Accurate physics-based prediction of binding affinities of RNA and DNA targeting ligands

Ara M. Abramyan[1*], Anna Bochicchio[2,6*], Chuanjie Wu[3], Wolfgang Damm[3], David R. Langley[4], Devleena Shivakumar[3,7], Dmitry Lupyan[5], Lingle Wang[3], Edward Harder[3], Eliud O. Oloo[5]

[1]Schrödinger Inc, San Diego, CA, USA; [2]Schrödinger GmbH, Munich, Germany; [3]Schrödinger Inc, New York, NY, USA; [4]Arvinas Inc, New Haven, CT, USA; [5]Schrödinger Inc, Cambridge, MA, USA; [6]Present Address: Novartis BioMedical Research, 4002 Basel, Switzerland; [7] Present Address: MRL, Merck & Co., Inc., South San Francisco, CA, USA

*Co-first authors
Contacts: ara.abramyan@schrodinger.com, anna.bochicchio@novartis.com, eliud.oloo@schrodinger.com

ABSTRACT

Accurate prediction of the affinity of ligand binding to nucleic acids represents a significant challenge for current computational approaches. This limitation has hindered the use of computational methods to develop small molecule drugs that modulate the activity of nucleic acids, including those associated with anticancer, antiviral and antibacterial effects. In recent years, significant scientific and technological advances as well as easier access to compute resources have contributed to free-energy perturbation (FEP) becoming one of the most consistently reliable approaches for predicting relative binding affinities of ligands to proteins. Nevertheless, FEP's applicability to nucleic-acid targeting ligands has remained largely undetermined. In this work, we present a systematic assessment of the accuracy of FEP, as implemented in the FEP+ software and facilitated by improvements in the OPLS4 force field, in predicting relative binding free energies of congeneric series of ligands interacting with a variety of DNA/RNA systems. The study encompassed more than 100 ligands exhibiting diverse binding modes, some partially exposed and others deeply buried. Using a consistent simulation protocol, more than half of the predictions are within 1 kcal/mol of the experimentally measured values. Across the dataset, we report a combined average pairwise root-mean-square-error (RMSE) of 1.39 kcal/mol, which falls within one log unit of the experimentally measured

dissociation constants. These results suggest that FEP+ has sufficient accuracy to guide the optimization of lead series in drug discovery programs targeting RNA and DNA.

INTRODUCTION

The development of small molecule modulators of nucleic acids has remained a topic of growing interest in drug discovery over the years. Efforts to drug nucleic acids are, in part, motivated by the fact that the number of possible nucleic acid targets far exceeds the number of druggable protein targets. It is estimated that only 1.5% of the human genome encodes for proteins of which a small fraction, approximately 12%, has been linked to diseases, and an even smaller fraction of 3% is targeted by existing drugs (Djebali et al., 2012; Lander et al., 2001). Due to the important role nucleic acids play in transcription, replication and a host of other functions, modulating them and their pathways has the potential to greatly expand the space of viable targets essential to key cellular functions and diseases associated biological processes (Cech, 2012; Cech and Steitz, 2014).

As drug targets, nucleic acids are widely acknowledged to be substantially more challenging to design ligands for than proteins. As a result, only a handful of compounds modulating RNA have so far been approved by the FDA: ataluren, linezolid, ribocil and risdiplam (Childs-Disney et al., 2022; Kovachka et al., 2024; Yu et al., 2020). At the moment, phenotypic screening is the dominant approach for discovering nucleic acid drugs with scant reports of successful structure-based design available; an attestation to the intractability of these molecules to conventional methods for structure determination and computational modeling and simulation. The difficulty in dealing with nucleic acids stems from several factors. From a physicochemical perspective, nucleic acid molecules harbor inherent chemical and structural characteristics that often confound traditional technologies and approaches that have been applied with greater success to proteins like X-crystallography and NMR (Ma et al., 2022; Zhang et al., 2022). For one, they possess a highly flexible ribose and phosphate backbone, which contains more degrees of freedom than proteins thus requiring more rigorous sampling. In addition, they exhibit weak long-range interactions that insufficiently hold the tertiary structure together making them heterogeneous and thus difficult to structurally characterize, particularly RNA. Even when more structured, as is more likely to be the case for DNA, the polyionic nature of nucleic acids generally precludes deep hydrophobic pockets of the kind that accommodate ligands in protein targets. While cationic ligands could in theory be readily designed to interact with nucleic acids,

such ligands tend to be promiscuous and prone to off-target binding (Kovachka et al., 2024). The safety risk and other consequences of non-specific DNA damage caused by chemotherapeutic treatment are particularly concerning. This reinforces the urgent need for new approaches that can improve the design of potent ligands while also maintaining specificity and other properties essential for safety and efficacy.

The challenge of simultaneously achieving potency and safety may be facilitated through structure-based rational drug design using computational techniques capable of accelerating the accurate identification of promising hits and speeding up the optimization of candidate molecules. Although at present there are only about 14,000 nucleic acid-containing structures in the PDB and NDB repositories, advancements in technologies and methods continue to expand the number of available structures (Berman et al., 2022; Lawson et al., 2024; Ma et al., 2022; Westhof and Leontis, 2021). In parallel, robust computational approaches need to be created to maximize the benefit of the expected growth in RNA and DNA structural data. Notably, the same properties that make nucleic acid molecules recalcitrant to structural characterization by experimental techniques also pose difficulties in computational modeling (Schneider et al., 2023; Zhang et al., 2022). The inherent flexibility of these molecules increases the vastness of the conformation space to be sampled computationally while the charges in the sugar-phosphate backbone result in a highly charged simulation environment, including counterions, whose parameterization is far from straightforward. A robust and reliable force field thus needs to account for long-range electrostatic and solvation models that include explicit metal ions and water.

Among the available frameworks for detailed modeling of the binding free energies of biomolecules and ligands, free energy perturbation (FEP) has proven to be the most consistently accurate(Kuhn et al., 2020; Ross et al., 2023; Song et al., 2019; Wang et al., 2019, 2015; Zhu et al., 2022). FEP methods employ atomistic molecular dynamics or Monte Carlo simulations to calculate the relative free-energy between two related ligands whereby one is perturbed to the other throughout the simulation. Enthalpic and entropic effects are accounted for by explicitly modeling conformational flexibility, solvent molecules, ions and other important conditions like temperature. The method has been validated extensively in both retrospective studies and prospective commercial applications (Ross et al., 2023; Wang et al., 2015). FEP+ is a particular free-energy perturbation protocol that builds upon the advantages mentioned above by offering a robust force field, enhanced sampling techniques and an automated workflow that

simplifies calculation set up and analysis. Over the last few years, FEP+ has undergone a continuous stream of improvements. These include enhancements to non-charged R-group perturbations in small molecules (Wang et al., 2015, 2013), more accurate handling of charge-charge perturbations (Chen et al., 2018), core-hopping (Wang et al., 2017), pKa correction (de Oliveira et al., 2019), macrocyclization (Yu et al., 2017), fragment-linking (Yu et al., 2021) and covalent FEP (Yu et al., 2019). Most recently the method has been extended to estimating binding free energies of non-congeneric small molecules and also absolute binding FEP (ABFEP) (Chen et al., 2023) as well as for the assessment of crystalline and amorphous small molecule solubilities. Beyond small molecules, recent advances have improved FEP+ accuracy in predicting the impact of protein mutation on stability and affinity, and ligand selectivity (Clark et al., 2019; Duan et al., 2020; Hauser et al., 2018; Scarabelli et al., 2022).

To our knowledge, the applicability of FEP to nucleic-acid targeting ligands remains largely unexplored. Only a few examples reporting such use cases can be found in the literature (Chen et al., 2019). Despite the repeated success of FEP+ in predicting the relative binding free energies of protein-ligand and protein-protein systems, the protocol's performance on ligand binding to nucleic acids has, up to this point, not been ascertained. We therefore sought to quantitatively assess the accuracy of FEP in calculating the relative free energies of binding of congeneric series of ligands interacting with DNA and RNA. One of the motivations for this effort is the fact that in addition to prior enhancements to FEP+, more recent work has been done to improve OPLS4 force field parameters for RNA and DNA modeling. Improvements to atomic charges, co-valence terms and torsions for nucleic acids were shown to result in measurably more stable molecular dynamics simulation trajectories for various nucleic acid types including B-DNA, A-DNA, Z-DNA, RNA, PNA, LNA, and nucleic acid/protein complex systems (see supporting information). The integration of these parameter enhancements into OPLS4 has broadly improved force field coverage for nucleic acids. This was verified by reproducing the dynamics and physicochemical properties of various forms of nucleic acid structures. Taking advantage of the aforementioned force field improvements, we performed a retrospective validation study involving 8 targets across over 100 ligands spanning a variety of DNA and RNA systems representing a total of 270 perturbations.

