Title
Predicting Binding Free Energies in a Large Combinatorial Chemical Space Using Multisite λ Dynamics

Journal
The Journal of Physical Chemistry Letters, 9(12)

ISSN
1948-7185

Authors
Vilseck, Jonah Z
Armacost, Kira A
Hayes, Ryan L
et al.

Publication Date
2018-06-21

DOI
10.1021/acs.jpclett.8b01284

Peer reviewed

Published in final edited form as:

J Phys Chem Lett. 2018 June 21; 9(12): 3328–3332. doi:10.1021/acs.jpclett.8b01284.

Predicting Binding Free Energies in a Large Combinatorial Chemical Space Using Multisite λ Dynamics

Jonah Z. Vilseck[1,†], **Kira A. Armacost**[1,†], **Ryan L. Hayes**[1], **Garrett B. Goh**[1], and **Charles L. Brooks III**[1,2,*]

[1]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

[2]Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109, United States

Abstract

In this study, we demonstrate the extensive scalability of the biasing potential replica exchange multisite λ dynamics (BP-REX MSλD) free energy method by calculating binding affinities for 512 inhibitors to HIV Reverse Transcriptase (HIV-RT). This is the largest exploration of chemical space using free energy methods known to date, requires only a few simulations, and identifies 55 new inhibitor designs against HIV-RT predicted to be at least as potent as a tight binding reference compound (i.e., as potent as 56 nM). We highlight that BP-REX MSλD requires an order of magnitude less computational resources than conventional free energy methods while maintaining a similar level of precision, overcomes the inherent poor scalability of conventional free energy methods, and enables the exploration of combinatorially large chemical spaces in the context of *in silico* drug discovery.

Graphical abstract

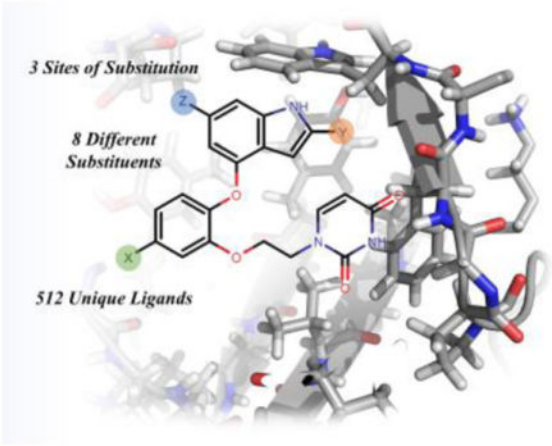
*Corresponding Author: brookscl@umich.edu. Address: Department of Chemistry and Biophysics, University of Michigan, 930 N. University Ave., Ann Arbor, MI 48109.

†J.Z.V. and K.A.A. contributed equally to this work.

The authors declare no competing financial interest.

Supporting Information.

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpclett.8b01284. Computational details, all 512 indole $\Delta\Delta G_{bind}$ results, and supplementary analyses (PDF)



Recent studies suggest that the development of new pharmaceutical small-molecule drugs currently requires $1–2 billion and 10+ years before drug approval and market launch.[1,2] Pre-phase I discovery alone can account for as much as 30% of this cost, with lead optimization requiring 17% and an average of 2 years of active research.[1–3] In an effort to reduce costs and accelerate drug discovery, computational chemistry tools are frequently employed in the preclinical drug development stages of hit identification and lead optimization,[4–6] with overall success.[7–10] Alchemical free energy calculations, including Free Energy Perturbation theory[11] and Thermodynamic Integration[12] (FEP/TI), have been shown to be particularly effective in lead optimization by allowing novel chemical modifications to a lead compound to be evaluated computationally prior to experimental synthesis and activity determination. Provided sufficient statistical sampling and correct representation of the energetics of a system, these calculations can yield a rigorous determination of protein-ligand binding free energy differences for localized, pairwise molecular perturbations.[13–16] Recent advances in computer hardware technology in the form of graphics processing units (GPUs) and enhanced sampling techniques have provided effective solutions to problems of speed and convergence encountered historically.[6,14–17] However, the fact remains that current alchemical free energy methodologies scale poorly with the total number of compounds modeled, i.e., linearly for pairwise perturbations. Therefore, there is an inherent scalability limitation that has yet to be addressed, impeding computational explorations of very large chemical spaces, e.g., several hundred ligands.

