Computational Alchemy Using Accelerated GPU Calculations: Fine Structural Tuning Inhibitors of L,D-transpeptidase 2 from *Mycobacterium tuberculosis*

Semen M. Baldin[1,2] iD *, Vsevolod O. Shegolev*[1,3] iD *, Vytas K. Švedas*[1,2,4] iD

The computational alchemy is based on equilibrium methods of molecular dynamics (MD), including thermodynamic integration (TI) or free energy perturbation (FEP). It makes it possible to evaluate the energy characteristics of a subtle alchemical transformation of the prototype compound structures in order to modulate their functional properties, what can be used searching for new analogues of enzyme inhibitors. In recent years, the method is gaining new opportunities due to high-performance computing as different MD engines got the ability to perform TI or FEP calculations using GPU acceleration. We have tested the method on two experimentally studied enzyme-inhibitor systems: penicillin acylase from *E. coli* (PaEc) and Influenza A virus neuraminidase (NaIAv) and performed computational alchemy calculations for pairs of their known inhibitors. Thus, the validated methodology was then applied to estimate binding of L,D-transpeptidase 2 from *M. tuberculosis* (Ldt$_{Mt2}$) to the structural analogs of inhibitors previously discovered in our group. We have found that the performance of computational alchemy predictions is highly dependent on properly chosen simulation parameters and at optimal performance the difference between the calculated and experimentally determined relative binding energies is less than 1 kcal/mol. Thus, for optimal application to novel enzyme—inhibitor or receptor—ligand systems, we recommend prior adaptation of this method, based on validation of force field combinations as well as water models, to experimentally determined binding data of structurally related inhibitors/ligands to the molecular target of interest, if available, and then to proceed to the study of new compounds. High-performance computational alchemy can be used as a useful integrative part of the methodology searching for new enzyme inhibitors or receptor ligands and optimizing their structure at drug design.

Keywords: computational alchemy, thermodynamic integration, protein—ligand binding, binding free energy, drug design, L,D-transpeptidase.

Introduction

In classical molecular modeling, the veracity of calculation is believed to be a result of physically accurate approximations; contrariwise the molecular alchemy approach is dealing with non-physically, but mathematically correct models, which, on the one hand, can be even more precise than strict-physical models, and, on the other hand, adds new options to decrease computational efforts. The computational alchemy approach can be applied to solve some problems of computational chemistry (by so-called *alchemical transformation*, or "transmutation") such as calculation of the energy of hydration, estimation of relative changes of binding constants K_i for substrates or inhibitors, all this by changing structures of ligands or amino acid residues even within a binding site. The purpose of this work is to show how this approach can be applied to process a structure fine-tuning of known inhibitors in case to find new analogues with lower (better) inhibition constants K_i.

[1]Lomonosov Moscow State University, Belozersky Institute of Physicochemical Biology, Moscow, Russia
[2]Lomonosov Moscow State University, Research Computing Center, Moscow, Russia
[3]Lomonosov Moscow State University, Faculty of Chemistry, Moscow, Russia
[4]Lomonosov Moscow State University, Faculty of Bioengineering and Bioinformatics, Moscow, Russia

1. Theoretical Aspects

1.1. Statistical Mechanics

In canonical ensemble the number of microstates accecible to the system is described by the partition function Q. For the state $\mathcal{A}$ of the system,

$$Q_{\mathcal{A}}(N, V, T) = \frac{1}{N! h^{3N}} \iint e^{-\beta \mathcal{H}_{\mathcal{A}}(\mathbf{p}, \mathbf{q})} \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}, \tag{1}$$

where N is the number of (strictly speaking, indistinguishable) particles in the system, $\mathcal{H}_{\mathcal{A}}(p, q)$ is the Hamiltonian depending on generalized momenta $\mathbf{p} = \mathbf{p}_1, \cdots, \mathbf{p}_N$ and coordinates $\mathbf{q} = \mathbf{q}_1, \cdots, \mathbf{q}_N$ of all particles in the system, h is the Planck's constant, $\beta = 1/kT$ is the "thermodynamic beta", and k is the Boltzmann's constant.

Free energy A of the system in a state $\mathcal{A}$, and its difference between two states $\mathcal{A}$ and $\mathcal{B}$ of the system can be written using (1) as

$$A_{\mathcal{A}} = -\frac{1}{\beta} \ln Q_{\mathcal{A}} \quad \text{and} \tag{2}$$

$$\Delta A_{\mathcal{A}\mathcal{B}} = A_{\mathcal{B}} - A_{\mathcal{A}} = -\frac{1}{\beta} \ln \frac{Q_{\mathcal{B}}}{Q_{\mathcal{A}}} \quad \text{accordingly.} \tag{3}$$

In molecular dynamics, the expression of the Hamiltonian $\mathcal{H}$ of the system of N particles with masses m_i and Cartesian coordinates $\mathbf{r} = \mathbf{r}_1, \cdots, \mathbf{r}_N$ [21] is

$$\mathcal{H}(\mathbf{p}, \mathbf{r}) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{r}_1, \cdots, \mathbf{r}_N), \tag{4}$$

potential term U of the Hamiltonian is usually defined[5] [6] as

$$U = \sum_b k_b(l_b - \bar{l}_b)^2 + \sum_a k_a(\theta_a - \bar{\theta}_a)^2 \tag{5}$$

$$+ \sum_d \sum_n k_{d,n} \left\{ 1 + \cos\left(n\phi_d - \phi_d'\right) \right\} \tag{6}$$

$$+ \sum_{i<j} \left\{ \frac{A_{i,j}}{r_{i,j}^{12}} - \frac{B_{i,j}}{r_{i,j}^{6}} + k_e \frac{q_i q_j}{r_{i,j}} \right\}, \tag{7}$$

where the energy of deformation from the average value is approximated by Hooke's quadratic potentials (5) for bond length l_b and valence angles θ_a, and by periodic Fourie series (6) for dihedrals ϕ_d and its harmonics n; pairwise particles i, j interactions (7) are described by Lennard-Jones "6-12" potential and Coulomb's law and decrease in distance $r_{i,j}$; $k_b, \bar{l}_\theta, k_a, \bar{\theta}_a, k_{d,n}, \phi_d', A_{i,j}, B_{i,j}, k_e, q_i$, and q_k are the relevant empirical parameters which consist the *force filed*.