MATERIALS AND METHODS

Dataset curation

To embark on this study, we aimed to assemble a comprehensive benchmark dataset composed of high quality, publicly available binding affinity measurements performed on experimentally well-resolved structures and congeneric series. Criteria for selection included the requirement that an experimental structure of the nucleic acid-ligand complex exists for at least one ligand in the congeneric series. We also ensured that there were no significant structural ambiguities in the receptor structure, and that the binding data for the congeneric series were *in vitro* measurements of dissociation constants (Kd), inhibition constants (Ki), or ligand concentrations that achieve 50% inhibition (IC50s). While one of the series we identified from literature, the Influenza A virus RNA promoter binders (Bottini et al., 2015), met the above criteria, we excluded it from our analysis due to several reasons: all of its ligands were weak binders (from ~34 μM to 540 μM); the range of affinities was very narrow only spanning 1 log unit; the affinities obtained from two experimental measurements, IC50 and Kd had virtually no correlation (see Figure S10). Lastly, we omitted machine-learning (ML) predicted nucleic acid structures for this work; while their number is growing, the limited number and quality of the existing experimentally determined nucleic structures (especially compared to proteins) currently present considerable challenges (Schneider et al., 2023).

Structure preparation and FEP+ calculation setup

The calculations for this study were conducted using the Schrödinger molecular modeling software suite, release 2023-3. The reference structures used were obtained from the Protein Data Bank (PDB) (Berman et al., 2000) and prepared using the Protein Preparation Workflow (PPW) in the Schrödinger software suite (Sastry et al., 2013). During structure preparation, force field atom types, bond orders and tautomer/ionization states were assigned, water orientations were sampled and constrained energy minimization using the default threshold of 0.3 Å RMSD was performed. For the integrase system, which contains a protein in addition to DNA, the Asparagine, Glutamine, and Histidine residue side chains were flipped to optimize the hydrogen bond networks.

Ligand structures and affinity data were obtained from the literature. The structures were prepared using LigPrep (“Schrödinger Release 2023–3: Ligprep,” 2023). Chiral centers were retained as specified in the literature, ionization states and tautomer forms were enumerated at

pH 7.4 ± 2.0 with Epik (Shelley et al., 2007). All relevant protonation and tautomeric states were added to both Glide (Friesner et al., 2004; Halgren et al., 2004) docking and the relative binding FEP map. For the latter, the relative free energies were calculated and accounted for using a previously described pKa correction protocol (de Oliveira et al., 2019).

The OPLS4 force field was used for the calculations (Lu et al., 2021). This force field version incorporates parameter improvements specific to nucleic acids. Briefly, RNA/DNA force field improvements incorporated in OPLS4 encompassed optimization of partial charges as well as bond stretching, bending and torsion parameters. Partial charges were fitted to the electrostatic potential at HF/6-31G* level with the model compounds. Torsional energies were fitted to the LMP2/cc-pVTZ(-f) single-point energies on the structures minimized with B3LYP/6-31G*. For χ, ε and ζ torsions, each of them were fitted with three rotamer sets to represent the B-DNA, A-DNA, and RNA structures. The three rotamer sets were fitted together for one of the torsions. To reproduce correct sugar puckering, sugar torsions were fitted by sampling the sugar ring conformations with both the O3'-methyl and O3'-phosphate groups with different nucleic bases for both deoxyribose and ribose rings. Further details on the force field parameterization and fitting procedure are provided in the Supporting Information. The Force Field Builder tool (Schrodinger release 2023-2) was used to automatically generate accurate force field torsional parameters for all ligands containing substructures not fully covered by the standard OPLS4 parameters (Lu et al., 2021).

Starting poses for relative binding FEP simulations were generated by docking ligands into the binding site using the Maximum Common Substructure (MCS) Docking Alignment protocol (Cappel et al., 2020) in Glide, with core constraints on the reference ligand core atoms to ensure that the core made similar interactions for all ligands in the series.

Binding affinities were predicted using the relative binding FEP methodology in FEP+, which has been described elsewhere (Wang et al., 2015). In short, relative binding FEP combines free energy perturbation (FEP) as described originally by Zwanzig (Zwanzig, 1954) with an accurate modern force field (OPLS4), efficient GPU-enabled parallel molecular dynamics with Desmond (Bowers et al., 2006), REST enhanced sampling (Wang et al., 2013), cycle-closure correction (Wang et al., 2013) to combine redundant information into robust free energy estimates, and the FEP+ Mapper GUI to automate setup and analysis of the calculations.

The systems were solvated in an orthorhombic box of water molecules with a buffer width of 5 Å for neutral and 8 Å for charge-changing perturbations in the complex leg and 10 Å in the solvent leg simulations. All the simulation stages used the simple point charge (SPC) water model (Berendsen et al., 1987) with sodium chloride ions added at 0.15 M concentration and Na+

counterions added to neutralize the systems. It is worth noting that in the standard FEP+ protocol explicit ions are only added for charge-changing perturbations, whereas in cases where there is no alchemical charge change, the standard Desmond method is used wherein a small charge imbalance is spread over the entire system. This method usually produces good results in terms of the convergence for protein-based systems. However, for nucleic acid systems it has been shown that adding explicit ions resulted in greater stability of the systems and better results, hence we included explicit ions in our systems studied here (Cheatham III and Case, 2013; Sharp and Honig, 1995; Tan and Chen, 2011). The default FEP+ relaxation protocol was used as described previously (Wang et al., 2015). In the FEP+ production run, 12/16/24 λ windows were used for neutral/core-hopping/charge-changing types of the perturbations. The MD production stage was run for 20 ns per λ-window for both complex and solvent simulations, which was longer than the software default minimum of 5ns. The temperature was kept at 300 K and constant volume maintained for the complex simulations and constant pressure for the solvent simulations. The pressure was controlled using the Langevin barostat using τ = 50 ps. The temperature was controlled using the Langevin thermostat using τ = 0.1 ps. Electrostatic interactions were treated using u-series with a cutoff at 9 Å(Shaw et al., 2014). No cutoff was used for van der Waals interactions. The grand canonical Monte Carlo water sampling protocol was used to sample water molecules that are buried within binding sites, as described previously (Ross et al., 2023). Replica exchanges between neighboring λ windows were attempted every 1.2 ps.

Application of pKa and binding mode group corrections

Common nucleic acid binders possess pyridine, pyrazole, imidazole, amines, or guanidinium moieties that exhibit significant populations of various protonated and tautomeric states (Davis et al., 2004; Giorgio and Duca, 2019; Kovachka et al., 2024) resulting in significant modulation of binding affinity. Moreover, in modeling nucleic acids, variations in salt concentration due to experimental conditions can impact the protonation states of the ligands (Kovachka et al., 2024). With this in mind, we explicitly considered all possible protonated/deprotonated and tautomeric states, determined using LigPrep and Epik, in our FEP simulations. We then applied a previously published protocol for ΔG correction based on ligand state populations, accessible in the FEP+ GUI (de Oliveira et al., 2019). Additionally, as described below, due to the narrow binding pocket of the DNA minor groove, all poses suggested by docking series DB75 and DB921 compounds were evaluated with FEP . We applied the binding pose correction workflow for these cases using a pre-existing protocol (Ross et al., 2023).

Metrics used for the analysis

We assess FEP accuracy by comparing the mean unsigned error (MUE) and root-mean-square error (RMSE) between the calculated and experimental $\Delta\Delta G$ for each ligand pair in the FEP perturbation graph. Pairwise RMSE and MUE were calculated for each FEP map in the benchmark and reported in Table 1. The $\Delta\Delta G$ values were converted into absolute binding free energies ΔGs up to an unknown constant as described in the supporting information of Wang et al., 2013. These ΔG values were then used to determine the Pearson correlation coefficient R^2 through linear-least squares regression against experimental ΔG values as well as Kendall's τ. The weighted average of the RMSEs and of R^2, where the weights are equal to the number of compounds in each map, was computed for the collection of maps in the dataset (Figure 1b).

RESULTS & DISCUSSION

Brief description of systems studied

The dataset for this study was assembled from targets and ligands of pharmaceutical interest and consisted of five RNA/ligand complexes, two DNA/ligand systems and one ligand-bound DNA/protein(HIV Integrase) system as given in Table 1. To further evaluate the robustness of the computational predictions described in the methods section, we ensured that the dataset of eight systems contained a variety of secondary and tertiary structural forms of DNA and RNA. The DNA systems included two duplexes, while the RNA systems comprised three riboswitches (lysine, guanine, and flavin) as well as two viral RNA targets (HIV-1 trans-activation response [TAR] RNA, and hepatitis C virus [HCV] internal ribosomal entry site). The DNA/protein(HIV Integrase) complex mentioned above was notable due to the fact that it involved small molecule ligand binding at DNA/Protein interface and thus offered a unique context to test the performance of FEP+. Six of the eight systems had structural data derived from x-ray crystallography with resolutions ranging between 1.64 Å and 3 Å. The remaining systems were structurally characterized by NMR and Cryo-EM. All structures of the individual ligands and targets used in the study are detailed in the Supporting Information section (Figure S11). Also provided, in Table 1, are the literature references from which binding affinity data for each ligand were obtained. The experimental binding constants in these references varied from low millimolar to low nanomolar thus covering a suitable range of low and high affinity binders.

Overall relative binding affinity prediction performance

The summary performance statistics for all 270 perturbations on 103 ligands, as reported in Table 1 and Figure 1, indicate that the level of accuracy achieved by FEP+ on DNA and RNA systems compares favorably with recently reported results for protein-ligand use cases (Ross et al., 2023).