In this Letter, we address these scalability issues by demonstrating the effectiveness of an innovative free energy method, λ dynamics, designed to study large chemical spaces within the context of a pharmaceutics design problem.[18–22] Unlike conventional free energy methods, which calculate the relative binding affinity of pairs of compounds separately,[11–15] λ dynamics, as implemented in the CHARMM molecular simulation package,[23,24] can simultaneously calculate relative binding affinities between several substituents attached at a single site[18,19] or on multiple sites.[20–22] This is achieved by making λ, the parameter that alchemically transforms one ligand into another, a continuous variable that is propagated along with the Cartesian coordinates of the system using extended Lagrangian methods.[18] Thus, λ fluctuates dynamically throughout a molecular dynamics (MD) fsimulation based

on a ligand's interactions with its environment, analogous to an experimental competitive binding assay against a drug target.[19] In an unprecedented application of free energy methodologies, we show that many *hundreds* of inhibitors to HIV Reverse Transcriptase (HIV-RT)[7,25,26] can be evaluated *simultaneously* within a *single* calculation using multisite λ dynamics (MSλD). MSλD requires 20–30 times less computational resources compared to FEP/TI methods, and no loss of statistical precision is observed.

MSλD simulations were performed with the biasing potential replica exchange algorithm (BP-REX MSλD)[21] on the indole scaffold of a catechol diether inhibitor of HIV-RT reported by Lee *et al.* (Figure 1).[25] The availability of high-quality experimental data, including activity assay results and crystallographic structures, from the Jorgensen group make this an appealing benchmark for the present demonstration of MSλD scalability.[7,25–30] Simulations were performed as described previously,[21] with the addition of adaptive landscape flattening variable biases and a soft-core Lennard-Jones interaction potential.[22] Eight substituents at three sites (X, Y, and Z) around the inhibitor core yielded 512 possible permutations of unique inhibitor designs (Figure 1). These substituents resemble previous perturbations made by Jorgensen and co-workers[7,25,26,30] and occupy similar volumes, a metric previously identified to optimize MSλD sampling.[21] We note, however, that by utilizing a soft-core interaction potential reliable results can be obtained with ease regardless of substituent volumetric differences.[22]

Simulations were performed in duplicate with five replicas for both the unbound ligand (30 ns/replica) and the protein–ligand complex (40 ns/replica), for a total of 3.75 μs of sampling, to provide the free energy changes for the vertical arms in an alchemical thermodynamic cycle (Figure S1).[13] All ligands remained bound in the same pocket, and binding free energies were calculated relative to the reference compound X = H, Y = CN, Z = H (H/CN/H), where the experimental EC_{50} was determined to be 56 nM.[25] Within the λ dynamics framework, the relative free energy difference ($\Delta\Delta G_{\text{bind}}$) for each pair of compounds can be calculated from

$$\Delta\Delta G_{\mathrm{X}(i);\,\mathrm{Y}(j);\,\mathrm{Z}(k)\rightarrow\mathrm{X}(l);\,\mathrm{Y}(m);\,\mathrm{Z}(n)} = -k_B T \ln \frac{P(\lambda_{\mathrm{X}(l)}=1;\lambda_{\mathrm{Y}(m)}=1;\lambda_{\mathrm{Z}(n)}=1)}{P(\lambda_{\mathrm{X}(i)}=1;\lambda_{\mathrm{Y}(j)}=1;\lambda_{\mathrm{Z}(k)}=1)} \quad (1)$$