1.2. Free Energy Perturbation

Substituting the expression (5–7) for U in (4), then in (1), and then in (3), reducing the fraction $Q_{\mathcal{A}}/Q_{\mathcal{B}}$ in (3) by the same kinetic energy term $\exp \sum_i \frac{\mathbf{p}_i^2}{2m_i}$ leads to

$$\Delta A_{\mathcal{A}\mathcal{B}} = -\frac{1}{\beta} \ln \frac{Z_{\mathcal{B}}}{Z_{\mathcal{A}}}, \tag{8}$$

---

[5]in AMBER family of force fields

where configurational partition function Z depends only on coordinates $\mathbf{r}$, e.g., for the $\mathcal{B}$ state:

$$Z_{\mathcal{B}}(N, V, T) = \int e^{-\beta U_{\mathcal{B}}(\mathbf{r})} \mathrm{d}\mathbf{r}, \tag{9}$$

this equation can be rewritten in terms of the ensemble averages as

$$Z_{\mathcal{B}} = \int e^{-\beta U_{\mathcal{B}}} \mathrm{d}\mathbf{r} = \int e^{-\beta U_{\mathcal{B}}} e^{\beta U_{\mathcal{A}}} e^{-\beta U_{\mathcal{A}}} \mathrm{d}\mathbf{r} = \int e^{-\beta(U_{\mathcal{B}} - U_{\mathcal{A}})} \underbrace{e^{-\beta U_{\mathcal{A}}}}_{\rho_{\mathcal{A}}} \mathrm{d}\mathbf{r} = \left\langle e^{-\beta(U_{\mathcal{B}} - U_{\mathcal{A}})} \right\rangle_{\mathcal{A}}, \tag{10}$$

where $\rho_{\mathcal{A}}$ is a distribution function of the state $\mathcal{A}$, $\langle X \rangle_{\mathcal{A}}$ denotes the average of X in ensemble with the distribution $e^{-\beta U_{\mathcal{A}}}$, and finally, (8) takes the form named *free energy perturbation formula* [26]:

$$\Delta A_{\mathcal{AB}} = -\frac{1}{\beta} \ln \langle e^{-\beta(U_{\mathcal{B}} - U_{\mathcal{A}})} \rangle_{\mathcal{A}}. \tag{11}$$

Physically, (11) means that microstates of the state $\mathcal{A}$ can be used to estimate the average of the function $e^{-\beta \Delta_{\mathcal{AB}} U}$ and then calculate transitional free energy $\Delta A_{\mathcal{AB}}$ between two states $\mathcal{A}$ and $\mathcal{B}$, it means that microstates of state $\mathcal{A}$ must also be the microstates of the state $\mathcal{B}$ with a sufficiently high probability to get an accurate estimate of the average $\langle \cdots \rangle_{\mathcal{A}}$, i.e., the state $\mathcal{B}$ must be only the *small perturbation* of the state $\mathcal{A}$.

1.3. Thermodynamic Integration

Equation (11) can be generalized to estimate $\Delta A_{\mathcal{AB}}$ between arbitrary states $\mathcal{A}$ and $\mathcal{B}$, close or not. It is only enough to maintain intermediate states s (with distrubutions $e^{-\beta U_s}$) between $\mathcal{A}$ and $\mathcal{B}$, satisfying the "closeness" condition:

$$\Delta A_{\mathcal{AB}} = -\frac{1}{\beta} \sum_s \ln \langle e^{-\beta(U_{s+1} - U_s)} \rangle_s, \tag{12}$$

or, in continuous case, integration through the transformational path λ can be proceeded by defining a sufficient transformation between two states in potential term of the Hamiltonian:

$$U(\lambda) = \lambda U_{\mathcal{A}} + (1 - \lambda) U_{\mathcal{B}}, \tag{13}$$

from hence one can obtain a formula

$$\Delta A_{\mathcal{AB}} = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda \mathrm{d}\lambda, \tag{14}$$

by using (2) and the following:

$$\frac{\partial A}{\partial \lambda} = -\frac{1}{\beta Q} \frac{\partial Q}{\partial \lambda} = -\frac{1}{\beta Z} \frac{\partial Z}{\partial \lambda} = -\frac{1}{\beta Z} \frac{\partial}{\partial \lambda} \int e^{-\beta U(\mathbf{q};\lambda)} \mathrm{d}\mathbf{q} =$$
$$= -\frac{1}{\beta Z} \int \left(-\beta \frac{\partial U(\mathbf{q};\lambda)}{\partial \lambda} \right) e^{-\beta U(\mathbf{q};\lambda)} \mathrm{d}\mathbf{q} = \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda. \tag{15}$$

The resulting relation (14) is called the *thermodynamic integration* formula. In practice, from integration again switch to summation:

$$\Delta A_{\mathcal{AB}} \approx \sum_s \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda_s} \delta\lambda_s. \tag{16}$$

1.4. Computational Alchemy

As mentioned above, thermodynamic integration (14) can be performed through a transformational path λ connecting any two states of the system. Let us suppose, just conceptually, at the start, we have a system of a ligand in water, at the state $\mathcal{A}$, the ligand has a chemical formula L_1, during the all "alchemical" path λ, the ligand is continuously transforming into the chemical formula L_2, corresponding to the state $\mathcal{B}$. In the other words, a chemical molecule undergoes non-physical transformation, or "transmutation" during which the whole groups of its atoms are changing their chemical nature. The described procedure can be easily implemented *in silico* in two different ways.

1.4.1. The single-topology paradigm

Empirical parameters of the force filed (5–7), can be transformed during the molecular dynamics simulation, e.g., phenol molecule can be transmutated to thiophenol by changing the numerical parameters of the oxygen atom O into the sulfur S through the "fictitious" atom $X(\lambda)$ which is in the functional dependence on path λ (Fig. 1).
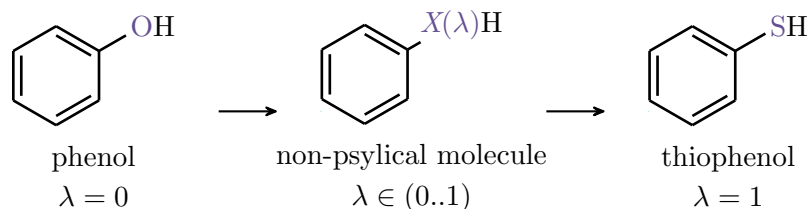

Figure 1. Scheme of alchemical transformation in single-topology paradigm

Mathematically, all force field parameters of the transforming atoms are changing along the path $\lambda = 0..1$:

$$\begin{cases} m_X(\lambda) = \lambda m_O + (1 - \lambda)m_S, \\ \quad\vdots \\ q_X(\lambda) = \lambda q_O + (1 - \lambda)q_S. \end{cases} \tag{17}$$

Then, calculated parameters involve into forces calculation as (5–7), and integration of motion equation according to the Newton's 2nd law. In that case, the one molecule topology is transforming.

1.4.2. The dual-topology paradigm

Alternatively, Equation (13) can be implemented more straightforwardly by combining two different topologies of the physical molecules at the start $\mathcal{A}$ and at the end $\mathcal{B}$ of the alchemical path λ. Pairs of atoms, common for both topologies, share their coordinates and are treated as a single atom in all interactions with the environment. In contrast, differing atoms do not undergo a transformation as in the single-topology paradigm, instead, they do not sense the existence of each other, but their interactions with the environment are scaled according to covered path λ (Fig. 2). So, from the outside, the molecule looks like a chimera of two.