A strong correlation was observed between the predicted binding affinities and experimentally measured values with an R^2 of 0.61 and Kendall τ of 0.62 (Figure S12). Importantly, when distinguishing between DNA and RNA systems, the R^2 values are 0.57 and 0.64 , for DNA and RNA respectively; indicating similar performance for both classes of nucleic acids. It should be noted that here and in the following statistics we include the DNA/protein integrase complex as part of the DNA analysis for simplicity.

	DNA		**DNA/protein**	**RNA**				
Series Number. of compounds Number of perturbations	Db921 (Liu et al., 2011) 19 90	Db75 (Wilson et al., 2008) 11 42	Insist (Zhao et al., 2016) 6 7	rbt (Davis et al., 2004) 14 41	Guanine (Kim et al., 2009) 16 33	Lysine (Garst et al., 2012) 10 14	Flavins (Vicens et al., 2018) 12 23	Benzimidazole (Dibrov et al., 2012) 15 20
Construct	Duplex		HIV integrase/ DNA	Hairpin HIV-1 Tar	Riboswitch			IRES element
PDB method, resolution	2B0K X-Ray, 1.64 Å	1VZK X-Ray, 1.8 Å	6PUZ EM, 2.8 Å	1UUD NMR	1Y27 X-Ray, 2.4 Å	4ERL X-Ray, 3.0 Å	6DN1-3 X-Ray, 2.8 Å	3TZR X-Ray, 2.2 Å
pka correction	✔			✔	✔		✔	
binding mode correction	✔	✔						
Exp binding affinity range lowest/highest	-11.8/-8.5	-10.6/-6.8	-12.6/-10.5	-7.9/-5.9	-12.7/-7.5	-5.8/-2.9	-11.3/-7.2	-8.4/-5.0

R2	0.47	0.80	0.45	0.76	0.57	0.48	0.46	0.86
Kendall τ	0.65	0.75	0.55	0.58	0.69	0.45	0.28	0.83
RMSE/MUE pairwise	1.81/1.46	1.58/1.24	1.56/1.35	0.69/0.56	1.38/1.10	1.30/1.05	1.33/1.11	1.14/0.96
RMSE/MUE edgewise	1.90/1.52	1.21/0.90	1.31/1.05	0.76/0.64	1.42/1.07	1.11/1.02	1.37/1.15	1.04/0.73
RMSE/MUE Pairwise before correction	4.03/3.08	2.58/2.06	N/A	1.42/1.15	1.90/1.46	N/A	2.00/1.55	N/A
RMSE/MUE edgewise before correction	3.97/3.01	2.75/2.08	N/A	1.36/1.16	1.79/1.31	N/A	1.77/1.37	N/A

Table 1. Summary of the eight types of systems used in this study and the results of the relative binding FEP+ predictions performed on them. Eight different nucleic acid receptors, including one DNA/protein complex, with a diverse range of conformations were studied. The nucleic acid types, ligand series names and the original publication where the data sets were obtained from, number of ligands and perturbations for each system as well the experimental and predicted affinity data and their statistics are shown. The experimental binding affinity values and the RMSE/MUE for the FEP+ predictions are in the units of kcal/mol.

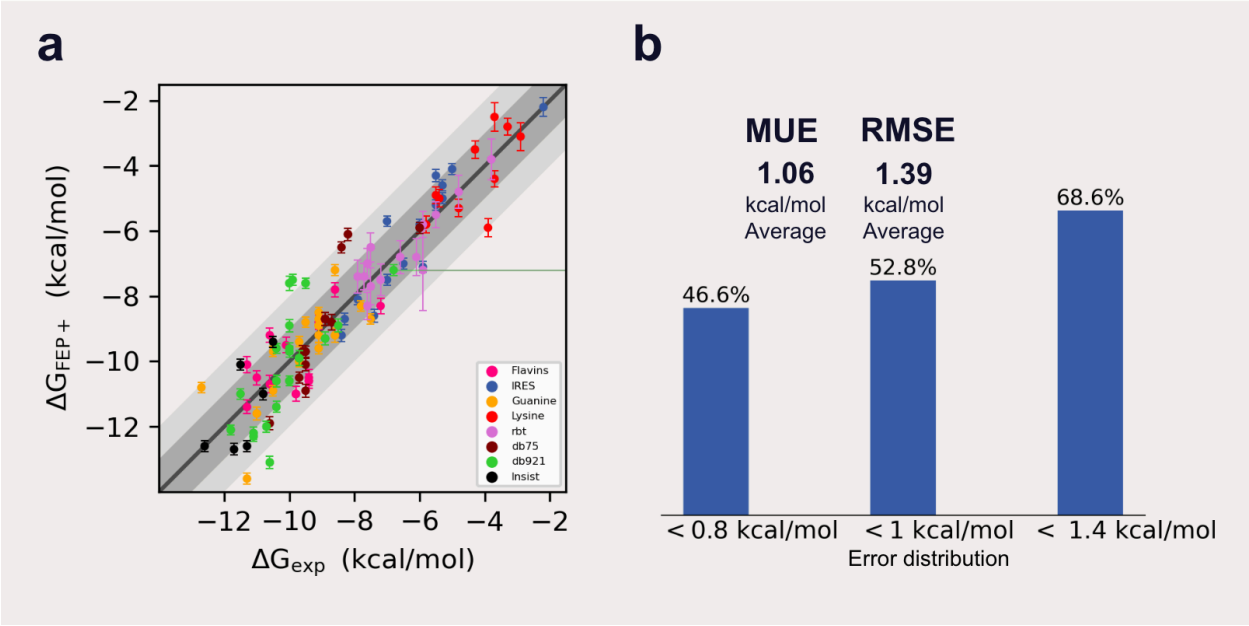

Figure 1. (**a**) A correlation plot between the FEP+ predicted binding affinities and experimentally determined affinities for all the 8 systems studied. The dark gray area represents a 1 kcal/mol error between the predicted and experimental data, whereas the light gray area represents a 2 kcal/mol error. Only four out of 103 studied ligands fall outside the 2 kcal/mol area. (**b**) The error distribution for all 270 perturbations shows that approximately 70% of FEP+ relative binding free energy predictions are within 1.4 kcal/mol of the experimental values, which is equivalent to 1 log unit. More than half of the predictions achieved remarkably low error, below 1 kcal/mol.

Overall, the average pairwise RMSE for all of our systems before applying pKa and/or binding mode correction was 2.48 kcal/mol, which improved to 1.39 kcal/mol after applying the groups feature in FEP+ (Ross et al., 2023), underscoring the importance of considering multiple states for compounds in FEP simulations of nucleic acids. Upon decomposing this metric into DNA and RNA system contributions, the RMSE values for DNA was 1.63 kcal/mol, which was noticeably higher than the 1.16 kcal/mol realized for RNA. This difference could be a reflection of variability in experimental accuracy. It could also be due to the fact that the RNA binding sites are generally more distinctly defined structurally than the DNA ones. An analysis of the error distribution revealed that a large majority, approximately 69%, of the predicted relative binding affinities were below 1.4 kcal/mol of the experimental error (Figure S13), which corresponds to less than one logarithmic unit of the dissociation constant (Kd). Only 19.7% differed by more than 2 kcal/mol. Below, we describe some common features observed for all of the series as well as representative examples of various types of nucleic acid binders, categorized by the nature of binding, e.g. minor groove, riboswitch, viral RNA binders, etc.

DNA Minor groove binders

Helical forms of RNA and DNA consist of a major and minor groove each of which possesses peculiar chemical and geometrical characteristics. For instance, the major groove of the most common form of helical RNA, A-RNA, is deep and narrow while the minor groove is shallow and wide. In comparison, the major groove of the most common form of helical DNA, B-DNA, is wide and shallow and the minor groove narrow and deep (Saenger, 1984; Shakked and Rabinovich, 1986). The type of polar and non-polar elements presented at the base edges facing the minor and major groove determine the binding characteristics of either groove.

Representative examples of DNA minor groove binders in this study are found in the DB921 (Liu et al., 2011) and DB75 (Wilson et al., 2008) series which bind to the minor groove of the AT-rich

sites in parasite mitochondrial kinetoplast DNA, destroying kinetoplast and causing cell death. In both series, the challenge of simulating these types of binders is due to the narrow nature of the minor groove, as stated above. Given the planar types of the ligands in these series, the relatively short FEP simulation time frame is insufficient to thoroughly sample the alternative binding modes of the compounds including anticipated ring flips and various orientations of the terminal charged amidine groups. Hence, we concluded that for this series, starting the FEP runs with all of the binding poses as well as ring orientations generated through Glide docking was necessary. Fortunately, FEP+ has a built-in post-processing protocol for correcting the predicted ΔGs based on the most dominant binding pose/orientation (Ross et al., 2023). A specific example of the application of this procedure was compound DB829 (Figure 2). For this compound, the individual predicted ΔGs associated with poses exhibiting various nitrogen substituted phenyl rings orientations did not show good correlation with the experimental ΔG. The four possible orientations were off by a magnitude of between 1.4kcal/mol and 4.6 kcal/mol relative to experiment. However, after the FEP+ binding pose correction was applied, the predicted ΔG of -8.9 kcal/mol matches very closely with the experimental ΔG of -8.7 kcal/mol.

