Journal of Molecular Graphics and Modelling 60 (2015) 15–23

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Qgui: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems

Geir Villy Isaksen[a], Tor Arne Heim Andberg[a], Johan Åqvist[b], Bjørn Olav Brandsdal[a,*]

[a] The Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø, N9037 Tromsø, Norway
[b] Department of Cell and Molecular Biology, Biomedical Center, Uppsala University, Box 596, S-75124 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 12 February 2015
Received in revised form 13 May 2015
Accepted 14 May 2015
Available online 21 May 2015

Keywords:
Thermodynamic activation parameters
Free energies
Q
Empirical valence bond
Linear interaction energy
Molecular dynamics
Free energy perturbation

ABSTRACT

Structural information and activity data has increased rapidly for many protein targets during the last decades. In this paper, we present a high-throughput interface (Qgui) for automated free energy and empirical valence bond (EVB) calculations that use molecular dynamics (MD) simulations for conformational sampling. Applications to ligand binding using both the linear interaction energy (LIE) method and the free energy perturbation (FEP) technique are given using the estrogen receptor (ERα) as a model system. Examples of free energy profiles obtained using the EVB method for the rate-limiting step of the enzymatic reaction catalyzed by trypsin are also shown. In addition, we present calculation of high-precision Arrhenius plots to obtain the thermodynamic activation enthalpy and entropy with Qgui from running a large number of EVB simulations.

1. Introduction

Computer modeling and simulations have now become important tools to study complex systems in many areas of chemistry and biology. With the tremendous increase in speed and availability of high performance computers it is nowadays possible to perform quantitative calculations on highly complex systems such as the ribosome [1,2] and virus capsids [3–5]. Biochemical topics where computational modeling is frequently used to obtain novel insights include among others ligand binding and design, folding of proteins, enzyme catalysis and protein-membrane interactions. The ability to accurately calculate free energies enables us to characterize the structure and energetics of molecular complexes, and is often the key to understand many biological functions. In this respect, it is especially worth emphasizing that energetics often provides the most important and useful link between structure and function of biomolecular systems.

Molecular dynamics simulations represent an efficient way to sample the thermally accessible region of conformational space with microscopic models of molecular systems. Together with the potential energy (given by the force field), the ensemble of sampled structures form the basis for further calculations of, for example, free energies. Thermodynamic and kinetic experiments provide valuable data in terms of binding free energies, solvation energies and activation free energies, which are directly comparable to results obtained with simulations techniques. It thus becomes possible not only to verify and validate calculated properties with experiments, but also to make predictions that can be examined with experiments.

Historically, classical MD simulations to calculate free energies have been a low-throughput technique. This picture is now changing mainly due to advancements in algorithms and the scaling performance of software together with the increase in computer power witnessed during the last decade. Several different methods to compute free energies that rely on MD for conformational sampling are presently available [6–11]. These methods are generally more complicated to apply when compared to for example simplified ligand docking and scoring approaches, especially when the ligands are structurally dissimilar [12,13]. Computational tools that simplify the use of conformational sampling techniques and make the entire process of computing free energies more efficient are therefore desirable. Graphical user interfaces have the potential not only to lower the barrier for new users, but also of making the process of calculating free energies more efficient.

* Corresponding author. Tel.: +47 77 64 40 57; fax: +47 77 64 47 65.
E-mail address: bjorn-olav.brandsdal@uit.no (B.O. Brandsdal).