Values of λ can lay anywhere on the interval $[0..1]$. Because Lennard-Jones's and electric potential tends to $+\infty$ when $r \to 0$, the transition from $\lambda \equiv 0$ to $\lambda \sim 0$ will be unstable because a potential of appearing atoms will grow explosively in regions of $r \sim 0$, so some "soft potential"
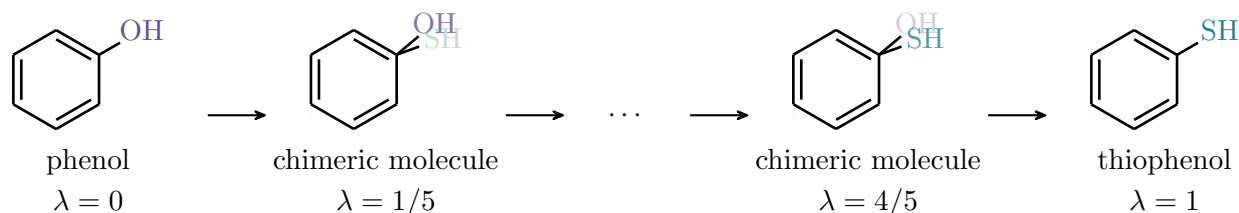
phenol
$\lambda = 0$

chimeric molecule
$\lambda = 1/5$

$\cdots$

chimeric molecule
$\lambda = 4/5$

thiophenol
$\lambda = 1$

Figure 2. Scheme of alchemical transformation in dual-topology paradigm

approximation is used [2, 18]:

$$U_{i,j}^{\mathrm{LJ}}(r_{i,j}, \lambda) = 4\epsilon_{i,j}\lambda^n \left\{ \frac{1}{[\alpha(1-\lambda)^2 + (r_{i,j}/\sigma_{i,j})^6]^2} - \frac{1}{\alpha(1-\lambda)^2 + (r_{i,j}/\sigma_{i,j})^6} \right\}, \quad (18)$$

where $\epsilon_{i,j}, \sigma_{i,j}$ are Lennard-Jones parameters, n and α are numerical constants; and similarly for electrostatics:

$$U_{i,j}^{\mathrm{Col}}(r_{i,j}, \lambda) = (1-\lambda)\frac{q_i q_j}{4\pi\varepsilon_0\sqrt{\beta\lambda + r_{i,j}^2}}, \quad (19)$$

where ε_0 is the permittivity, q_i and q_j are charges, β is a numeric constant.

1.5. Thermodynamic Cycle

Computational alchemy by itself is not very usable, however, the power of this technique is revealed when using a hybrid experimental-computational approach. In the previous section, we have described the transformation $L_1 \to L_2$ for a ligand in bulk. In the same way, the ligand can be transformed while being in a complex with a protein E, i.e. $EL_1 \to EL_2$. Thus, we can construct a thermodynamic cycle (Fig. 3): up and down edges of the cycle correspond to free energies $\Delta_{\mathrm{alc}}A_{\mathrm{lig}}$ (free ligand) and $\Delta_{\mathrm{alc}}A_{\mathrm{com}}$ (ligand in a complex) of alchemical transformations, left and right, to experimentally measurable ligands-protein binding energies $\Delta_b G$.

The expression of the free energy A is defined by the thermodynamic ensemble, for canonical NVT, $A \equiv F$ (the Helmholtz free energy); in isothermal–isobaric NPT ensemble, $A \equiv G$ (the Gibbs free energy). Nevertheless, in condensed phases, $\Delta G \approx \Delta F \approx \Delta A$ in a large pressure range [12], further, we will imply that.
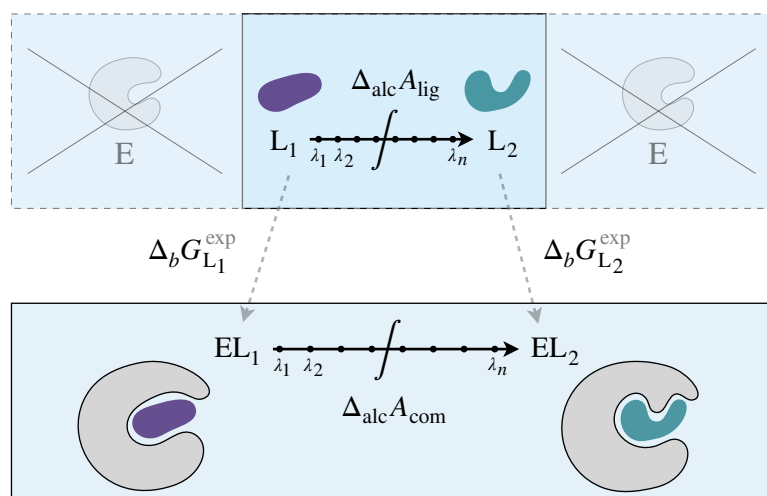

Figure 3. Thermodynamic cycle for calculation of the relative binding energy of two ligands L_1 and L_2 in the active site of the protein E

Having an experimental value $\Delta_b G_{L_1}$ for binding of L_1 one can estimate $\Delta_b G_{L_2}$ for an arbitrary ligand L_2 by using the thermodynamic cycle:

$$\Delta_b G_{L_2} = \Delta_b G_{L_1} + \Delta_{\text{alc}} A_{\text{com}} - \Delta_{\text{alc}} A_{\text{lig}}$$
$$= \Delta_b G_{L_1} + \underbrace{\Delta_{\text{alc}} G_{\text{com}} - \Delta_{\text{alc}} G_{\text{lig}}}_{\Delta\Delta_{\text{alc}} G}, \tag{20}$$

where $\Delta\Delta_{\text{alc}} G$ is the computed alchemical correction for the experimental $\Delta_b G_{L_1}$. Hence, the ratio for the relative binding energies is

$$\Delta\Delta_b G = \Delta\Delta_{\text{alc}} G. \tag{21}$$

2. Acceleration of TI MD Calculations Using GPU

Although the acceleration of GPU calculations for the molecular dynamics (MD) method has been implemented in many MD engines for a while, the GPU acceleration for the thermodynamic integration (TI) method has been tested relatively recently [9, 10]. It should be noted that the motivation for the implementation of GPU acceleration for TI method was to speed up calculations of relative binding energies of inhibitors with enzymes in order to adjust the inhibitor structure to the geometry of the binding site [7].

Let us discuss a typical computational task for estimation of the relative binding energy of a pair of inhibitors in a binding site of an enzyme. As it was previously revealed, to estimate the free energy of the alchemical transformation we need to do two types of calculations: transformation in the bulk (*ligands*) and in a binding site of an enzyme (*complex*).