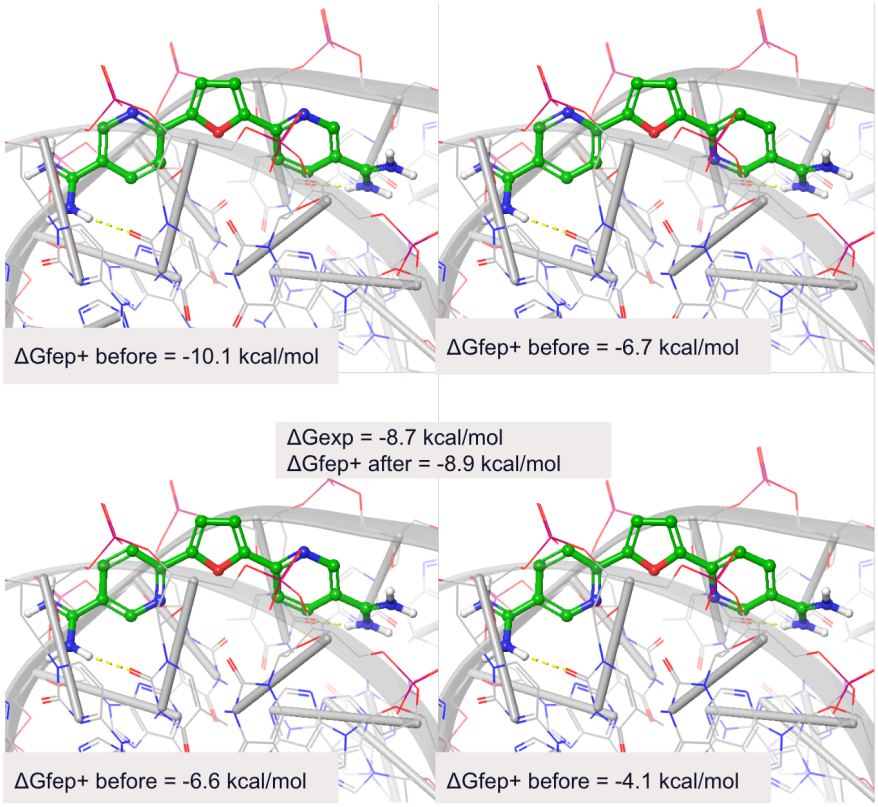

Figure 2. Four poses of compound DB829 representing the possible orientations of the nitrogen-substituted phenyls within the DNA minor groove. Whereas ΔGs for individual poses do not

correlate well with the experimental ΔG, the predicted ΔG, after FEP+ binding mode corrections, has an excellent matching with the experimental ΔG.

Another minor groove binding compound from the DB75 series, DB244 (Fig 3, left) possesses two hydrophobic cyclopentyl groups on the termini of the parent scaffold for this series, DB75 (Fig 3, right). These groups make favorable interactions with the floor of the minor groove while also displacing additional waters from the groove relative to DB75. Furthermore, as DB244 moves from an aqueous environment into the less polar DNA minor groove, the compound shows a more favorable solvation ΔG compared to DB75, as is correctly captured by FEP+. The FEP+ predicted solvation ΔΔG of ~15 kcal/mol between DB244 and DB75 is remarkably high, further indicating that the dehydration of DB244 is a driving factor in its favorable binding affinity.

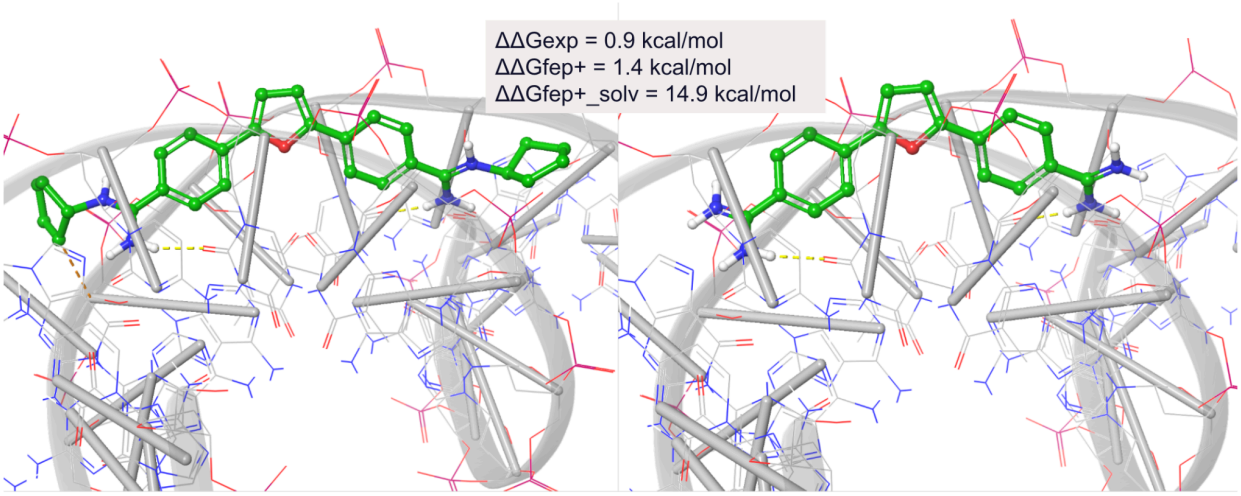

Figure 3. An example of a perturbation between a particularly hydrophobic compound in the DNA minor groove (DB244) and a less hydrophobic compound, DB75. FEP+ correctly predicts stronger binding affinity for the more hydrophobic compound.

A peculiarity of the DB921 series is that these compounds are more linear and less complementary to the contours of the DNA minor groove. Their affinities are thought to be driven by the water bridge interactions, specifically when compared to more curved ligands. The ability of dynamic water molecules to facilitate noncovalent interactions by serving as flexible donors or acceptors of hydrogen bonds, thereby stabilizing complexes between ligands and macromolecules is a phenomenon that has been noted in both protein and nucleic acids in complex with small molecules (Athri and Wilson, 2009; Cooper, 2005). An example of such binding can be found in the perturbation between the linear DB921 (Fig 4, left) and a more

curved DB911 (Fig 4, right). Despite the shape of DB911 being more DNA minor-groove complementary, DB921 has a more favorable ΔG due to its favorable water-mediated interactions. FEP+ is able to correctly capture these types of perturbations among other things due to incorporation of the grand canonical Monte Carlo (GCMC) method, which rigorously samples water molecules in the binding site and the bulk solvent and estimates their preferred positions (Ross et al., 2020).

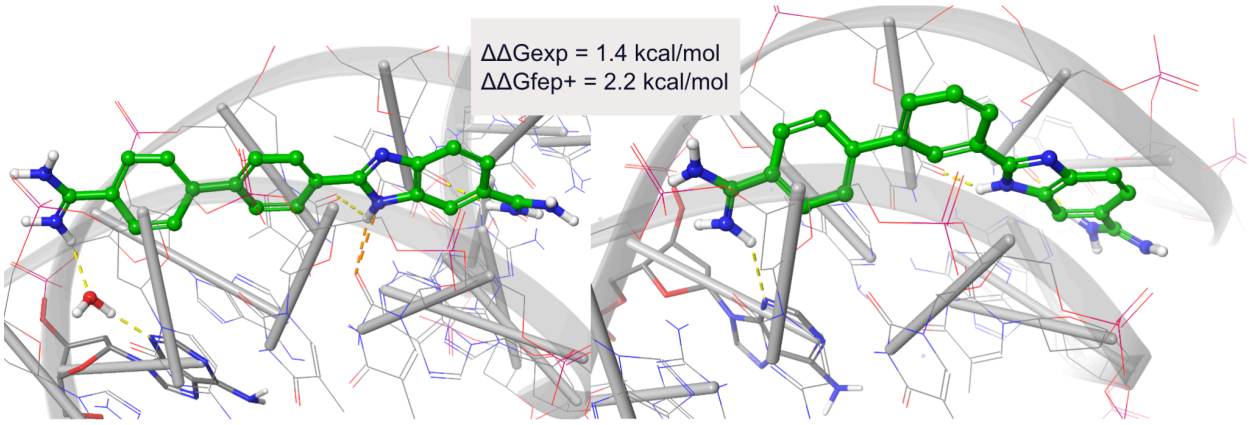

Figure 4. An example of a perturbation between a linear compound via a water bridge interaction with the DNA minor groove (left) and a curved compound (right) in the DB921 series.

Major groove binders

A representative example of viral hairpin RNA major groove binders is shown in Fig 5.

The rbt series target the trans-activation response element (TAR RNA) region of the HIV-1 and are involved in inhibition of the proliferation of the virus. The series contain elongated flexible side chains with charged guanidine ends and bind to a shallow binding pocket of the TAR RNA. The length of the side chains and the presence/absence of the guanidine groups influence the affinity. For instance, compound 541 (Fig 5, left) has long side chains with guanidine moieties on both ends, which are able to penetrate and make cation-π interactions with the RNA bases, as well as hydrogen bond and salt bridge interactions with the RNA bases/backbone atoms. Shortening the side chains and substituting a guanidine group with a smaller charged amine, results in lower affinity as can be seen in the case of the compound 393 (Fig 5, right), which is not able to make similar interactions as 541. This 2.7 kcal/mol decrease in affinity is correctly predicted by FEP+ (2.6 kcal/mol).