http://dx.doi.org/10.1016/j.jmgm.2015.05.007
1093-3263/

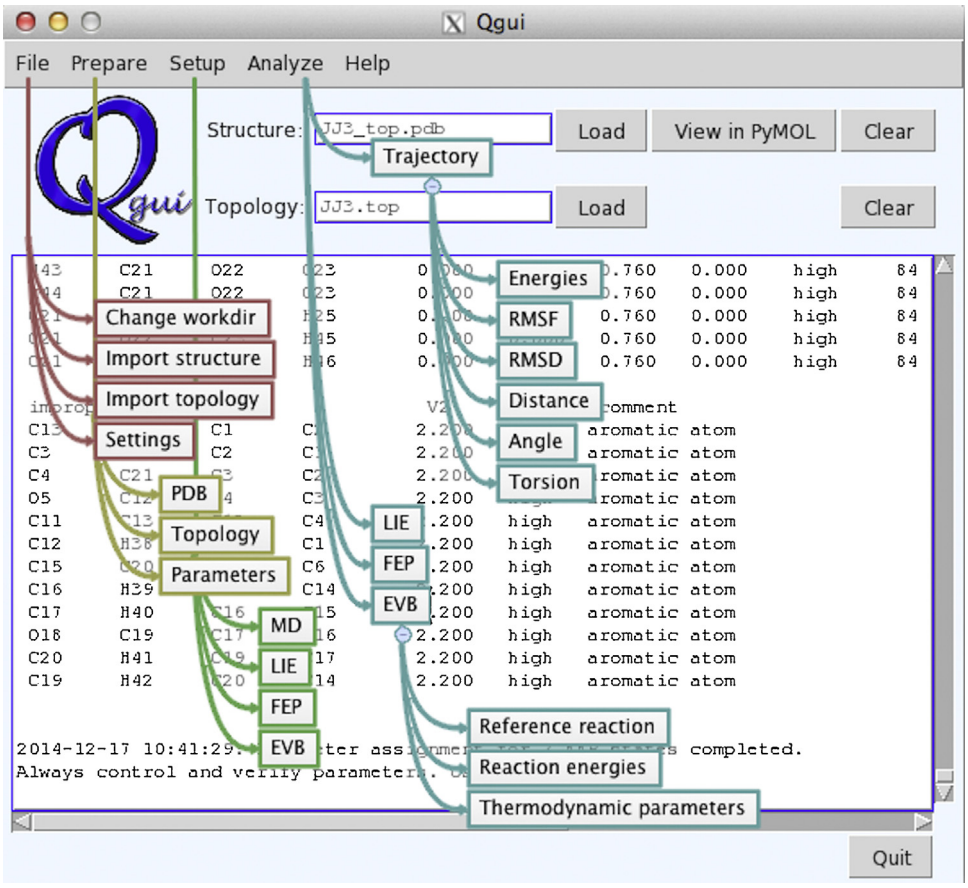

Fig. 1. The main window of the graphical user interface to Q (Qgui) with illustrated overview of dropdown menus and their contents.

Another issue related to computations of free energies is how to sample phase space most efficiently, and to ensure that the relevant parts are really captured by the simulations. It has become standard practice to conduct several replicas of the simulations to ensure that statistically significant results with adequate sampling of the relevant parts of phase space are achieved. This is typically done by repeating the simulations with different initial conditions, which may be done by assigning different initial velocities or different starting coordinates when running the MD simulations. The main challenge with this procedure is management and handling of the enormous amount of data generated.

Historically, most studies have focused on how to predict and calculate the Gibbs free energy (ΔG), as direct calculation of enthalpies and entropies involves large energy contributions. While the free energy typically converges within reasonable simulation time, obtaining converged enthalpic and entropic contributions pose a major computational challenge. Nonetheless, knowledge about the enthalpic and entropic contributions is desirable and provides important chemical information about molecular systems. We have recently shown that not only the activation free energy, but also the activation enthalpy and entropy can be calculated for enzymatic reactions using the EVB approach [14,15]. Of particular relevance here is determination of high precision computational Arrhenius plots in order to obtain thermodynamic activation parameters [16,17]. This was recently illustrated by Isaksen et al. [17] with more than 3000 independent simulations yielding a total simulation time of ~2 μs. When high precision is needed, sufficient sampling becomes critical. The development of a graphical user interface was initially driven by our need for an efficient way of setting up free energy simulations and to analyze large amounts of data produced using the MD program Q [18]. Obtaining reliable and converged reaction free energies typically involves averaging over several hundreds of MD trajectories. Furthermore, in order to obtain reliable thermodynamic activation parameters it is necessary to run the simulations at several temperatures (typically 6–10) over a suitable range for the enzyme of interest. This procedure is of course not limited to enzymatic reactions, and can easily be applied to van't Hoff plots and ligand binding.

Here we present an overview of the newly developed graphical user interface, Qgui. Examples of how it is used to build molecular topologies for MD simulations and to compute ligand binding free energies with the LIE [7,19] and FEP [6,20] method are given. We also show how Qgui enables efficient studies of chemical reactions and to obtain thermodynamics activation parameters for enzymatic reactions. It should be noted that Qgui is not limited to Q, and can in principle be extended to function with other MD software packages.

2. Methods

2.1. Qgui—Graphical user interface to Q

Qgui has been written in the Python language using object oriented design principles based on the TkInter GUI library (https://docs.python.org/2/library/tkinter.html). The Qgui consists of a main window, which provides feedback to the user, and four dropdown menus, as illustrated in Fig. 1. The Qgui main window (Fig. 1) is simplistic and mainly functions as a feedback/logging window for all the different tools available. PDB files can be imported directly from the RCSB Protein Data Bank (http://www.pdb.org). To visualize and manipulate structures, Qgui communicates with any