For the first system, considering the distance from the solute to the modeling cell boundaries is from 12 to 15 Å, the size of such a box will be about 5K to 10K atoms. For the second system, the size of the solvate box will be primarily determined by the size of the enzyme model. Using the examples described later in this paper, one can find the sizes of the simulated systems: for the model of L,D-transpeptidase 2 from *M. tuberculosis* (Ldt$_{\text{Mt2}}$) model (373 aa, or amino acid residues) the solvate box size will be around 66 K atoms, for penicillin acylase from *E. coli* (PaEc) model (764 aa) the solvate box size will be around 111 K atoms, for 4 subunit neuraminidase from Influenza A virus subtype H_2N_2 (NaIAv) model (1876 aa) the solvate box size will be around 178 K atoms. The latter one is especially big system to run molecular dynamic without any acceleration hardware.

To estimate $\Delta\Delta_{\text{alc}} G$ for each of the described systems, it is necessary to calculate the molecular dynamic trajectory for all λ in the course of transformations. Let us take the minimum number of λ windows equal to 9, which are used in some cases in this work. In addition, to check the reproducibility of the results it is necessary to perform at least three repetitions (runs) of the same calculations. Thus, according to the most conservative estimate, in order to reliably calculate one $\Delta\Delta_{\text{alc}} G$ value, it is necessary to calculate 27 molecular dynamics trajectories for the *ligands* system and 27 trajectories for the *complex* system. In case if $\partial U / \partial \lambda$ converges on a time between 10 and 50 nanoseconds (ns) in every MD trajectory, one could see that from 270 to 1350 ns in total of MD trajectories for both *complex* and *ligands* systems are needed to calculate in acceptable period of time. The number of available to compute ns of MD trajectories per day is given in Tab. 1 for a few desktop-class GPUs in comparison to 27x Tesla P100 [22].

Obviously, only with parallel computing on multiple HPC-class GPUs it is possible to achieve the required calculation speed and obtain one $\Delta\Delta_{\text{alc}} G$ value per day, which is necessary if the

Table 1. Comparison of speed in **ns/day** of MD TI calculations for 3 systems on desktop and HPC GPU

system and size	GTX 980	RTX 2070 Super	RTX 3080	27x Tesla P100 [22]
Ldt$_{Mt2}$ (66 K atoms)	24.3	49.1	87.4	1360.8
PaEc (111 K atoms)	12.6	26.1	45.5	—
NaIAv (178 K atoms)	—	15.7	27.9	537.8

task requires to estimate dozens of $\Delta\Delta_{alc}G$ values for potential analogues of inhibitors and their one-to-one transformations.

3. Systems Preparation and Simulation Protocol

The full atomic structure of penicillin acylase from *E. coli* (PaEc) in a complex with oxidised penicillin G (PDB 1GM9) was used as a starting structure for the construction of enzyme-inhibitor complexes of PaEc with *R*-mandelic acid, *S*-mandelic acids and phenylacetic acid.

The full atomic four subunit model of neuraminidase from Influenza A virus subtype H_2N_2 in a complex with inhibitors was built from a complex with zanamivir (PDB 3TIC).

The three-domain structure of L,D-transpeptidase 2 from *M. tuberculosis* without a signal peptide was assembled from two crystal structures: for domains A and B (PDB 4HU2), and B and C (PDB 4HUC). The resulting structure was equilibrated by classical MD followed by accelerated MD to obtain open conformations of a flexible lid covering the active site. MD frames with the open lid were used to construct enzyme-inhibitor complexes by molecular docking.

Partial charges on inhibitors atoms were obtained by geometry optimization by the B3LYP/6-31+G* method followed by calculation of the electrostatic potential by the HF/6-31G* method and charge fitting by the RESP method [1]. The types of inhibitors atoms were defined using GAFF [25] or GAFF2 force fields, enzymes were described using ff14SB [14] or ff19SB [20]. TIP3P or OPC [8] water models were used.

MD parameters of simulations were as follows: $\Delta t = 1$ fs, Langevin thermostat with collision frequency $\gamma = 2$ ps^{-1}, $T = 298$ K, Monte Carlo barostat with pressure relaxation time $\tau = 2$ ps, $P = 1.01325$bar, SHAKE was off, NTP ensemble, nonbonded cutoff for PME was 8.0 Å, for TI calculations softcore potentials were applied for Lennard-Jones and Coulomb potentials with $\alpha = 0.5$ and $\beta = 12.0$ Å^2 respectively. The Amber package was used to simulate MD trajectories [4].

The stresses in the constructed systems were eliminated by minimizing the energy using the steepest descent algorithm. The systems were heated to 298 K on NVT ensemble, equilibrated on NPT ensemble until the nominal values of water density were reached and 5 ns after that. Separate trajectories were created for each value of λ using the last frame of the previous trajectory. For each value of λ the velocities obtained at the previous step were not taken into account, but the systems were heated again at a given value of λ, then equilibrated again using NPT ensemble for the first 10% of the trajectory and during the last 90% of time, values of $\partial U/\partial \lambda$ were collected. The convergence of $\partial U/\partial \lambda$ values was monitored by the values of the moving average and the incremental average. The obtained values of $\partial U/\partial \lambda$ for all λ windows were integrated by the appropriate method, depending on the partition of the transformation path (trapezoidal or Gaussian).

4. Method Validation

4.1. Inhibitors of Penicillin Acylase from *E. coli* (PaEc)

The penicillin acylase is widely used in industry to catalyze an acyl transfer reaction from the activated acyl donors to the amino group of the penicillin nucleus 6-aminopenicillanic acid. Phenylacetic acid and its derivatives are also PaEc inhibitors. The inhibition constants for phenylacetic, *R*-mandelic and *S*-mandelic acids are 0.06 mM, 31 mM and 10 mM, respectively, which means (according to $\Delta\Delta_b G = -RT \ln\left(K_i^1/K_i^2\right)$) $\Delta\Delta_b G_{\mathrm{exp}} = -3.7$ kcal/mol for transformation from *R*-mandelic acid to phenylacetic acid and $\Delta\Delta_b G_{\mathrm{exp}} = -3.0$ kcal/mol for transformation from *S*-mandelic acid to phenylacetic acid.
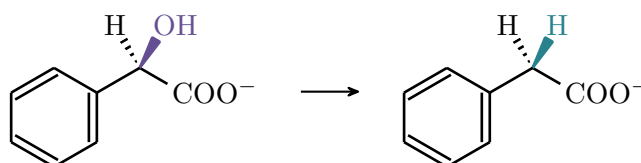

Figure 4. A scheme of alchemical transformation from *R*-**mandelic acid** (left) to **phenylacetic acid** (right). Colored atoms are those to which the soft-core potentials were applied

Firstly, for transformation from *R*-mandelic acid to phenylacetic acid (Fig. 4) we decided to use default force fields parameters for TI molecular dynamics (ff14SB, GAFF, TIP3P water) and 13 λ windows (0.0, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 1.0). One could see we add 2 additional λ values on the sides of the transformation path. Calculated values of $\Delta\Delta_{\mathrm{alc}} G$ are given in Tab. 2. After five runs of calculation, the estimated interval $\Delta\Delta_{\mathrm{alc}} G = -1.0 \pm 0.4$ kcal/mol does not contain the experimental value $\Delta\Delta_b G_{\mathrm{exp}} = -3.7$ kcal/mol. At the same time, the computed value agrees with the experimental trend, but the absolute error is 2.7 kcal/mol.