where $P(\lambda_{X(l)}=1;\lambda_{Y(m)}=1;\lambda_{Z(n)}=1)$ represents the population of the inhibitor of interest and $P(\lambda_{X(i)}=1;\lambda_{Y(j)}=1;\lambda_{Z(k)}=1)$ represents the reference inhibitor's population.[20] Alphabetic subscripts refer to individual substituents at each site (Figure 1). Inhibitor populations, $P(\lambda_X=1;\lambda_Y=1;\lambda_Z=1)$, are determined by counting the amount of time each unique ligand is sampled, reweighted across all replicas.[21] Additional computational details are presented in the Supporting Information (SI).

Relative binding affinities were calculated for all 512 inhibitors using eq 1 (Figure 2A, Table S1); statistical uncertainties (σ) were calculated as the standard error of the mean across the many duplicate simulations and were on average 0.13 kcal/mol. A cutoff of ±0.50 kcal/mol was used to select ligands with differing chemical properties or groups but that maintained a

similar binding affinity as the reference compound. As shown in Figure 2A, 55 out of 512 inhibitors were found to be more or as potent as the reference compound with $\Delta\Delta G_{bind} \leq$ 0.50 kcal/mol, and 18 were more potent by at least −0.50 kcal/mol. In comparison, in prior computationally guided studies by Jorgensen and co-workers, only a single compound, H/CN/CH_3, was reported that was experimentally more potent than the reference.[25] Of the 55 compounds identified in this work, we note that all but 1 have at least 1 fluorine (F) or chlorine (Cl) atom at any site, over half have a F or Cl atom at 2 out of 3 sites, and 5 of the top 10 have F or Cl at all three sites (Table S1). These results suggest that the addition of halogen atoms at X, Y, or Z sites is highly favorable for increasing inhibitor potency for the indole derivatized catechol diether ligand. Experimentally, this has been verified for several catechol diethers developed by Jorgensen and co-workers,[7,25–27,30] with many ligands displaying nanomolar to picomolar activities.

To evaluate MSλD convergence and verify the precision of the $\Delta\Delta G_{bind}$ results, conventional TI calculations were performed in conjunction with the multistate Bennett acceptance ratio (MBAR) method.[31] Sampling all 512 molecules would be computationally unfeasible with TI/MBAR; therefore, a subset of 34 molecules was investigated that had either experimental data available, the lowest predicted $\Delta\Delta G_{bind}$, or the highest σ (Figure S2). As shown in Figure 2B, this data set extends over a significant range of 9 kcal/mol in computed affinities and thus provides an adequate comparison set for evaluating the precision of the larger MSλD data set. A Pearson correlation of 0.968 and a mean unsigned error (MUE) of 0.52 kcal/mol are observed, suggesting that excellent agreement is achieved between the methods. It is worth reiterating that MSλD free energies were obtained from a *single* simulation of 512 ligands, whereas TI/MBAR required 49 separate pairwise perturbations to examine only 34 ligands, and that MSλD is able to accurately rank ligands from most to least active over this entire 9 kcal/mol range of comparison. As an additional check of precision, a Welch *t* test was performed to demonstrate the statistical significance of all MSλD $\Delta\Delta G_{bind}$ (Table S1). With a significance level of 0.01 (corresponding to a 99% confidence interval), 490 molecules show statistically meaningful $\Delta\Delta G_{bind}$ differences from 0.00, including all 18 ligands predicted to be more potent by at least −0.50 kcal/mol. The remaining 21 molecules had $\Delta\Delta G_{bind}$ within ±0.50 kcal/mol of the reference compound, thus supporting the claim that these molecules are as potent as the reference compound.