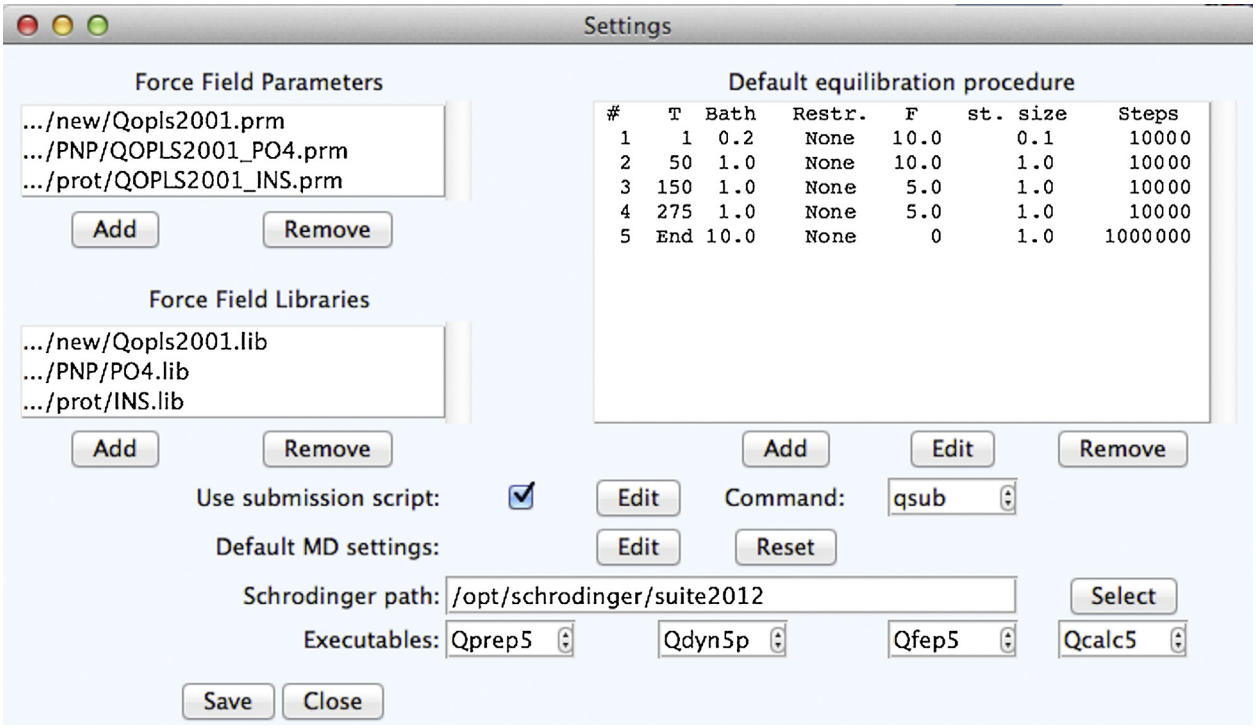

Fig. 2. The Qgui global settings panel that allows user specific flexibility.

version of the Open Source PyMOL (Schrödinger, LCC, New York), which is freely available. The File menu contains the settings option, which is normally accessed prior to the topology creation and setup of MD simulations. The Qgui settings (Fig. 2) have several useful global adjustments that can be modified according to each users personal preference. Force field parameter and library files specified in the settings menu are read in the topology building process and can be added or removed in the Qgui settings. There is no limitation on the number of files to include. Options to set up and edit the default equilibration procedure that is launched before any MD simulations are also available. If jobs are submitted on a cluster with a queuing system, the queuing commands can be prepared with the 'Use submission script', which also includes the cluster specific command for submitting jobs to the queue (Fig. 2). The settings of Qgui can be accessed and changed at anytime during the topology building session and setup process of MD simulations.

2.2. Linear interaction energy calculations

The linear interaction energy method relies on conformational sampling using either MD or Monte Carlo simulations of the ligand and the complex to estimate the absolute free energy of binding. Based on the simulations, the binding free energy is calculated according to the LIE approximation [7,19]:

$$\Delta G_{\text{bind}} = \alpha \Delta \left\langle U_{l-s}^{\text{vdw}} \right\rangle + \beta \Delta \left\langle U_{l-s}^{\text{el}} \right\rangle + \gamma \tag{1}$$

where $\langle\rangle$ denotes the MD averages of electrostatic (el) and van der Waals (vdw) interactions between the ligand and its surroundings (l–s), i.e., the solvated receptor binding site or just solvent. The Δ's denotes the difference between the averages in bound and free states. The parameters α and β are weight coefficients for the non-polar and polar binding energies, respectively. One may also choose to use different coefficients for the bound and free state, as illustrated by Sund et al. [21]. The last parameter, γ, is an additional constant which may be needed to reproduce absolute experimental binding energies. It should be noted that the additional constant γ is not needed when computing relative binding free energies.

Standard errors for the LIE energy terms are computed according to the method presented by Allen and Tildesley [22] based on the concept of statistical inefficiency. The method is a general scheme to compensate for correlated data that can be considered independent. This involves partitioning the data into consecutive blocks with progressively larger and larger lengths (t_b), taking the average for each block and calculating the variance of these averages (σ_b^2). The block lengths are finally extrapolated to infinite block lengths and the statistical inefficiency s is then defined by

$$s = \lim_{t_b \to \infty} \frac{t_b \sigma_b^2}{\sigma^2} \tag{2}$$

where t_b is the block length, σ_b^2 is the variance of a series of block averaged values and σ^2 the variance of the entire series of sampled interaction energies. Once the statistical inefficiency is given, the standard error (SE) is computed as:

$$\text{SE} = \frac{\sigma}{\sqrt{N/s}} \tag{3}$$

where N is the total number of points and $\sqrt{N/s}$ is the number of independent data points. In addition, the time development for each energy term can be plotted separately or together, which is useful to ensure convergence and stability of the simulations.