When analyzing the $\partial U/\partial\lambda$ of λ graphs, we found that the maximum change of $\partial U/\partial\lambda$ values occurs in the interval between 0 and 0.1 in λ scale. Afterwards, to sample this interval more rigorously and improve the estimation of $\Delta\Delta_{\mathrm{alc}} G$, we introduced 5 additional λ windows (0.01, 0.02, 0.03, 0.04 and 0.075), their total number was increased to 18. Having performed similar calculations, we obtained a new estimate of $\Delta\Delta_{\mathrm{alc}} G = -1.3 \pm 0.2$ kcal/mol. This estimate is better in terms of reproducibility and closer to the experimental value of $\Delta\Delta_b G_{\mathrm{exp}}$ but still differs from the experimental value by 2.4 kcal/mol.

After several trials, we decided to choose more modern force fields: ff19SB for the enzyme and GAFF2 for the inhibitors, relatively new OPC water model that is recommended [19] to be used with ff19SB was applied. Under new conditions we estimate $\Delta\Delta_{\mathrm{alc}} G = -2.6 \pm 0.7$ kcal/mol which is better than previous result and differ from the experimental value by 1.1 kcal/mol.
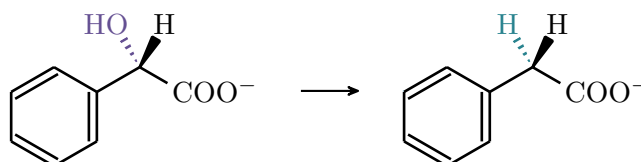

Figure 5. A scheme of alchemical transformation from *S*-**mandelic acid** (left) to **phenylacetic acid** (right). Colored atoms are those to which the soft-core potentials were applied

After applying a similar approach to modeling the conversion of S-mandelic acid to phenylacetic acid (Fig. 5) using only 13 λ windows we get estimate of $\Delta\Delta_{\mathrm{alc}}G = -3.7 \pm 1.2$ kcal/mol, while experimental value is -3.0 kcal/mol. Calculated values of $\Delta\Delta_{\mathrm{alc}}G$ are given in Tab. 3.

Table 2. Calculated thermodynamic integrals of alchemical transformation for a pair of R-**mandelic acid** and **phenylacetic acid** in the complex with PaEc ($\Delta_{\mathrm{alc}}G_{\mathrm{com}}$) and in the bulk ($\Delta_{\mathrm{alc}}G_{\mathrm{lig}}$) for different modeling parameters. All numerical values are given in **kcal/mol**

run	$\Delta_{\mathrm{alc}}G_{\mathrm{com}}$	$\Delta_{\mathrm{alc}}G_{\mathrm{lig}}$	$\Delta\Delta_{\mathrm{alc}}G$	$\overline{\Delta\Delta_{\mathrm{alc}}G} \pm \mathrm{CI}_{95\%}$	$\Delta\Delta_b G_{\mathrm{exp}}$
	ff14SB, GAFF, TIP3P, 13 λ windows				
1	-73.71	-72.73	-0.98		
2	-73.88	-72.74	-1.14		
3	-74.10	-72.77	-1.33	-1.0 ± 0.4	-3.7
4	-73.56	-72.75	-0.81		
5	-73.32	-72.75	-0.57		
	ff14SB, GAFF, TIP3P, 18 λ windows				
1	-74.21	-72.88	-1.33		
2	-74.13	-72.89	-1.24	-1.3 ± 0.2	-3.7
3	-74.30	-72.90	-1.40		
	ff19SB, GAFF 2, OPC, 13 λ windows				
1	-76.13	-73.32	-2.81		
2	-76.17	-73.56	-2.61	-2.6 ± 0.7	-3.7
3	-76.93	-74.69	-2.24		

Table 3. Calculated thermodynamic integrals of alchemical transformation for a pair of S-**mandelic acid** and **phenylacetic acid** in the complex with PaEc ($\Delta_{\mathrm{alc}}G_{\mathrm{com}}$) and in the bulk ($\Delta_{\mathrm{alc}}G_{\mathrm{lig}}$). Parameters of modelling: ff19SB, GAFF2, OPC, 13 λ windows. All numerical values are given in **kcal/mol**

run	$\Delta_{\mathrm{alc}}G_{\mathrm{com}}$	$\Delta_{\mathrm{alc}}G_{\mathrm{lig}}$	$\Delta\Delta_{\mathrm{alc}}G$	$\overline{\Delta\Delta_{\mathrm{alc}}G} \pm \mathrm{CI}_{95\%}$	$\Delta\Delta_b G_{\mathrm{exp}}$
1	-78.56	-74.38	-4.18		
2	-78.27	-74.70	-3.57	-3.7 ± 1.2	-3.0
3	-78.42	-75.22	-3.20		

As a conclusion for penicillin acylase inhibitors we found out that ff19SB+GAFF2+OPC performs better in comparison with ff14SB+GAFF+TIP3P.

4.2. Inhibitors of Neuraminidase from Influenza A virus (NaIAv)

The neuraminidase enzyme of influenza virus cleaves of sialic acid from human receptors to facilitate the release of newly formed viral particles from the cell surface and is a molecular target for the most effective anti-influenza drugs zanamivir and oseltamivir [11, 23, 24]. Several sialic acid analogue inhibitors of the enzyme have been synthesized and characterized experimentally [24].