To assess the accuracy of our calculations, MUE and root-mean-square errors (RMSEs) were calculated for the six indole inhibitors with reported experimental EC_{50} values.[25,26] We emphasize that this analysis reflects the underlying *accuracy of the force field* employed in the calculations and not the *precision of the methodology*. CHARMM can employ any number of existing force fields with MSλD, and thus, the accuracy of the current results may be improved via force field adjustment, a task that is, however, beyond the scope of this work. In addition, two other bicycle-containing catechol diethers were examined with BP-REX MSλD, specifically, indolizine and naphthalene derivatives (Figure S3), yielding 22 additional data points for comparison.[25,26] Experimental binding free energies were estimated from available EC_{50} values (eq 2), similar to previous work from the Jorgensen group.[32] We note that EC_{50} is not a direct measurement of binding affinity and the experimental uncertainties are unknown. To minimize errors introduced by choosing an arbitrary reference molecule, relative free energies were converted to absolute values by

adding a single offset free energy such that eq 3 is satisfied.[9] Absolute free energies of binding (ΔG_{bind}) for all ligand series are reported in Table S2 and plotted in Figure 3.

$$\Delta G_{bind} = RTln(EC_{50}) \quad (2)$$

$$\sum_{i=1}^{N} \Delta G^{i}_{expt} = \sum_{i=1}^{N} \Delta G^{i}_{bind} \quad (3)$$

The MUE and RMSE for all 28 ΔG_{bind} is 0.88 and 1.11 kcal/mol, respectively, commensurate with errors from other research groups using free energy methods and different force fields for computer-aided drug design,[9,14,16,25,33] i.e., between 1.0 and 1.5 kcal/mol, including previous MSλD calculations.[20,21] The 10 compounds with fluorine at X and Z sites displayed excellent agreement with experiment, yielding a MUE and RMSE of 0.50 and 0.72 kcal/mol, respectively. Thus, the fidelity of the predicted ΔG_{bind} results with experiment lends credence to the accuracy of the larger 512 indole results, especially for the many fluorine-containing inhibitors predicted to be more potent than the H/CN/H indole reference (Table S3).

With the ability to quantitatively evaluate 512 unique ligands collectively, patterns in the preferred substitution at each site can be analyzed and new structural insights can be gained. By independently sorting binding affinities by substituent and site (Figure S4), it becomes readily apparent that Cl, F, and H substituents are favorable attachments at all three sites. In contrast, hydroxy (OH), methoxy (OCH_3), and cyano (CN) substituents were favored at predominately one site in the most potent molecules. At site X, for example, trifluoromethyl (CF_3) is disfavored due to steric clashes with Val179, which pushes the peptide chain away and weakens aryl-aryl contacts between Tyr181 and the catechol ring (Figure 4A); however, OH and OCH_3 are solvent exposed and able to make favorable hydrogen bonds to water. At site Y, CN benefits from favorable ion-dipole interactions with Lys223 (Figure 4B), with CN–Lys distances of 4.0–7.0 Å frequently observed in the trajectories, in agreement with published crystal structures.[25] Finally, reported steric sensitivity with regards to the orientation of Tyr181 at site Z (Figure 4C) described crystallographically is recapitulated in the MSλD simulations.[25,28] Specifically, Tyr181 is observed to switch between edge-to-face and edge-to-edge conformations in conjunction with transitions between Z site substituents. Additional structural insights obtained from the MSλD trajectories are discussed in greater detail in the SI.

We suggest that the ability of BP-REX MSλD to predict potent inhibitors within a large chemical space, covering several hundreds of ligands, will dramatically improve the utility and facility with which free energy calculations can be employed to guide lead optimization in a typical drug discovery campaign. To demonstrate this point further, scalability between traditional free energy methods, including FEP and TI methods, can be compared to BP-REX MSλD in the context of this study of 512 inhibitors. For example, the 49 TI calculations performed herein, featuring triplicate production runs and multiple closed

thermodynamic cycles,[34] required ~8.1 μs of total sampling, far exceeding the resources expended with MSλD yet covering 15 times less chemical space! Extrapolation from 34 to 512 molecules suggests that a minimum of 84–122 μs of sampling would be required with TI/MBAR, while MSλD accomplished this task with 30 times less sampling! A more generalized scalability analysis presented in the SI shows that MSλD-based methods display superior scalability in both the number of simulations and simulation length.