We have implemented an automatic scheme for generating predictive LIE models by combining available experimental binding free energies with the MD energy terms. The automatic model generation first finds the sum of squared errors (SSE) between the computed and experimental binding free energies as a function of the selected parameter. This is done by mapping the selected parameter until sufficient information exist to compute the polynomial regression of the form:

$$\text{SSE} = b_2 x^2 + b_1 x + b, \qquad x \in [\alpha, \beta, \gamma] \tag{4}$$

Then the parameter value minimizing SSE is found by solving the partial derivative of SSE with respect to the selected parameter

$$\frac{\delta \mathrm{SSE}}{\delta x} = 0, \qquad x \in [\alpha, \beta, \gamma] \tag{5}$$

which has the general solution

$$x_{\min(\mathrm{SSE})} = \frac{-b_1}{2b_2}, \qquad x \in [\alpha, \beta, \gamma] \tag{6}$$

The selected parameter value obtained from this approach will ensure that the SSE is at its minimum.

2.3. Free energy perturbation

The free energy perturbation (FEP) method can be used for alchemical transformations, and is most commonly applied to two state situations driving a system from ϕ_1 to ϕ_2 [6,20]. The method is, however, not limited to two-state cases, and three- and four-state treatments can also be used (see e.g. Ref. [23]). For simplicity, we will here show how the two-state FEP is implemented. In the FEP scheme an alchemical transformation of a system from ϕ_1 to ϕ_2 is driven by a linear combination of the two states via a set of intermediate mapping potentials:

$$\varepsilon_m = \lambda_m \varepsilon_1 + (1 - \lambda_m)\varepsilon_2; \quad \lambda_m \in (0, 1) \tag{7}$$

Here ε_m denotes the effective potential energy function of a particular FEP window where the coupling parameter λ_m is stepwise decremented from 1 to 0. The free energy difference associated with the perturbation is calculated using the Zwanzig's formula [24]

$$\Delta G_m = \beta^{-1} \sum_{m=0}^{n-1} \ln \left\langle \exp\left[-\beta(\varepsilon_{m+1} - \varepsilon_m)\right]_m \right\rangle \tag{8}$$

where $\beta = 1/kT$, n is the number of intermediate states and $\langle \ldots \rangle_m$ denotes the average on the mapping surface ε_m. Future releases of Qgui will also contain alternatives to Zwanzig's exponential formula, as implemented in the Q program [18].

2.4. Empirical valence bond simulations

The EVB method [14,15] computes the free energy of a chemical reaction by calculating the adiabatic ground-state energy (E_g) and the corresponding eigenvector (C_g) at each sampled conformation by solving the corresponding secular equation

$$\mathrm{H}C_g = E_g C_g \tag{9}$$

The energetics associated with each resonance form is determined by a specific potential energy function, ε_i, describing the bonding arrangement and the charge distribution for each structure.

$$H_{ii} = \varepsilon_i = U^i_{\mathrm{bnd}} + U^i_{\mathrm{ang}} + U^i_{\mathrm{tor}} + U^i_{\mathrm{imp}} + U^i_{nb,rr} + U^i_{nb,rs} + U^i_{ss} + \alpha^i \tag{10}$$

here the subscripts bnd, ang, tor, imp and nb are abbreviations for bond, angle, torsion, improper and non-bonded, whereas r and s denotes the reacting fragments and the surroundings, respectively. The last term of the Hamiltonian, α^i, represents the intrinsic gas-phase energy of the given resonance structure with all fragments at infinite separation [14,15]. The off-diagonal matrix elements, H_{ij}, representing the adiabatic mixing of the VB states and are typically described by a constant or an exponential function [14,15,25]. The general solution of Eq. (9) for a 2-state reaction is

$$E_g = c_1^2 H_{11} + c_2^2 H_{22} + 2c_1 c_2 H_{12} \tag{11}$$

Calculation of the free energies follows the free energy perturbation (FEP) scheme as described above (Eqs. (7) and (8)). The free energy profile corresponding to trajectories moving on the actual ground state potential is further computed from the umbrella sampling expression

$$\Delta G(\Delta\varepsilon_n) = \frac{\sum_{m \supset \Delta\varepsilon_n} w_m \left(\Delta G_m - RT \ln \left\langle \exp - \left[(E_g(\Delta\varepsilon_n) - \varepsilon_m(\Delta\varepsilon_n))/RT \right] \right\rangle_m \right)}{\sum_{m \supset \Delta\varepsilon_n} w_m} \tag{12}$$

where $\Delta\varepsilon_n$ is the discretized reaction coordinate and $w_m / \sum w_m$ ensures that the different contributing vectors to the reaction coordinate interval are weighted proportionally to the total contribution to the interval. As mentioned, three- and four-state models can also be used if desired.