Let us look at the pair 4-amino-Neu5Ac2en and 4-guanidino-Neu5Ac2en (zanamivir) (Fig. 6), the inhibition constants K_i are 50.0 nM and 0.2 nM respectively [24], which means, $\Delta\Delta_b G_{\text{exp}} = -3.3$ kcal/mol for the transformation of 4-amino-Neu5Ac2en to 4-guanidino-Neu5Ac2en.
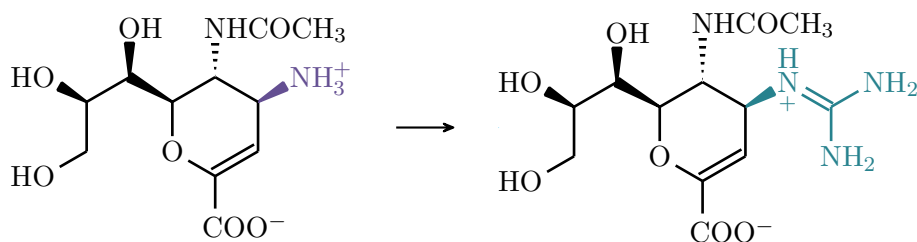

Figure 6. A scheme of alchemical transformation from **4-amino-Neu5Ac2en** (left) to **4-guanidino-Neu5Ac2en**, or **zanamivir**, (right). Colored atoms are those to which the soft-core potentials were applied

In this case, we used ff14SB+GAFF+TIP3P to describe the system. We split the transformation path into 9 λ windows (0.01592, 0.08198, 0.19331, 0.33787, 0.5, 0.66213, 0.80669, 0.91802, 0.98408) according to the n-point Gaussian quadrature ($n = 9$), then Gauss integration method was applied to estimate $\Delta\Delta_{\text{alc}} G$. As can be seen in Tab. 4, after three runs of the calculations, $\Delta\Delta_{\text{alc}} G = -3.3 \pm 0.7$ kcal/mol which is very close to the experimental (-3.3 kcal/mol) value calculated from known K_i values.

Table 4. Calculated thermodynamic integrals of alchemical transformation for a pair of **4-amino-Neu5Ac2en** and **4-guanidino-Neu5Ac2en** (zanamivir) in the complex with NaIAv ($\Delta_{\text{alc}} G_{\text{com}}$) and in the bulk ($\Delta_{\text{alc}} G_{\text{lig}}$). Parameters of modelling: ff14SB, GAFF, TIP3P, 9 λ windows. All numerical values are given in **kcal/mol**

run	$\Delta_{\text{alc}} G_{\text{com}}$	$\Delta_{\text{alc}} G_{\text{lig}}$	$\Delta\Delta_{\text{alc}} G$	$\overline{\Delta\Delta_{\text{alc}} G} \pm \text{CI}_{95\%}$	$\Delta\Delta_b G_{\text{exp}}$
1	-8.35	-4.94	-3.41		
2	-8.07	-4.98	-3.09	-3.3 ± 0.7	-3.3
3	-8.65	-5.06	-3.59		

4.3. Inhibitors of L,D-transpeptidase from *M. tuberculosis* (Ldt$_{\text{Mt2}}$)

The L,D-transpeptidase 2 from *M. tuberculosis* is an enzyme catalyzing cross-linking of peptidoglycan chains in the pathogenic cell wall [5]. This protein is a potential target for the new class of antitubercular medications (anti-TB) due to high importance for the growth of the pathogenic bacteria. Well-known β-lactams do not provide efficient anti-TB effect [16, 17], so an active search for new treatments is underway: new combinations of known anti-TB drugs are being studied [15] and new molecules are being sought [13].

Competitive inhibitors of L,D-transpeptidase 2 from *M. tuberculosis* are considered as perspective anti-TB preparations [3]. We have found a new representative by *in silico* screening of commercial ligand libraries. The selected compound, inhA (Fig. 7) was then experimentally studied, characterized as Ldt$_{\text{Mt2}}$ inhibitor and used as a molecule at computational alchemy optimization. The resulting improved structure inhB of the inhibitor is shown in Fig. 7 and the data

of the corresponding energy calculations are presented in Tab. 5. The computed value of relative binding energy ($\Delta\Delta_{\mathrm{alc}}G = -1.6 \pm 0.6$ kcal/mol) shows an appropriate fit with experimental data, which has confirmed more effective inhibition of $\mathrm{Ldt_{Mt2}}$ by the designed compound inhB ($K_i = 5$ μM) compared to the parent structure inhA ($K_i = 14$ μM); the absolute error between calculated and experimental relative free energy of binding ($\Delta\Delta_b G = -RT\ln\left(K_i^{\mathrm{inhA}}/K_i^{\mathrm{inhB}}\right) = -0.6$ kcal/mol) is in order of 1 kcal/mol.
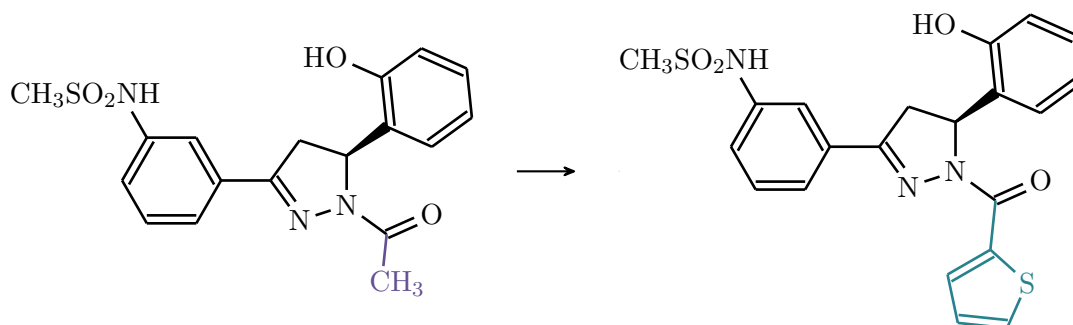

Figure 7. A scheme of alchemical transformation from **inhA** (left) to **inhB** (right). Colored atoms are those to which the soft-core potentials were applied
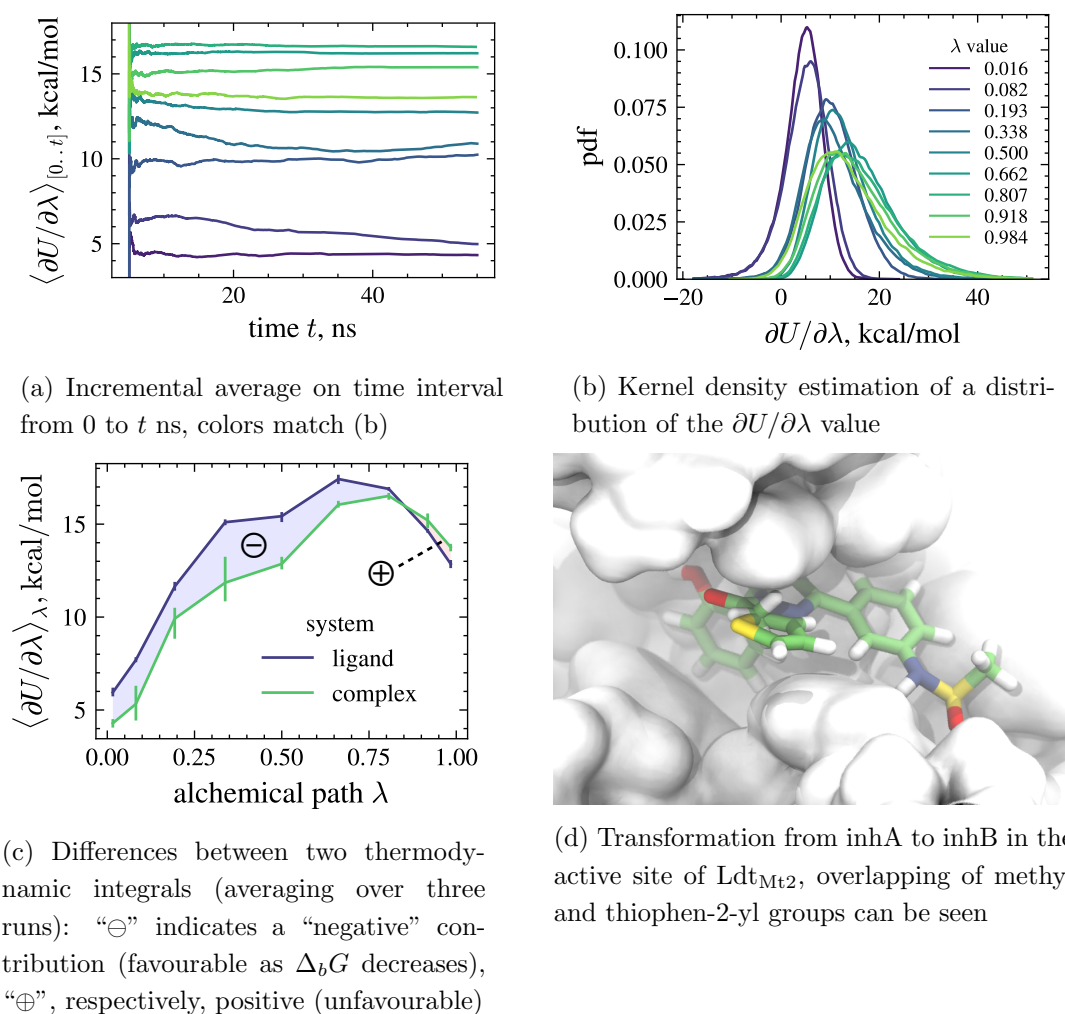