In conclusion, an exhaustive investigation of 512 possible permutations of inhibitor designs with BP-REX MSλD using only 3.75 μs of total simulation time identified 55 new inhibitor designs predicted to be more or as potent as a 56 nM reference inhibitor. Comparisons to independent TI/MBAR calculations and available experimental data have demonstrated that a high degree of precision and accuracy, within ± 0.5–1.0 kcal/mol, is achieved while also being 20–30 times more efficient than conventional free energy methods. A structural analysis of key interactions of the most favorable chemical groups has been discussed and corroborates well crystallographic data published by Jorgensen and co-workers.[7,25–30] Furthermore, this study utilized GPU accelerated MD code, enabling the longest protein calculations to finish within 2–3 days, approximately 4–5 times faster than previous central processing unit (CPU)-based simulations. This work initiates future λ dynamics applications by demonstrating the ability to efficiently explore highly dimensional ligand chemical spaces with minimal computational costs compared to FEP/TI methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank Dr. John Chodera and Xinqiang Ding for insightful discussions and their review of preliminary work. The National Institutes of Health, through Grants GM037554 and GM107233, is gratefully acknowledged for financial support.

References

1. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the Pharmaceutical Industry: New Estimates of R&D. J Health Econ. 2016; 47:20–33. [PubMed: 26928437]
2. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge. Nat Rev Drug Discovery. 2010; 9:203–214. [PubMed: 20168317]
3. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of Early Drug Discovery. Br J Pharmacol. 2011; 162:1239–1249. [PubMed: 21091654]
4. Jorgensen WL. The Many Roles of Computation in Drug Discovery. Science. 2004; 303:1813–1818. [PubMed: 15031495]
5. Warr WA. A CADD-alog of Strategies in Pharma. J Comput-Aided Mol Des. 2017; 31:245–247. [PubMed: 28315993]
6. Abel R, Mondal S, Masse C, Greenwood J, Harriman G, Ashwell MA, Bhat S, Wester R, Frye L, Kapeller R, et al. Accelerating Drug Discovery Through Tight Integration of Expert Molecular Design and Predictive Scoring. Curr Opin Struct Biol. 2017; 43:38–44. [PubMed: 27816785]
7. Jorgensen WL. Computer-Aided Discovery of Anti-HIV Agents. Bioorg Med Chem. 2016; 24:4768–4778. [PubMed: 27485603]

8. Kudalkar SN, Beloor J, Quijano E, Spasov KA, Lee WG, Cisneros JA, Saltzman WM, Kumar P, Jorgensen WL, Anderson KS. From *In Silico* Hit to Long-Acting Late-Stage Preclinical Candidate to Combat HIV-1 Infection. Proc Natl Acad Sci U S A. 2018; 115:E802–E811. [PubMed: 29279368]

9. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, et al. 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. 2015; 137:2695–2703. [PubMed: 25625324]

10. Ciordia M, Perez-Benito L, Delgado F, Trabanco AA, Tresadern G. Applications of Free Energy Perturbation for the Design of BACE1 Inhibitors. J Chem Inf Model. 2016; 56:1856–1871. [PubMed: 27500414]

11. Zwanzig RW. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J Chem Phys. 1954; 22:1420–1426.

12. Kirkwood JG. Statistical Mechanics of Fluid Mixtures. J Chem Phys. 1935; 3:300–313.

13. Kollman P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. Chem Rev. 1993; 93:2395–2417.

14. Chodera JD, Mobley DL, Shirts MR, Dixon RW, Branson K, Pande VS. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. Curr Opin Struct Biol. 2011; 21:150–160. [PubMed: 21349700]