A prerequisite for any enzyme free energy calculations is that the free energy surface of a suitable reference reaction is fitted to available ab initio or experimental data. This is typically done by adjusting the intrinsic gas phase shift, α_i, and the adiabatic mixing of the states, H_{ij}, until the desired activation and reaction free energies are reproduced. We have developed a procedure for automatic 2-state EVB reference reaction calibration (RRC). The RRC scheme is divided into 3 stages. The initial stage involves finding the correct sign for α_i. This is simply done by checking the reaction profile with a positive and a negative α_i value, where the correct sign should give the reactant state $\Delta G = 0$ as illustrated in Fig. S1-1. After finding the correct α_i sign the scheme moves on to part 2, which quickly locates the reaction profile within the target activation and reaction energies. Step 2 in the RRC scheme takes advantage of the close to linear relationship between α_i and H_{ij} against ΔG_0 and $\Delta G^{\ddagger}$, respectively. First a linear regression of $\Delta G^{\ddagger}$ as a function of H_{ij} is performed (Fig. S1-2A):

$$\Delta G^{\ddagger}\left(H_{ij}\right) = \beta_0 + \beta_1 H_{ij} \tag{13}$$

Then a linear regression is computed with ΔG_0 as a function of α_i with the new H_{ij} value kept constant (Fig. S4-2B):

$$\Delta G_0(\alpha_i) = \beta_0 + \beta_1 \alpha_i \tag{14}$$

The reaction profile now looks like it should, but the computed values are not in close agreement to the target values. This is due to the fact that the linear approximation is only good when close to the target values. Repeating the previous linear regression steps once more (Fig. S1-2C and S1-2D) usually brings the activation and reaction energies very close to the desired target values. Finally, in step 3 of the RRC scheme H_{ij} and α_i are fine-tuned in an iterative manner until the target values are reached (Fig. S1-3). The step size used for each iteration is controlled by the size of $\Delta\Delta G$ (difference between the target and current free energies). The RRC is considered converged when $\Delta\Delta G$ in part 3 is less then a certain threshold. This threshold is by default 0.2 kcal/mol, but will automatically be increased gradually if RRC part 3 for some reason struggles to find the target free energies. Calibrating a reference reaction with around 18 000 atoms and 40 EVB atoms using RRC takes ~10 s. It should be noted that we have currently restricted the RRC scheme to consider H_{ij} as a constant, but more complex functions (e.g. exponential or Gaussian) could easily be implemented. Furthermore, while Qgui automatically parameterizes the H_{ij} and α_i for two-state EVB, three- and four-state models must be manually parameterized as these are less amenable to automated parameterization.

Qgui efficiently computes the reaction free energy profiles, automatically locating the transition and product state free energies. In addition, the breakdown of the EVB potential activation energy terms given in Eq. (10), U, are computed and scaled accordingly to both the eigenvector coefficients, c_i, in Eq. (11) and the different lambda window contributions to the reactant and transition state. That is, when computing the reaction free energy profile as a function of the energy gap ($\Delta\varepsilon$) as the reaction coordinate, several intermediate mapping potentials typically contributes to the

same energy gap. Qgui calculates the exact EVB potential activation energy terms by collecting all lambda values contributing to the transition and reactant state. The energies originating from each lambda value are then scaled by w_i according to the actual state contribution so that

$$U = \sum_i w_i \left(U_{1,i} + U_{2,i}\right); \quad i \in \lambda_m \tag{15}$$

where $\sum w_i = 1$.

In addition, the total potential activation energy within the reacting region and to the surroundings, $\Delta U^{\ddagger}_{rr+rs}$, is computed exactly as

$$\Delta U^{\ddagger}_{rr+rs} = \Delta\left(c_1^2 U_{1,rr+rs}\right)^{\ddagger} + \Delta\left(c_2^2 U_{2,rr+rs}\right)^{\ddagger} + \Delta\left(c_2^2 \alpha_2\right)^{\ddagger} + 2\Delta(c_1 c_2 H_{12})^{\ddagger} \tag{16}$$

where all terms are scaled according to Eq. (15) and $U_{rr+rs} = U_{rr} + U_{rs}$ in Eq. (10). This term is particularly useful for evaluating the interactions within the surroundings, U_{ss}, which is not obtained directly from the EVB calculations (suffers from slow convergence), but rather obtained through the simple relationship (neglecting the $P\Delta V$ term)

$$\Delta H^{\ddagger} = \Delta U^{\ddagger}_{rr} + \Delta U^{\ddagger}_{rs} + \Delta U^{\ddagger}_{ss} \tag{17}$$

Qgui is written with a particular focus on automating the task of obtaining thermodynamic activation parameters ($\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$). We have implemented specialized setup and analyze tools automatically generating and recognizing directory hierarchies consisting of a range of temperatures with a series of subdirectories containing all the EVB runs per temperature. The thermodynamics is simply obtained by computing the activation free energy as a function of the inversed temperature. The resulting Arrhenius plots are then computed with standard linear regression methods.

