cameron, C. E. (2011) Human mitochondrial RNA polymerase: evaluation of the single-nucleotide-addition cycle on synthetic RNA/DNA scaffolds. *Biochemistry* **50**, 5016–5032 [CrossRef Medline](#)