15. Chipot C. Frontiers in Free-Energy Calculations of Biological Systems. WIREs Comput Mol Sci. 2014; 4:71–89.

16. Steinbrecher TB, Dahlgren MK, Cappel D, Lin T, Wang L, Krilov G, Abel R, Friesner RA, Sherman W. Accurate Binding Free Energy Predictions in Fragment Optimization. J Chem Inf Model. 2015; 55:2411–2420. [PubMed: 26457994]

17. Liu P, Kim B, Friesner RA, Berne BJ. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. Proc Natl Acad Sci U S A. 2005; 102:13749–13754. [PubMed: 16172406]

18. Kong X, Brooks CL III. λ Dynamics: A New Approach to Free Energy Calculations. J Chem Phys. 1996; 105:2414–2423.

19. Knight JL, Brooks CL III. λ Dynamics Free Energy Simulation Methods. J Comput Chem. 2009; 30:1692–1700. [PubMed: 19421993]

20. Knight JL, Brooks CL III. Multisite λ Dynamics for Simulated Structure-Activity Relationship Studies. J Chem Theory Comput. 2011; 7:2728–2739. [PubMed: 22125476]

21. Armacost KA, Goh GB, Brooks CL III. Biasing Potential Replica Exchange Multisite λ Dynamics for Efficient Free Energy Calculations. J Chem Theory Comput. 2015; 11:1267–1277. [PubMed: 26579773]

22. Hayes RL, Armacost KA, Vilseck JZ, Brooks CL III. Adaptive Landscape Flattening Accelerates Sampling of Alchemical Space in Multisite λ Dynamics. J Phys Chem B. 2017; 121:3626–3635. [PubMed: 28112940]

23. Brooks BR, Brooks CL III, MacKerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al. CHARMM: The Biomolecular Simulation Program. J Comput Chem. 2009; 30:1545–1614. [PubMed: 19444816]

24. Brooks BR, Bruccoleri RE, Olafson BD, States BD, Swaminathan S, Karplus M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. J Comput Chem. 1983; 4:187–217.

25. Lee WG, Gallardo-Macias R, Frey KM, Spasov KA, Bollini M, Anderson KS, Jorgensen WL. Picomolar Inhibitors of HIV Reverse Transcriptase Featuring Bicyclic Replacement of a Cyanovinylphenyl Group. J Am Chem Soc. 2013; 135:16705–16713. [PubMed: 24151856]

26. Lee WG, Frey KM, Gallardo-Macias R, Spasov KA, Bollini M, Anderson KS, Jorgensen WL. Picomolar Inhibitors of HIV 1 Reverse Transcriptase: Design and Crystallography of Naphthyl Phenyl Ethers. ACS Med Chem Lett. 2014; 5:1259–1262. [PubMed: 25408842]

27. Bollini M, Domaoal RA, Thakur VV, Gallardo-Macias R, Spasov KA, Anderson KS, Jorgensen WL. Computationally-Guided Optimization of a Docking Hit to Yield Catechol Diethers as Potent Anti-HIV Agents. J Med Chem. 2011; 54:8582–8591. [PubMed: 22081993]

28. Frey KM, Bollini M, Mislak AC, Cisneros JA, Gallardo-Macias R, Jorgensen WL, Anderson KS. Crystal Structures of HIV-1 Reverse Transcriptase with Picomolar Inhibitors Reveal Key Interactions for Drug Design. J Am Chem Soc. 2012; 134:19501–19503. [PubMed: 23163887]

29. Frey KM, Puleo DE, Spasov KA, Bollini M, Jorgensen WL, Anderson KS. Structure-Based Evaluation of Non-Nucleoside Inhibitors with Improved Potency and Solubility that Target HIV Reverse Transcriptase Variants. J Med Chem. 2015; 58:2737–2745. [PubMed: 25700160]

30. Lee WG, Frey KM, Gallardo-Macias R, Spasov KA, Chan AH, Anderson KS, Jorgensen WL. Discovery and Crystallography of Bicyclic Arylaminoazines as Potent Inhibitors of HIV-1 Reverse Transcriptase. Bioorg Med Chem Lett. 2015; 25:4824–4827. [PubMed: 26166629]