3. Results and discussion

Before any simulations can be carried out a three-dimensional model of the system must be available. Typically, atomic level models of the molecular systems are obtained from X-ray crystallography, NMR spectroscopy or homology modeling. Software packages for classical MD simulations of biological molecules contain the necessary force field parameters and topologies to describe standard amino acid residues, DNA and RNA. However, in many cases the system of interest contains non-standard residues, ligands, substrate etc., which must be properly parameterized before running simulations. This process can be very tedious and time-consuming, and automatic procedures are desirable to make the parameterization process more efficient and consistent. We will now first present the overall process of building the topology and how to set up and run MD simulations using Qgui, and then illustrate how Qgui makes more advanced free energy calculations more efficient and automated.

3.1. Creating topologies and running molecular dynamics simulations

The Topology Prepare tool makes it efficient to prepare any system for MD simulations, and the Qgui main window (Fig. 1) is used for feedback to the user. Qgui presently only supports spherical boundary conditions, but periodic boundary can be added. The simulation sphere center is automatically set to the center of the structure loaded and the radius of the sphere is by default 10 Å greater than the radius if the molecular system. The PDB file is further automatically checked against the force field parameter and library files defined in the Qgui settings for missing parameters. If any parameters are missing, an automated scheme is available to generate force field parameters and library files. This scheme relies on the ffld_server utility that comes with Maestro (Schrödinger, LCC, New York), and currently only supports OPLS-2001 and OPLS-2005 [26,27]. However, the assignment scheme can be extended to cover other force fields, but at present parameters for other force fields must be prepared manually before generating the topology.

As has been illustrated previously, proper treatment of electrostatic interactions is critical when calculating binding free energies for charged ligands to avoid potential Born-terms to enter into the free energy [28]. The charged state for ionizable amino acids (e.g. LYS, ARG, GLU, ASP and HIS) can be turned on or off (i.e. charged or neutral) by selecting them from the list in the Topology prepare window (Fig. S2). In addition, the distance from the charged group to the selected simulation center is given, which is convenient when deciding to neutralize charges close to the simulation boundary. For simulations using spherical boundary conditions, only charged residues that are properly solvated, that are well within the simulation boundary (~80%), are kept charged and residues close to and outside the boundary are kept neutral. N- and C-terminals are also automatically detected, and the charge of these can be toggled in the same manner as described above. The effect of neutralizing distant charges can be corrected for after the calculations are finished [29]. The molecular system can of course be solvated, and the desired water model is also selected in the Topology prepare window. Disulfide bridges are also automatically detected. After successfully generating a topology, both the topology and the corresponding PDB file will automatically load in the Qgui main window (Fig. 1). The molecular system can also be visualized and examined using PyMOL (Schrödinger, LCC, New York).

Once the topology for the system has been successfully generated, the options for controlling the MD simulations can be set. Heating and equilibration of the system are controlled with the settings menu, and can be customized according to the preferences of the users. The default settings heats the system stepwise to the target temperature before it is equilibrated prior to the production run. The MD setup panel allows users to change parameters that control the production part of the simulations, which typically include simulations time, stepsize, non-bonded cutoffs, recording intervals of energies, structures and output frequency among others.

3.2. Linear interaction energy calculations

The LIE method [19] enables the absolute binding free energy to be determined by running thermal sampling of the bound and unbound state of the ligand. Based on the simulations of the physical relevant states, the difference in the interaction energies between the ligand and its surroundings are scaled according to the LIE equation (Eq. (1)), and thus yielding the free energy of binding.

Qgui contains tools for both preparing and analyzing LIE from MD simulations. Preparing MD simulations for LIE calculations through the Setup LIE tool only requires a single PDB file with the complex (ligand and receptor). The ligand can be selected and extracted from the PDB file of the complex and topologies for both can be generated directly. After configuring the desired MD settings, Setup LIE allows the user to simultaneously submit an unlimited number of unique jobs by adjusting the 'Runs' (Fig. S3). This is achieved using different initial velocities.

The average energies with corresponding standard errors are presented in the Analyze LIE window. The time development for the different potential energy terms can be plotted and visualized all at once or individually. Computed ΔG_{bind} are saved together with corresponding experimental values and using the Fit LIE parameters tool generates the predictive binding free energy models (Fig. S4). The Fit LIE parameters tool visualizes the computed and

Table 1
Binding free energies (kcal/mol) obtained using the LIE method for the estrogen receptor.