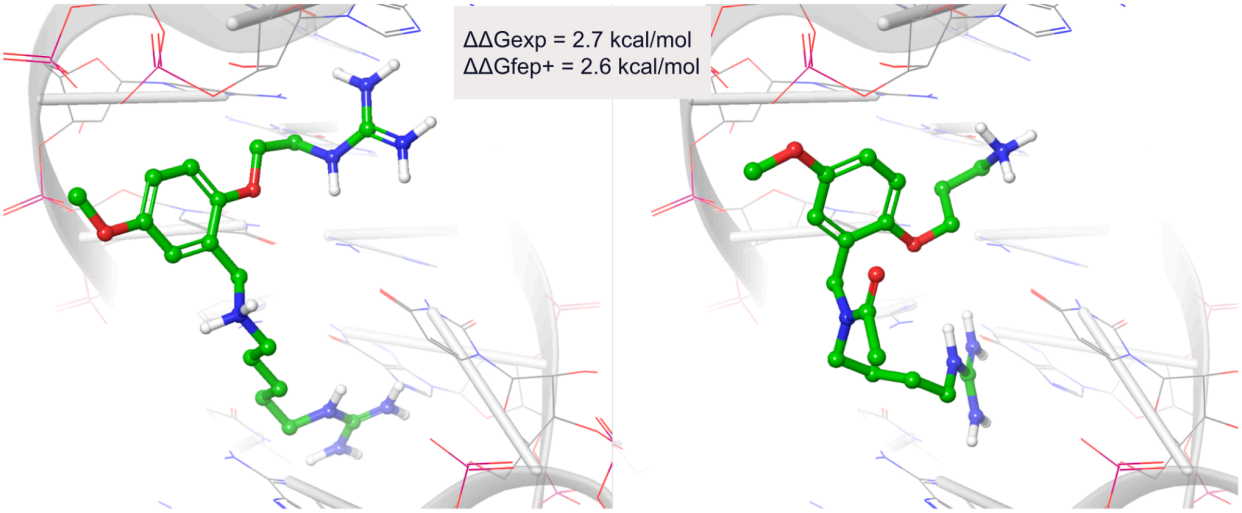

Figure 5. An example of a perturbation between a compound with longer guanidinium side chains (left) and a compound with shorter side chains, where one of the guanidinium is replaced by an amine. These compounds bind the major groove of the HIV-1 TAR domain.

Binding to riboswitches with more well-defined binding pockets

Next, we examined several small molecule compounds binding to the riboswitches. It is known that riboswitches possess a well-conserved aptamer domain that is very specific for its cognate ligand (Kim et al., 2009; Mandal et al., 2003). The RNA completely engulfs the ligands and makes interactions with virtually every polar atom on the ligand. This way the aptamer discriminates against the non native ligands. Hence, designing compounds that can compete against the native ligands is challenging. Nevertheless, several series of compounds have been designed and we have studied here the series that bind to the guanine, lysine and flavin riboswitches. For an FEP method to be successful a thorough sampling of the ligands both within the complex and solvent legs are necessary. Our results using FEP+ demonstrate reasonable RMSE for these series, indicating FEP+ is a powerful method for a rational structure based drug design at the riboswitch binding sites.

One such tight riboswitch binding-site drug design example can be found in the flavin mononucleotide (FMN) riboswitch series. Here, we see that the lipophilic interactions drive the specificity and potency, which has been observed for the compounds that bind this riboswitch (Howe et al., 2016; Kovachka et al., 2024). In this series the compounds are inserted between

two adenine bases (A48 and A85) and form π-stacking interactions with those residues as well as hydrogen bond contacts with A99 and with G10, G11 and G62. Whereas the hydrogen bond contacts are present in both the compounds with weaker and stronger affinities, the hydrophobic groups seem to be driving the potency of the higher affinity compounds. An example is a perturbation between the compound 6DN1 (Fig 6. left) and BRX897 (Fig 6. right), wherein the cyclohexane moiety in 6DN1 is inserted into a hydrophobic cleft in the FMN riboswitch formed by residues G62, A48, U61. This moiety likely drives both the affinity and specificity of 6DN1, overall resulting in a ~2.4 kcal/mol gain in potency, which is correctly predicted by FEP+ to be 2.8 kcal/mol.

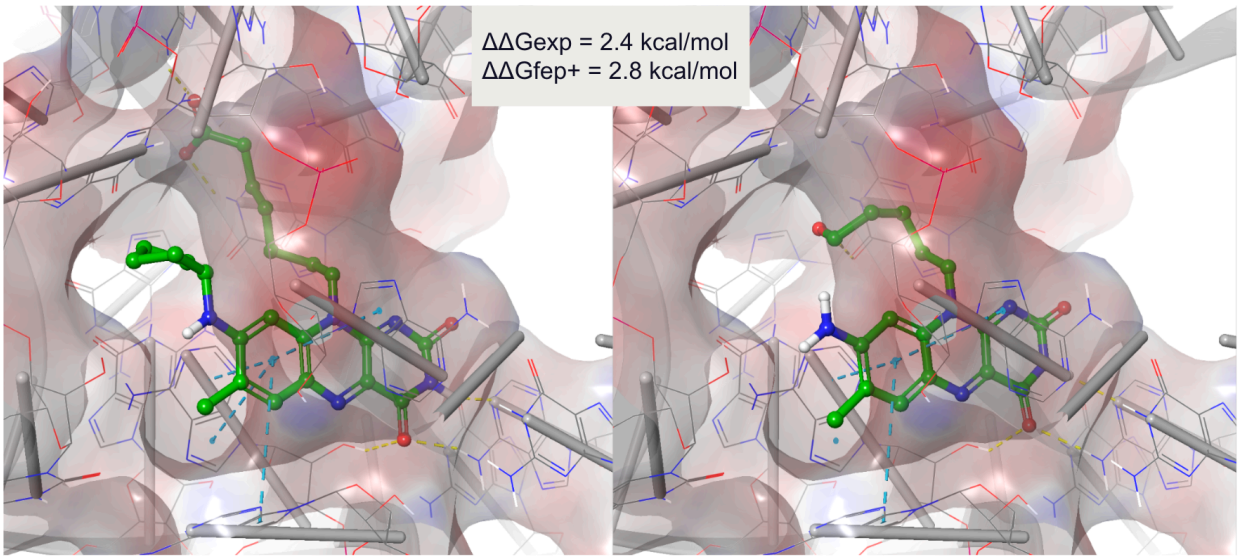

Figure 6. An example of a tight riboswitch binding site where a lipophilic moiety of the ligand on the left contributes to its specificity and potency.

In another riboswitch example, an addition of a fragment to the native ligand displaces an ion-solvent coordination within the binding site. Specifically, when the native lysine ligand (Fig 7, left) is mutated into a lysine-glycine dipeptide (Fig 7, right), the glycine moiety displaces Na+ coordinated water network, resulting in a decrease in affinity, correctly predicted by FEP+. The ion-coordinated water-mediated contacts are important for both the structure of the RNA binding site and the binding of the ligands (Serganov et al., 2009, 2008, 2006; Thore et al., 2006). The compounds in this series with similar substituent on the carboxylic site of lysine, exhibit lower affinities compared to lysine.

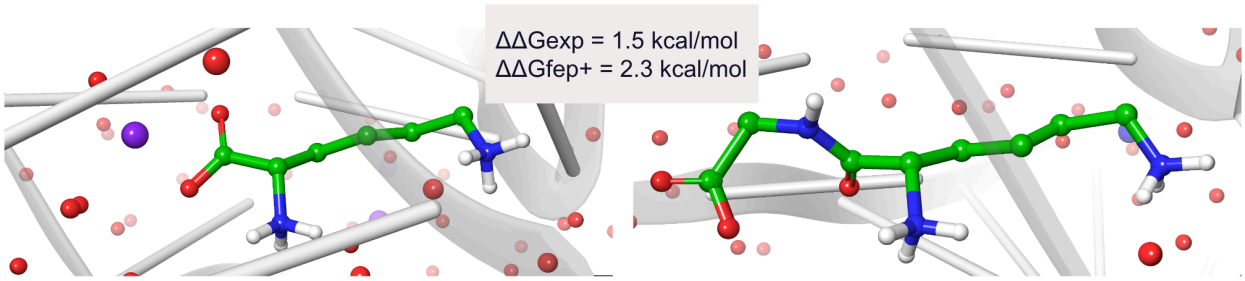

Figure 7. An example of a displacement of a sodium-water coordination site within the lysine riboswitch, leading to a decrease in affinity. Water oxygen atoms are shown in red spheres, whereas Na+ ions are in purple spheres.