31. Shirts MR, Chodera JD. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. J Chem Phys. 2008; 129:124105. [PubMed: 19045004]

32. Kroeger Smith MB, Hose BM, Hawkins A, Lipchock J, Farnsworth DW, Rizzo RC, Tirado-Rives J, Arnold E, Zhang W, Hughes SH, et al. Molecular Modeling Calculations of HIV-1 Reverse Transcriptase Nonnucleoside Inhibitors: Correlation of Binding Energy with Biological Activity for Novel 2-Aryl-Substituted Benzimidazole Analogues. J Med Chem. 2003; 46:1940–1947. [PubMed: 12723956]

33. Yin J, Henriksen NM, Slochower DR, Shirts MR, Chiu MW, Mobley DL, Gilson MK. Overview of the SAMPL5 Host–Guest Challenge: Are We Doing Better? J Comput-Aided Mol Des. 2017; 31:1–19. [PubMed: 27658802]

34. Liu S, Wu Y, Lin T, Abel R, Redmann JP, Summa CM, Jaber VR, Lim NM, Mobley DL. Lead Optimization Mapper: Automating Free Energy Calculations for Lead Optimization. J Comput-Aided Mol Des. 2013; 27:755–770. [PubMed: 24072356]

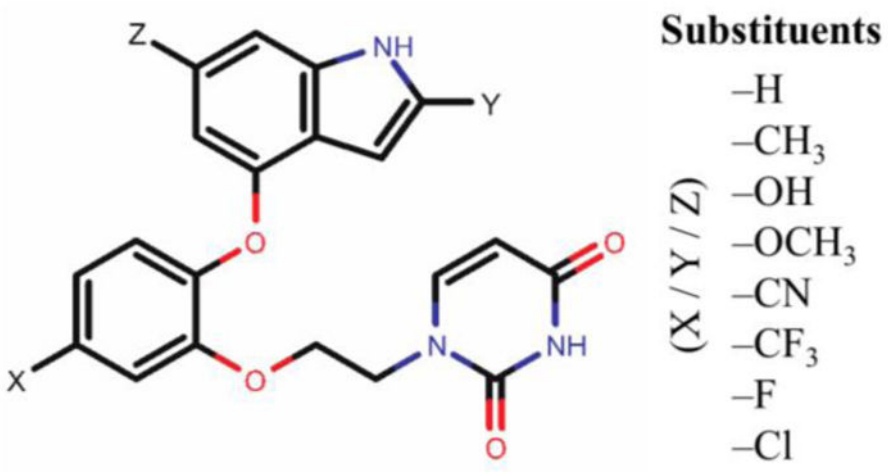


Figure 1.

Indole derivatized inhibitor simulated with BP-REX MSλD. Eight substitutions were investigated at three sites, X/Y/Z, for a total of 512 unique inhibitor designs.

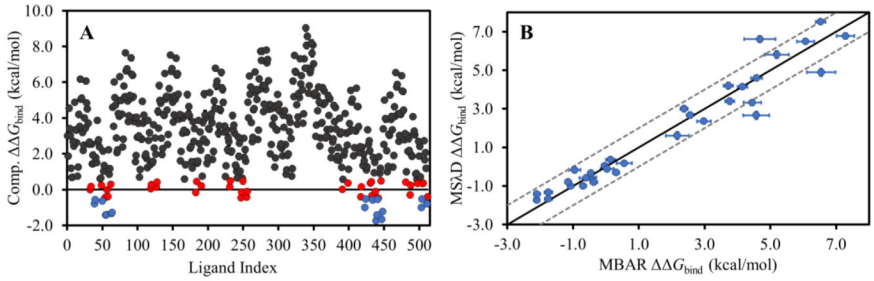


Figure 2.