R-group[a]	$\langle U_{l-s}^{vdW} \rangle_p$	$\langle U_{l-s}^{el} \rangle_p$	$\langle U_{l-s}^{vdW} \rangle_w$	$\langle U_{l-s}^{el} \rangle_w$	γ	ΔG_{bind}^{LIE} [b]	ΔG_{exp} [c]
H	−38.7 ± 0.3	−28.8 ± 0.3	−22.5 ± 1.2	−34.5 ± 0.1	−11.0	−12.1	−11.8
Me	−42.9 ± 0.4	−25.8 ± 0.3	−23.7 ± 1.0	−35.2 ± 0.1	−11.0	−11.4	−11.7
CH_2OMe	−45.7 ± 0.4	−26.9 ± 0.3	−25.6 ± 1.3	−38.8 ± 0.1	−11.0	−10.7	−10.9
CH_2OEt	−47.7 ± 0.2	−25.6 ± 0.3	−27.1 ± 0.8	−38.7 ± 0.1	−11.0	−10.4	−10.3
CH_2OH	−41.4 ± 0.3	−32.7 ± 0.3	−22.0 ± 0.7	−47.8 ± 0.1	−11.0	−9.5	−9.3

Error bars are standard error of the mean from five independent simulations.
[a] Functional groups are attached in the position highlighted in Fig. 3.
[b] $\alpha = 0.18$ and $\beta = 0.33$.
[c] Experimental energies from Ref. [30].

experimental data in real time together with the corresponding linear regression lines. For original LIE models only the γ parameter is optimized to enhance the binding free energy correlations, but the Fit LIE tool offers the option to also optimize α and β, which in some cases can be useful to generate predictive models. The models can be generated by manually or automatically adjusting the parameters. It is however useful to visually compare the regression lines for the computed and experimental binding energies before trusting the predictive ability of a model based on the minimized SSE. This is easily achieved with the Fit LIE parameters tool (Fig. S4). Qgui is not restricted to binding free energies, and it is for example possible to automate calculations of absolute solvation energies.

As an example of binding energy calculations using the LIE method with Qgui, we present results for binding of a series of five ligands to the estrogen receptor α (Table 1). Atomic coordinates for ligand 2 in complex with the estrogen receptor α were obtained from the protein data bank entry 2QE4 [30]. The remaining ligands were constructed based on this template assuming similar binding mode, as illustrated in Fig. 3. A 25 Å spherical system centered on the ligand was constructed using Qgui, and described using OPLS-AA 2001 [26]. Ionic residues beyond 22 Å from the ligand were treated as neutral yielding a total net charge of -3 within the simulation sphere. The systems were subjected to a stepwise heating protocol where heavy atoms were restrained and then a short unrestrained equilibration at 300 K. The production phase was run at 300 K for 12.5 ns using a 1 fs time step, during which ligand-surrounding energies were sampled every 10 fs. Five independent simulations were run for each ligand yielding a simulation time of 62.5 ns for the bound and unbound state, and the average energies are presented in Table 1. A γ parameter was optimized with Qgui and the results are presented in Table 1. Fig. 4 show the data unfitted and fitted to the experimental data using the procedure outlined in Section 2. It is also clear that the predictive power of the LIE method obtained here will become weaker when dealing with ligands that are outside the binding range used here with more than ~4 kcal/mol.

3.3. Efficient setup of FEP and EVB simulations

Qgui offers a new concept for efficient FEP and EVB simulations, and currently supports up to four-state FEP/EVB. Both FEP and EVB require that the changing force field terms are explicitly defined for each state, and the terms for each are manually defined in the FEP file. Qgui makes this process much easier and more efficient through automatic parameter and charge assignment and automatic mapping of the corresponding parameter numbers to the correct atom numbers in the topology. This is particularly useful as it opens up for visualization of the entire FEP/EVB process with PyMOL (e.g. atoms, bonds, angles, torsions and impropers for the various states). Qgui communicates with PyMOL so that bonds formed or broken in EVB or changing atoms (or disappearing/appearing) in FEP can be defined by clicking (Fig. 5).

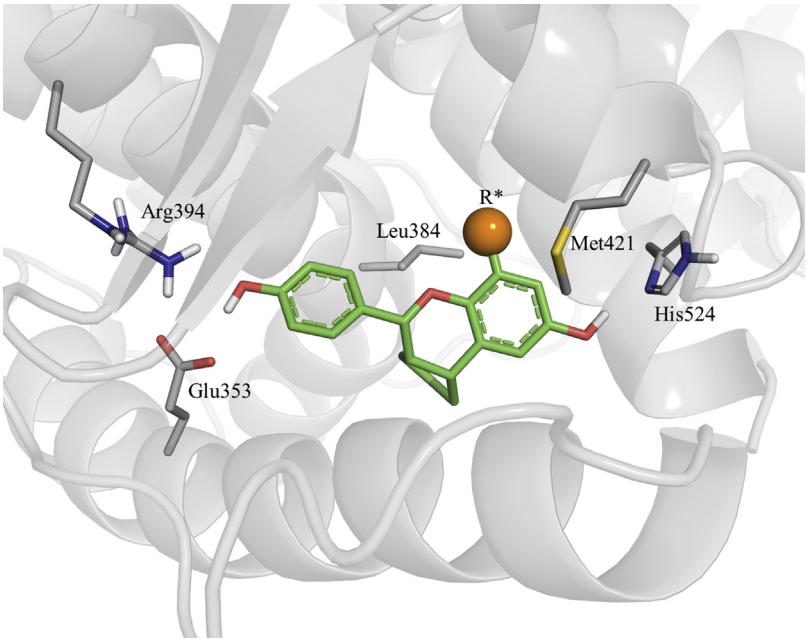

Fig. 3. The ligand scaffold in the active site of the estrogen receptor used for LIE and FEP calculations. The site of site for chemical modifications of the ligand is highlighted with an organge sphere.