Protein/DNA interface binding

A representative example of a protein/DNA interface binding can be found in the integrase complex studied here, where a series of integrase strand transfer inhibitors bind in between the DNA and the HIV integrase. In addition, this series of ligands are coordinated by Mg2+ atoms which bind the ligand and the protein residues. The perturbations in this series are on the methylene chains of varying lengths that possess the hydroxyl group. The compounds with longer methylene chains, such as compound 4d (Fig 8, left), can adopt conformations in which they interact both directly and through water-mediated interactions with three protein residues within a loop domain. Meanwhile compounds with shorter methylene chains, such 4b (Fig 8, right), are not able to fully penetrate the pocket formed by the loops of the protein and as a consequence interact with only one protein residue. This results in a lower affinity of 4b relative to 4d, as correctly captured by FEP+.

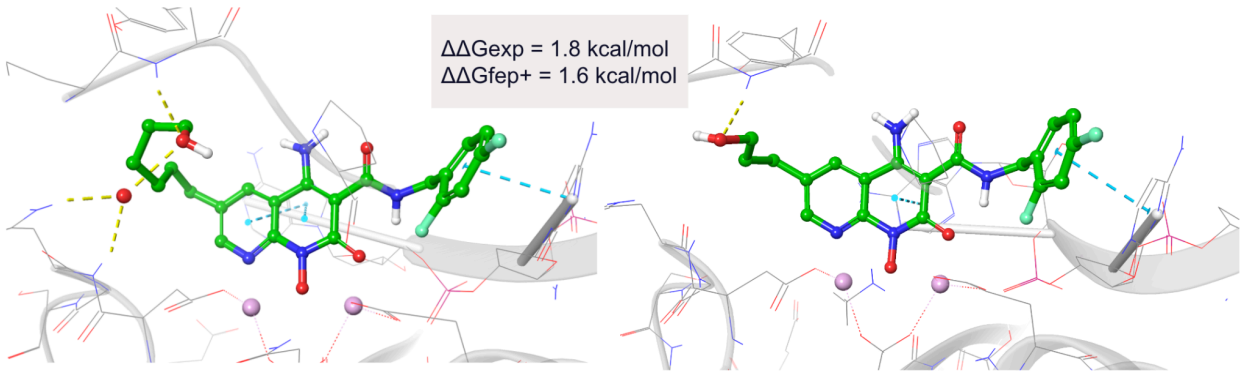

Figure 8. An example of ligand binding to the interface between a DNA and a protein (HIV integrase) coordinated by Mg2+ ions.

CONCLUSIONS

The design and development of medicines that target RNA or DNA lags behind similar efforts targeting proteins. Computational approaches can accelerate advances in this area by speeding up the identification and optimization of candidate molecules to improve properties that elicit on-target effectiveness without causing unintended off-target effects. This study demonstrates that with the improved parameterization of the OPLS4 force field, FEP+ is an accurate predictor of relative ligand binding affinities when applied to nucleic acid targets. Good performance was sustained across a diverse set of target classes as well as ligand chemistries, indicating that the method is robust enough to serve as a powerful practical tool in nucleic acid target lead optimization campaigns. One important difference between small molecules binding to proteins compared to nucleic acids is the nature of the binding site. Whereas in proteins most of the binding sites have pockets deep enough to engulf the ligand, in nucleic acids the sites can vary widely from deep minor grooves to shallow major grooves to nucleobase π-stacking in intercalation complexes and to tridimensional protein-like binding pockets found in riboswitches. Given this diversity of nucleic acid binding sites, as well that of the integrase complex where the ligand binds in between a DNA and protein molecule, the good overall performance achieved by FEP+ shows rigorousness in the breadth of binding sites and interaction types for which it can reliably predict binding affinities. The following factors were identified as important in obtaining reliable FEP+ predictions: 1) adding explicit salt in FEP+ is necessary to keep the nucleic acid structures stable throughout the FEP simulations; 2) carefully examining all states of ligands, including possible protomer and tautomeric states at a given pH, as well as various binding modes especially in narrow binding sites of minor grooves, where extensive sampling maybe difficult and later apply the pKa/binding mode correction via the Groups in the FEP+ panel; 3) use of OPLS4 for parameterizations of all ligand states; 4) running sufficiently long simulations - in our study we used 20ns runs, which is more than the default 5ns in FEP+. In future studies, it will be desirable to systematically investigate the minimum simulation length needed to give comparable results.

In addition, it is worth mentioning that the experimental affinities throughout the studied datasets were from diverse sets of experiments. For example, the IC50s for the integrase system measure the strand transfer when the integrase unzips the DNA; for the DB series binding to the minor groove of the DNA, the SPR experiments are measured on the hairpin variant of the DNA sequence attached to a biotin molecule; for the HIV-1 TAR RNA inhibitors, the Ki values were obtained via fluorescence resonance energy transfer (FRET) based assay; for the lysine

riboswitch, the Kds were measured by the 2AP fluorescence assay, whereas for the guanine and flavine riboswitch, in-line probing assays were performed to obtain the Kd values, and finally for the benzimidazole ligands binding to the hepatitis C virus RNA domain, Kds were determined via mass spectrometry assay. In addition, as a result of variability in the experimental accuracy of various assays used for measuring relative binding affinities, there is a maximal accuracy that can be expected from FEP predicted affinities (Ross et al., 2023). An illustration of this is found in the DB75 series, where binding affinities were obtained by SPR as well as ITC measurements. When plotting the correlation between these two experimentally obtained affinities (Fig S3a), we obtain an RMSE of 1.60 kcal/mol and a R^2 of 0.53. Our FEP+ predicted affinities correlate well with the SPR measurement and give an RMSE of 1.58 kcal/mol and an R^2 of 0.77. While for this series FEP performs well, in general these differences between experiential assays demonstrate the maximal accuracy an FEP method can achieve. Given this vast variety and differences of experimental methods, the statistics obtained from our FEP+ calculations are reasonable.

Since most of the current computational drug design tools have traditionally been first developed and optimized for protein-ligand systems, it is worthwhile to compare the results obtained here against those for protein-ligand systems. Our most recent benchmark validation for those systems found the overall pairwise RMSE accuracy of FEP+ using the same version of the program and OPLS4 force field to be 1.25 kcal/mol across diverse ligand and protein target series (Ross et al., 2023). Thus, the 1.39 kcal/mol accuracy reported here as the first attempt to study nucleic acid-ligand systems with FEP+ is encouraging. Nevertheless, future work would be necessary to further improve the current FEP+ method. Avenues for improvement include the following:

1. The current OPLS4 force field has been parameterized for nucleic acids, but there are limitations in representing the polarizability of the charged environment in the nucleic acid systems. Further validation of the predictions may be improved by running the simulations with OPLS5, which includes explicit description of polarization effects..
2. Charged systems are sensitive to the type and concentration of the ions used in simulations. While adding a physiological concentration (0.15M) of NaCl improves the results, further investigation of varying salt concentrations and the nature of the cation (K^+ vs Na^+) is needed, especially because potassium is more prevalent in bacterial cells.
3. Although significant improvements have been made in the preparation and docking of the nucleic acids systems in the latest suite, further use of these and FEP+/OPLS4 will

determine whether additional improvements are necessary to enhance the reported statistics.

4. Our future work will focus on optimizing and validating the absolute binding FEP protocol, which is crucial in virtual screening/hit discovery campaigns, as well as studying nucleic acid residue (base and/or backbone) mutations, given the broad applicability of this protocol.

The findings discussed in this work demonstrate the extended range of applicability of FEP+ and introduce a robust approach to address drug design challenges in nucleic acid-based systems, particularly in the lead optimization phase of drug-discovery programs.

ACKNOWLEDGEMENTS

We are very grateful to Matthew Repasky and Robert Abel for providing valuable feedback on the manuscript.

SUPPORTING INFORMATION

Supporting figures and details on the force field parameterization and fitting procedure for nucleic acids
SI - Figures and FF Doc

REFERENCES

Athri, P., Wilson, W.D., 2009. Molecular Dynamics of Water-Mediated Interactions of a Linear Benzimidazole−Biphenyl Diamidine with the DNA Minor Groove. J. Am. Chem. Soc. 131, 7618–7625. https://doi.org/10.1021/ja809249h

Berendsen, H.J.C., Grigera, J.R., Straatsma, T.P., 1987. The missing term in effective pair potentials. J. Phys. Chem. 91, 6269–6271. https://doi.org/10.1021/j100308a038

Berman, H.M., Lawson, C.L., Schneider, B., 2022. Developing Community Resources for Nucleic Acid Structures. Life 12, 540. https://doi.org/10.3390/life12040540

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Res. 28, 235–242. https://doi.org/10.1093/nar/28.1.235

Bottini, A., De, S.K., Wu, B., Tang, C., Varani, G., Pellecchia, M., 2015. Targeting Influenza A Virus RNA Promoter. Chem. Biol. Drug Des. 86, 663–673. https://doi.org/10.1111/cbdd.12534

Bowers, K.J., Chow, D.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., Shaw, D.E., 2006. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters, in: SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. Presented at the SC '06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, pp. 43–43. https://doi.org/10.1109/SC.2006.54

Cappel, D., Jerome, S., Hessler, G., Matter, H., 2020. Impact of Different Automated Binding Pose Generation Approaches on Relative Binding Free Energy Simulations. J. Chem. Inf. Model. 60, 1432–1444. https://doi.org/10.1021/acs.jcim.9b01118