(a) Incremental average on time interval from 0 to t ns, colors match (b)

(b) Kernel density estimation of a distribution of the $\partial U/\partial\lambda$ value

(c) Differences between two thermodynamic integrals (averaging over three runs): "$\ominus$" indicates a "negative" contribution (favourable as $\Delta_b G$ decreases), "$\oplus$", respectively, positive (unfavourable)

(d) Transformation from inhA to inhB in the active site of $\mathrm{Ldt_{Mt2}}$, overlapping of methyl and thiophen-2-yl groups can be seen

Figure 8. Visualization of $\partial U/\partial\lambda$ during the alchemical transformation of found and tested competitive inhibitors inhA and inhB of $\mathrm{Ldt_{Mt2}}$

Table 5. Calculated thermodynamic integrals of alchemical transformation for a pair **inhA**, **inhB** in the complex with Ldt_{Mt2} ($\Delta_{alc}G_{com}$) and in the bulk ($\Delta_{alc}G_{lig}$). Parameters of modelling: ff14SB, GAFF, TIP3P, 9 λ windows. All numerical values are given in **kcal/mol**

run	$\Delta_{alc}G_{com}$	$\Delta_{alc}G_{lig}$	$\Delta\Delta_{alc}G$	$\overline{\Delta\Delta_{alc}G} \pm CI_{95\%}$	$\Delta\Delta_b G_{exp}$
1	12.69	14.06	-1.37		
2	12.36	14.18	-1.81	-1.6 ± 0.6	-0.6
3	12.48	14.15	-1.67		

The parameters of computational experiment were as follows: nine λ values were chosen in an interval from 0 to 1 (according to 9-point Gaussian quadrature integration), at least 50 ns of MD trajectories were recorded for all 9 λ on three runs, i.e. [50 ns $\times$ 9 λ MDs] $\times$ 3 runs.

Let us take a closer look at how we processed the raw calculated data to obtain estimates of the relative binding free energy. Values of $\partial U/\partial \lambda$ obtained during the simulations were threated to compute two TIs (Gaussian quadrature integration was used) and then $\Delta\Delta_{alc}G$. Using an example of one run (9 MDs for each λ value) we can briefly describe the results (Fig. 8). Estimates of the averages $\langle \partial U/\partial \lambda \rangle_{[0..t]}$ reach a constant value in the first 20–30 ns of MD trajectories (Fig. 8a), all observations after 20 ns were used to get the estimation (for all runs separately). Values of $\partial U/\partial \lambda$ are distributed around their means, forming symmetric Gaussian-like "bell" curves (Fig. 8b). Since the difference of two TIs is required to calculate relative energies $\Delta\Delta_{alc}G$ the "negative" and "positive" areas are shown in Fig. 8c, it can be seen that the larger in absolute value term corresponds to a decrease in the free energy of binding.

The computational alchemy allows to estimate relative binding energies (21) which makes it useful for structure optimization. It should be noted, that a straightforward transformation with interpretable results can only be performed between close structural analogues (see subsection 1.2) which were found by methods of chemoinformatics (fingerprints, maximum common subgraph). Thus, we can conclude that the method can be used to optimize the structure of inhibitors.

Conclusion

The computational alchemy is based on equilibrium methods of molecular dynamics (MD), including thermodynamic integration (TI) or free energy perturbation (FEP). It makes it possible to evaluate the energy characteristics of a subtle alchemical transformation of the prototype compound structures in order to modulate their functional properties, what can be used searching for new analogues of enzyme inhibitors. In recent years, the method is gaining new opportunities due to high-performance computing as different MD engines got the ability to perform TI or FEP calculations using GPU acceleration. We have tested the method on two experimentally studied enzyme-inhibitor systems: penicillin acylase from *E. coli* (PaEc) and Influenza A virus neuraminidase (NaIAv) and performed computational alchemy calculations for pairs of their known inhibitors. Thus, the validated methodology was then applied to estimate binding of L,D-transpeptidase 2 from *M. tuberculosis* (Ldt_{Mt2}) to the structural analogs of inhibitors previously discovered in our group. We have found that the performance of computational alchemy predictions is highly dependent on properly chosen simulation parameters.

In our case, various combinations of force fields and water models (ff14SB+GAFF+TIP3P and ff19SB+GAFF2+OPC) were used, and at the optimal performance, the difference between the calculated and experimentally determined relative binding energies was less than 1 kcal/mol. Thus, for optimal application to novel enzyme—inhibitor or receptor—ligand systems, we recommend prior adaptation of this method, based on validation of force field combinations, to experimentally determined binding data of structurally related inhibitors/ligands to the molecular target of interest, if available, and then to proceed to the search for new compounds. High-performance computational alchemy can be used as a useful integrative part of the methodology searching for new enzyme inhibitors or receptor ligands and optimizing their structure at drug design.