(A) Relative free energies of binding for 512 indole inhibitors to HIV-RT computed with MSλD. Red data points highlight inhibitors predicted to be *as potent as* the reference compound (index 33); blue points indicate *more potent* affinities ($\Delta\Delta G_{\text{bind}} \leq -0.50$ kcal/mol). (B) Correlation between MBAR and MSλD computed $\Delta\Delta G_{\text{bind}}$ (kcal/mol) for 34 ligands. The solid black line represents y = x, and gray dashed lines represent y = x ± 1. Error bars are shown for all data points.

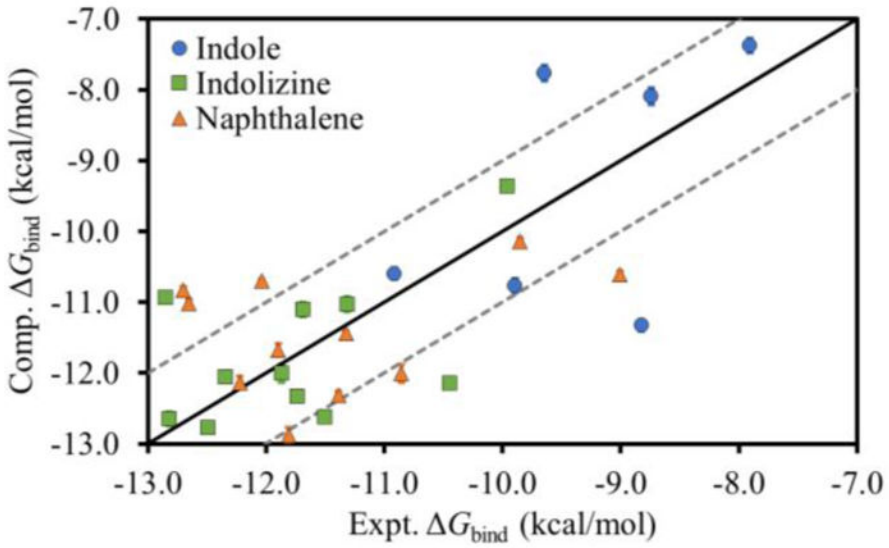


Figure 3.

Correlation between computed and experimental ΔG_{bind} for three ligand series bound to HIV-RT: indole (blue ●), indolizine (green ■), and naphthalene (orange ▲). Error bars show computed uncertainties. The solid black line is the y = x line, and gray dashed lines represent y = x ± 1.

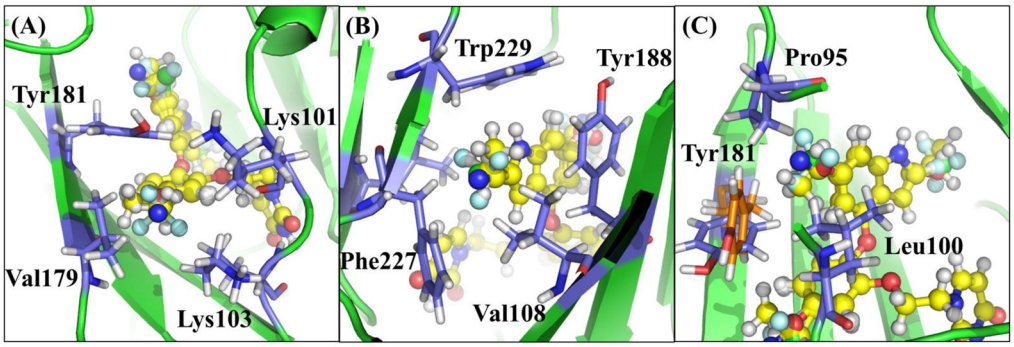


Figure 4.

Structural depictions of the non-nucleoside inhibitor binding pocket. Site X (A), site Y (B) and site Z (C) substituents with nearby residues are highlighted in yellow and indigo. An alternate conformation of Tyr181 is also shown in orange (C).