Cech, T.R., 2012. The RNA Worlds in Context. Cold Spring Harb. Perspect. Biol. 4, a006742. https://doi.org/10.1101/cshperspect.a006742

Cech, T.R., Steitz, J.A., 2014. The noncoding RNA revolution-trashing old rules to forge new ones. Cell 157, 77–94. https://doi.org/10.1016/j.cell.2014.03.008

Cheatham III, T.E., Case, D.A., 2013. Twenty-five years of nucleic acid simulations. Biopolymers 99, 969–977. https://doi.org/10.1002/bip.22331

Chen, J., Wang, X., Pang, L., Zhang, J.Z.H., Zhu, T., 2019. Effect of mutations on binding of ligands to guanine riboswitch probed by free energy perturbation and molecular dynamics simulations. Nucleic Acids Res. 47, 6618–6631. https://doi.org/10.1093/nar/gkz499

Chen, W., Cui, D., Jerome, S.V., Michino, M., Lenselink, E.B., Huggins, D.J., Beautrait, A., Vendome, J., Abel, R., Friesner, R.A., Wang, L., 2023. Enhancing Hit Discovery in Virtual Screening through Absolute Protein–Ligand Binding Free-Energy Calculations. J. Chem. Inf. Model. https://doi.org/10.1021/acs.jcim.3c00013

Chen, W., Deng, Y., Russell, E., Wu, Y., Abel, R., Wang, L., 2018. Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. J. Chem. Theory Comput. 14, 6346–6358. https://doi.org/10.1021/acs.jctc.8b00825

Childs-Disney, J.L., Yang, X., Gibaut, Q.M.R., Tong, Y., Batey, R.T., Disney, M.D., 2022. Targeting RNA structures with small molecules. Nat. Rev. Drug Discov. 21, 736–762. https://doi.org/10.1038/s41573-022-00521-4

Clark, A.J., Negron, C., Hauser, K., Sun, M., Wang, L., Abel, R., Friesner, R.A., 2019. Relative Binding Affinity Prediction of Charge-Changing Sequence Mutations with FEP in Protein–Protein Interfaces. J. Mol. Biol. 431, 1481–1493. https://doi.org/10.1016/j.jmb.2019.02.003

Cooper, A., 2005. Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics. Biophys. Chem., BIFI 2004

International Conference Biology after the Genoma: A Physical View 115, 89–97. https://doi.org/10.1016/j.bpc.2004.12.011
Davis, B., Afshar, M., Varani, G., Murchie, A.I.H., Karn, J., Lentzen, G., Drysdale, M., Bower, J., Potter, A.J., Starkey, I.D., Swarbrick, T., Aboul-ela, F., 2004. Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic “hot spots.” J. Mol. Biol. 336, 343–356. https://doi.org/10.1016/j.jmb.2003.12.046
de Oliveira, C., Yu, H.S., Chen, W., Abel, R., Wang, L., 2019. Rigorous Free Energy Perturbation Approach to Estimating Relative Binding Affinities between Ligands with Multiple Protonation and Tautomeric States. J. Chem. Theory Comput. 15, 424–435. https://doi.org/10.1021/acs.jctc.8b00826
Dibrov, S.M., Ding, K., Brunn, N.D., Parker, M.A., Bergdahl, B.M., Wyles, D.L., Hermann, T., 2012. Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. Proc. Natl. Acad. Sci. 109, 5223–5228. https://doi.org/10.1073/pnas.1118699109
Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigó, R., Gingeras, T.R., 2012. Landscape of transcription in human cells. Nature 489, 101–108. https://doi.org/10.1038/nature11233
Duan, J., Lupyan, D., Wang, L., 2020. Improving the Accuracy of Protein Thermostability Predictions for Single Point Mutations. Biophys. J. 119, 115–127. https://doi.org/10.1016/j.bpj.2020.05.020
Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K., Shaw, D.E., Francis, P., Shenkin, P.S., 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 47, 1739–1749. https://doi.org/10.1021/jm0306430
Garst, A.D., Porter, E.B., Batey, R.T., 2012. Insights into the Regulatory Landscape of the Lysine Riboswitch. J. Mol. Biol. 423, 17–33. https://doi.org/10.1016/j.jmb.2012.06.038
Giorgio, A.D., Duca, M., 2019. Synthetic small-molecule RNA ligands: future prospects as therapeutic agents. MedChemComm 10, 1242–1255. https://doi.org/10.1039/C9MD00195F
Halgren, T.A., Murphy, R.B., Friesner, R.A., Beard, H.S., Frye, L.L., Pollard, W.T., Banks, J.L., 2004. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. J. Med. Chem. 47, 1750–1759. https://doi.org/10.1021/jm030644s
Hauser, K., Negron, C., Albanese, S.K., Ray, S., Steinbrecher, T., Abel, R., Chodera, J.D., Wang, L., 2018. Predicting resistance of clinical Abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. Commun. Biol. 1. https://doi.org/10.1038/s42003-018-0075-x
Howe, J.A., Xiao, L., Fischmann, T.O., Wang, H., Tang, H., Villafania, A., Zhang, R., Barbieri, C.M., Roemer, T., 2016. Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the E. coli FMN riboswitch. RNA Biol. 13, 946–954.

https://doi.org/10.1080/15476286.2016.1216304
Kim, J.N., Blount, K.F., Puskarz, I., Lim, J., Link, K.H., Breaker, R.R., 2009. Design and Antimicrobial Action of Purine Analogs that Bind Guanine Riboswitches. ACS Chem. Biol. 4, 915–927. https://doi.org/10.1021/cb900146k
Kovachka, S., Panosetti, M., Grimaldi, B., Azoulay, S., Di Giorgio, A., Duca, M., 2024. Small molecule approaches to targeting RNA. Nat. Rev. Chem. 8, 120–135. https://doi.org/10.1038/s41570-023-00569-9
Kuhn, M., Firth-Clark, S., Tosco, P., Mey, A.S.J.S., Mackey, M., Michel, J., 2020. Assessment of Binding Affinity via Alchemical Free-Energy Calculations. J. Chem. Inf. Model. 60, 3120–3130. https://doi.org/10.1021/acs.jcim.0c00165
Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, Christina, Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blöcker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, Christopher, Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, C. for G.R., The Sanger Centre:, Washington University Genome Sequencing Center, US DOE Joint Genome Institute:, Baylor College of Medicine Human Genome Sequencing Center:, RIKEN

Genomic Sciences Center:, Genoscope and CNRS UMR-8030:, Department of Genome Analysis, I. of M.B., GTC Sequencing Center:, Beijing Genomics Institute/Human Genome Center:, Multimegabase Sequencing Center, T.I. for S.B., Stanford Genome Technology Center:, University of Oklahoma's Advanced Center for Genome Technology:, Max Planck Institute for Molecular Genetics:, Cold Spring Harbor Laboratory, L.A.H.G.C., GBF—German Research Centre for Biotechnology:, *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):, Scientific management: National Human Genome Research Institute, U.N.I. of H., Stanford Human Genome Center:, University of Washington Genome Center:, Department of Molecular Biology, K.U.S. of M., University of Texas Southwestern Medical Center at Dallas:, Office of Science, U.D. of E., The Wellcome Trust:, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921. https://doi.org/10.1038/35057062

Lawson, C.L., Berman, H.M., Chen, L., Vallat, B., Zirbel, C.L., 2024. The Nucleic Acid Knowledgebase: a new portal for 3D structural information about nucleic acids. Nucleic Acids Res. 52, D245–D254. https://doi.org/10.1093/nar/gkad957

Liu, Y., Kumar, A., DEPAUW, S., NHILI, R., DAVID-CORDONNIER, M.-H., Lee, M.P., Ismail, M.A., Farahat, A.A., Say, M., Chackal-Catoen, S., Batista-Parra, A., Neidle, S., Boykin, D.W., Wilson, W.D., 2011. Water-Mediated Binding of Agents that Target the DNA Minor Groove. J. Am. Chem. Soc. 133, 10171–10183. https://doi.org/10.1021/ja202006u

Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., Ross, G.A., Dahlgren, M.K., Russell, E., Von Bargen, C.D., Abel, R., Friesner, R.A., Harder, E.D., 2021. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. J. Chem. Theory Comput. 17, 4291–4300. https://doi.org/10.1021/acs.jctc.1c00302

Ma, H., Jia, X., Zhang, K., Su, Z., 2022. Cryo-EM advances in RNA structure determination. Signal Transduct. Target. Ther. 7, 1–6. https://doi.org/10.1038/s41392-022-00916-0

Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., Breaker, R.R., 2003. Riboswitches Control Fundamental Biochemical Pathways in Bacillus subtilis and Other Bacteria. Cell 113, 577–586. https://doi.org/10.1016/S0092-8674(03)00391-X

Ross, G.A., Lu, C., Scarabelli, G., Albanese, S.K., Houang, E., Abel, R., Harder, E.D., Wang, L., 2023. The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. Commun. Chem. 6, 1–12. https://doi.org/10.1038/s42004-023-01019-9

Ross, G.A., Russell, E., Deng, Y., Lu, C., Harder, E.D., Abel, R., Wang, L., 2020. Enhancing Water Sampling in Free Energy Calculations with Grand Canonical Monte Carlo. J. Chem. Theory Comput. 16, 6061–6076. https://doi.org/10.1021/acs.jctc.0c00660

Saenger, W., 1984. Principles of Nucleic Acid Structure, Springer Advanced Texts in Chemistry. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-5190-3

Sastry, G.M., Adzhigirey, M., Day, T., Annabhimoju, R., Sherman, W., 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. J. Comput. Aided Mol. Des. 27, 221–234. https://doi.org/10.1007/s10822-013-9644-8

Scarabelli, G., Oloo, E.O., Maier, J.K.X., Rodriguez-Granillo, A., 2022. Accurate Prediction of Protein Thermodynamic Stability Changes upon Residue Mutation using Free Energy Perturbation. J. Mol. Biol. 434, 167375. https://doi.org/10.1016/j.jmb.2021.167375

Schneider, B., Sweeney, B.A., Bateman, A., Cerny, J., Zok, T., Szachniuk, M., 2023. When will RNA get its AlphaFold moment? Nucleic Acids Res. gkad726. https://doi.org/10.1093/nar/gkad726

Schrödinger Release 202–3: Ligprep, 2022.