Acknowledgements

References

1. Bayly, C., Cieplak, P., Cornell, W., Kollman, P.: A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 97, 10269–10280 (1993). https://doi.org/10.1021/j100142a004

2. Beutler, T.C., Mark, A.E., van Schaik, R.C., *et al.*: Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chemical Physics Letters 222(6), 529–539 (Jun 1994). https://doi.org/10.1016/0009-2614(94)00397-1

3. Billones, J.B., Carrillo, M.C.O., Organo, V.G., *et al.*: Toward antituberculosis drugs: In silico screening of synthetic compounds against Mycobacterium tuberculosis l,d-transpeptidase 2. Drug design, development and therapy 10, 1147–1157 (2016). https://doi.org/10.2147/DDDT.S97043

4. Case, D.A., Cheatham III, T.E., Darden, T., *et al.*: The amber biomolecular simulation programs. Journal of computational chemistry 26(16), 1668–1688 (2005). https://doi.org/10.1002/jcc.20290

5. Cordillot, M., Dubée, V., Triboulet, S., *et al.*: In vitro cross-linking of Mycobacterium tuberculosis peptidoglycan by L,D-transpeptidases and inactivation of these enzymes by carbapenems. Antimicrobial agents and chemotherapy 57(12), 5940–5945 (Dec 2013). https://doi.org/10.1128/AAC.01663-13

6. Cornell, W.D., Cieplak, P., Bayly, C.I., *et al.*: A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. Journal of the American Chemical Society 117(19), 5179–5197 (May 1995). https://doi.org/10.1021/ja00124a002

7. He, X., Liu, S., Lee, T.S., *et al.*: Fast, accurate, and reliable protocols for routine calculations of protein–ligand binding affinities in drug design projects using AMBER GPU-TI with ff14SB/GAFF. ACS omega 5(9), 4611–4619 (2020). `https://doi.org/10.1021/acsomega.9b04233`

8. Izadi, S., Anandakrishnan, R., Onufriev, A.V.: Building water models: a different approach. The journal of physical chemistry letters 5(21), 3863–3871 (2014). `https://doi.org/10.1021/jz501780a`

9. Lee, T.S., Cerutti, D.S., Mermelstein, D., *et al.*: GPU-accelerated molecular dynamics and free energy methods in Amber18: performance enhancements and new features. Journal of chemical information and modeling 58(10), 2043–2050 (2018). `https://doi.org/10.1021/acs.jcim.8b00462`

10. Lee, T.S., Hu, Y., Sherborne, B., *et al.*: Toward fast and accurate binding affinity prediction with pmemdGTI: an efficient implementation of GPU-accelerated thermodynamic integration. Journal of chemical theory and computation 13(7), 3077–3084 (2017). `https://doi.org/10.1021/acs.jctc.7b00102`

11. Lehnert, R., Pletz, M., Reuss, A., Schaberg, T.: Antiviral medications in seasonal and pandemic influenza: A systematic review. Deutsches Ärzteblatt International 113(47), 799–807 (2016). `https://doi.org/10.3238/arztebl.2016.0799`

12. Lin, Y.H., Wessén, J., Pal, T., *et al.*: Numerical Techniques for Applications of Analytical Theories to Sequence-Dependent Phase Separations of Intrinsically Disordered Proteins. In: Phase-Separated Biomolecular Condensates: Methods and Protocols, pp. 51–94. Springer US, New York, NY (2023). `https://doi.org/10.1007/978-1-0716-2663-4_3`

13. Ma, Z., Lienhardt, C., McIlleron, H., *et al.*: Global tuberculosis drug development pipeline: The need and the reality. The lancet 375(9731), 2100–2109 (2010). `https://doi.org/10.1016/S0140-6736(10)60359-9`

14. Maier, J.A., Martinez, C., Kasavajhala, K., *et al.*: ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. Journal of chemical theory and computation 11(8), 3696–3713 (2015). `https://doi.org/10.1021/acs.jctc.5b00255`

15. Martelli, G., Pessatti, T.B., Steiner, E.M., *et al.*: N-Thio-β-lactams targeting L,D-transpeptidase-2, with activity against drug-resistant strains of Mycobacterium tuberculosis. Cell Chemical Biology 28(9), 1321–1332.e5 (2021). `https://doi.org/10.1016/j.chembiol.2021.03.008`

16. Rullas, J., Dhar, N., McKinney, J.D., *et al.*: Combinations of β-Lactam Antibiotics Currently in Clinical Trials Are Efficacious in a DHP-I-Deficient Mouse Model of Tuberculosis Infection. Antimicrobial agents and chemotherapy 59(8), 4997–4999 (Aug 2015). `https://doi.org/10.1128/AAC.01063-15`

17. Solapure, S., Dinesh, N., Shandil, R., *et al.*: In vitro and in vivo efficacy of β-lactams against replicating and slowly growing/nonreplicating Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy 57(6), 2506–2510 (Jun 2013). `https://doi.org/10.1128/AAC.00023-13`

18. Steinbrecher, T., Mobley, D.L., Case, D.A.: Nonlinear scaling schemes for Lennard-Jones interactions in free energy calculations. The Journal of chemical physics 127(21), 214108 (2007). `https://doi.org/10.1063/1.2799191`

19. Tian, C., Kasavajhala, K., Belfon, K.A.A., *et al.*: ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. Journal of Chemical Theory and Computation 16(1), 528–552 (Jan 2020). `https://doi.org/10.1021/acs.jctc.9b00591`

20. Tian, C., Kasavajhala, K., Belfon, K.A., *et al.*: ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. Journal of chemical theory and computation 16(1), 528–552 (2019). `https://doi.org/10.1021/acs.jctc.9b00591`

21. Tuckerman, M.E., Martyna, G.J.: Understanding Modern Molecular Dynamics: Techniques and Applications. The Journal of Physical Chemistry B 104(2), 159–178 (Jan 2000). `https://doi.org/10.1021/jp992433y`

22. Voevodin, V.V., Antonov, A.S., Nikitenko, D.A., *et al.*: Supercomputer Lomonosov-2: Large scale, deep monitoring and fine analytics for the user community. Supercomputing Frontiers and Innovations 6(2), 4–11 (2019). `https://doi.org/10.14529/jsfi190201`

23. Von Itzstein, M.: The war against influenza: discovery and development of sialidase inhibitors. Nature reviews Drug discovery 6(12), 967–974 (2007). `https://doi.org/10.1038/nrd2400`

24. Von Itzstein, M., Wu, W.Y., Kok, G.B., *et al.*: Rational design of potent sialidase-based inhibitors of influenza virus replication. Nature 363(6428), 418–423 (1993). `https://doi.org/10.1038/363418a0`

25. Wang, J., Wolf, R.M., Caldwell, J.W., *et al.*: Development and testing of a general amber force field. Journal of Computational Chemistry 25(9), 1157–1174 (Jul 2004). `https://doi.org/10.1002/jcc.20035`

26. Zwanzig, R.W.: High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. The Journal of Chemical Physics 22(8), 1420–1426 (Aug 1954). `https://doi.org/10.1063/1.1740409`