Serganov, A., Huang, L., Patel, D.J., 2009. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. Nature 458, 233–237. https://doi.org/10.1038/nature07642

Serganov, A., Huang, L., Patel, D.J., 2008. Structural insights into amino acid binding and gene

control by a lysine riboswitch. Nature 455, 1263–1267. https://doi.org/10.1038/nature07326
Serganov, A., Polonskaia, A., Phan, A.T., Breaker, R.R., Patel, D.J., 2006. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nature 441, 1167–1171. https://doi.org/10.1038/nature04740
Shakked, Z., Rabinovich, D., 1986. The effect of the base sequence on the fine structure of the DNA double helix. Prog. Biophys. Mol. Biol. 47, 159–195. https://doi.org/10.1016/0079-6107(86)90013-1
Sharp, K.A., Honig, B., 1995. Salt effects on nucleic acids. Curr. Opin. Struct. Biol. 5, 323–328. https://doi.org/10.1016/0959-440X(95)80093-X
Shaw, D.E., Grossman, J.P., Bank, J.A., Batson, B., Butts, J.A., Chao, J.C., Deneroff, M.M., Dror, R.O., Even, A., Fenton, C.H., Forte, A., Gagliardo, J., Gill, G., Greskamp, B., Ho, C.R., Ierardi, D.J., Iserovich, L., Kuskin, J.S., Larson, R.H., Layman, T., Lee, L.-S., Lerer, A.K., Li, C., Killebrew, D., Mackenzie, K.M., Mok, S.Y.-H., Moraes, M.A., Mueller, R., Nociolo, L.J., Peticolas, J.L., Quan, T., Ramot, D., Salmon, J.K., Scarpazza, D.P., Schafer, U.B., Siddique, N., Snyder, C.W., Spengler, J., Tang, P.T.P., Theobald, M., Toma, H., Towles, B., Vitale, B., Wang, S.C., Young, C., 2014. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer, in: SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. Presented at the SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 41–53. https://doi.org/10.1109/SC.2014.9
Shelley, J.C., Cholleti, A., Frye, L.L., Greenwood, J.R., Timlin, M.R., Uchimaya, M., 2007. Epik: a software program for pKaprediction and protonation state generation for drug-like molecules. J. Comput. Aided Mol. Des. 21, 681–691. https://doi.org/10.1007/s10822-007-9133-z
Song, L.F., Lee, T.-S., Zhu, C., York, D.M., Merz, K.M.Jr., 2019. Using AMBER18 for Relative Free Energy Calculations. J. Chem. Inf. Model. 59, 3128–3135. https://doi.org/10.1021/acs.jcim.9b00105
Tan, Z.-J., Chen, S.-J., 2011. Salt Contribution to RNA Tertiary Structure Folding Stability. Biophys. J. 101, 176–187. https://doi.org/10.1016/j.bpj.2011.05.050
Thore, S., Leibundgut, M., Ban, N., 2006. Structure of the Eukaryotic Thiamine Pyrophosphate Riboswitch with Its Regulatory Ligand. Science 312, 1208–1211. https://doi.org/10.1126/science.1128451
Vicens, Q., Mondragón, E., Reyes, F.E., Coish, P., Aristoff, P., Berman, J., Kaur, H., Kells, K.W., Wickens, P., Wilson, J., Gadwood, R.C., Schostarez, H.J., Suto, R.K., Blount, K.F., Batey, R.T., 2018. Structure–Activity Relationship of Flavin Analogues That Target the Flavin Mononucleotide Riboswitch. ACS Chem. Biol. 13, 2908–2919. https://doi.org/10.1021/acschembio.8b00533
Wang, L., Chambers, J., Abel, R., 2019. Protein–Ligand Binding Free Energy Calculations with FEP+, in: Bonomi, M., Camilloni, C. (Eds.), Biomolecular Simulations: Methods and Protocols. Springer, New York, NY, pp. 201–232. https://doi.org/10.1007/978-1-4939-9608-7_9
Wang, L., Deng, Y., Knight, J.L., Wu, Y., Kim, B., Sherman, W., Shelley, J.C., Lin, T., Abel, R., 2013. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative Binding Affinity Predictions of CDK2 Inhibitors. J. Chem. Theory Comput. 9, 1282–1293. https://doi.org/10.1021/ct300911a
Wang, L., Deng, Y., Wu, Y., Kim, B., LeBard, D.N., Wandschneider, D., Beachy, M., Friesner, R.A., Abel, R., 2017. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. J. Chem. Theory Comput. 13, 42–54. https://doi.org/10.1021/acs.jctc.6b00991

Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., Lupyan, D., Robinson, S., Dahlgren, M.K., Greenwood, J., Romero, D.L., Masse, C., Knight, J.L., Steinbrecher, T., Beuming, T., Damm, W., Harder, E., Sherman, W., Brewer, M., Wester, R., Murcko, M., Frye, L., Farid, R., Lin, T., Mobley, D.L., Jorgensen, W.L., Berne, B.J., Friesner, R.A., Abel, R., 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. https://doi.org/10.1021/ja512751q

Westhof, E., Leontis, N.B., 2021. An RNA-centric historical narrative around the Protein Data Bank. J. Biol. Chem. 296. https://doi.org/10.1016/j.jbc.2021.100555

Wilson, W.D., Tanious, F.A., Mathis, A., Tevis, D., Hall, J.E., Boykin, D.W., 2008. Antiparasitic compounds that target DNA. Biochimie, Targeting DNA Part 1 90, 999–1014. https://doi.org/10.1016/j.biochi.2008.02.017

Yu, A.-M., Choi, Y.H., Tu, M.-J., 2020. RNA Drugs and RNA Targets for Small Molecules: Principles, Progress, and Challenges. Pharmacol. Rev. 72, 862–898. https://doi.org/10.1124/pr.120.019554

Yu, H.S., Deng, Y., Wu, Y., Sindhikara, D., Rask, A.R., Kimura, T., Abel, R., Wang, L., 2017. Accurate and Reliable Prediction of the Binding Affinities of Macrocycles to Their Protein Targets. J. Chem. Theory Comput. 13, 6290–6300. https://doi.org/10.1021/acs.jctc.7b00885

Yu, H.S., Gao, C., Lupyan, D., Wu, Y., Kimura, T., Wu, C., Jacobson, L., Harder, E., Abel, R., Wang, L., 2019. Toward Atomistic Modeling of Irreversible Covalent Inhibitor Binding Kinetics. J. Chem. Inf. Model. 59, 3955–3967. https://doi.org/10.1021/acs.jcim.9b00268

Yu, H.S., Modugula, K., Ichihara, O., Kramschuster, K., Keng, S., Abel, R., Wang, L., 2021. General Theory of Fragment Linking in Molecular Design: Why Fragment Linking Rarely Succeeds and How to Improve Outcomes. J. Chem. Theory Comput. 17, 450–462. https://doi.org/10.1021/acs.jctc.0c01004

Zhang, J., Fei, Y., Sun, L., Zhang, Q.C., 2022. Advances and opportunities in RNA structure experimental determination and computational modeling. Nat. Methods 19, 1193–1207. https://doi.org/10.1038/s41592-022-01623-y

Zhao, X.Z., Smith, S.J., Maskell, D.P., Metifiot, M., Pye, V.E., Fesen, K., Marchand, C., Pommier, Y., Cherepanov, P., Hughes, S.H., Burke, T.R.Jr., 2016. HIV-1 Integrase Strand Transfer Inhibitors with Reduced Susceptibility to Drug Resistant Mutant Integrases. ACS Chem. Biol. 11, 1074–1081. https://doi.org/10.1021/acschembio.5b00948

Zhu, F., Bourguet, F.A., Bennett, W.F.D., Lau, E.Y., Arrildt, K.T., Segelke, B.W., Zemla, A.T., Desautels, T.A., Faissol, D.M., 2022. Large-scale application of free energy perturbation calculations for antibody design. Sci. Rep. 12, 12489. https://doi.org/10.1038/s41598-022-14443-z

Zwanzig, R.W., 1954. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J. Chem. Phys. 22, 1420–1426. https://doi.org/10.1063/1.1740409