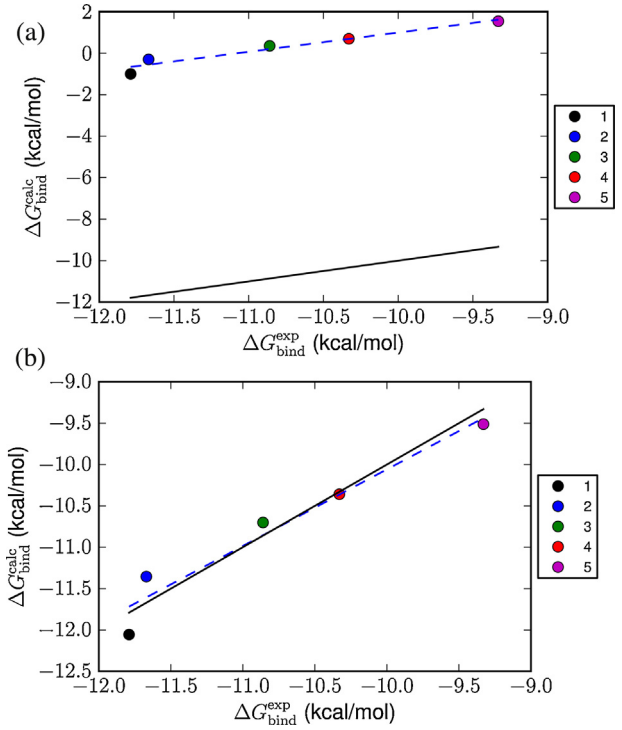

Fig. 4. Fitting of the LIE binding data for five ligands bound to the estrogen receptor with (a) unfitted data points and (b) data fitted to experimental values.

3.4. FEP application: Relative ligand binding free energies

The FEP simulations were carried out on the estrogen receptor α and series of ligands. As with the LIE set up the starting point was the 2QE4 protein data bank entry [30], and the ligands were built manually into the binding site assuming similar conformations. Qgui was used to set up a forward two state FEP simulation, and OPLS-2005 force field parameters [27] for state two was generated automatically. The simulations included a slow stepwise heating with restraints on heavy atoms, followed by a 100 ps unrestrained equilibration at 300 K. The perturbation is further divided into three stages, where the first involves turning "dummy atoms" into soft-core Lennard-Jones potentials, followed by changing the electrostatic interactions and finally changing from soft-core to Lennard-Jones potentials. Each of these stages used a 55 λ-step perturbation with a sigmoidal distribution was used (denser spacing near the end-points), and the simulation was run at 300 K using a 1 fs time step. Each λ-state was simulated for 10 ps of which the first 0.5 ps were discarded for equilibration and the energy was sampled every fifth femtosecond. 5 parallels were run for each system and the relative binding free energies were calculated and are presented in Table 2.

Table 2
Relative binding free energies (kcal/mol) obtained with the FEP method for the estrogen receptor.

Modification[a]	$\Delta\Delta G_{FEP1}$	$\Delta\Delta G_{FEP2}$	$\Delta\Delta G_{FEP3}$	$\Delta\Delta G^{FEP}_{bind}$ [b]	$\Delta\Delta G^{exp}_{bind}$ [c]
$H \rightarrow CH_3$	0.1 ± 0.1	0.3 ± 0.1	-1.1 ± 0.2	−0.7	+0.1
$OH \rightarrow OCH_3$	-0.2 ± 0.1	-1.9 ± 0.2	1.4 ± 0.3	−0.7	−1.6
$OCH_3 \rightarrow OCH_2CH_3$	0.0 ± 0.1	0.0 ± 0.1	-0.2 ± 0.1	−0.2	+0.6

Error bars are standard error of the mean from five independent simulations.
[a] Functional groups are attached in the position highlighted in Fig. 3.
[b] $\Delta\Delta G^{FEP}_{bind}$ (bound-free state) is obtained through three stages where stage 1 corresponds to turning on soft-core Lennard–Jones potential, stage 2 changing of electrostatics and stage 3 corresponds to transforming soft-core to Lennard–Jones potential.
[c] Experimental energies from Ref. [30].

3.5. Automated reference reaction calibrations

A key criterion for doing successful reaction dynamics with EVB is that a free energy surface for a suitable reference reaction must be fitted to existing experimental or ab initio data. In practice, this typically involves manually adjusting α_i and H_{ij} until the desired activation and reaction free energies are reproduced. The resulting α_i and H_{ij} parameters often vary between multiple reference reaction calibrations. Variations in these parameters will of course give different activation and reaction free energies in the complex. A more reliable approach is to calibrate several reference reactions and use the average parameters for the enzyme reaction. This

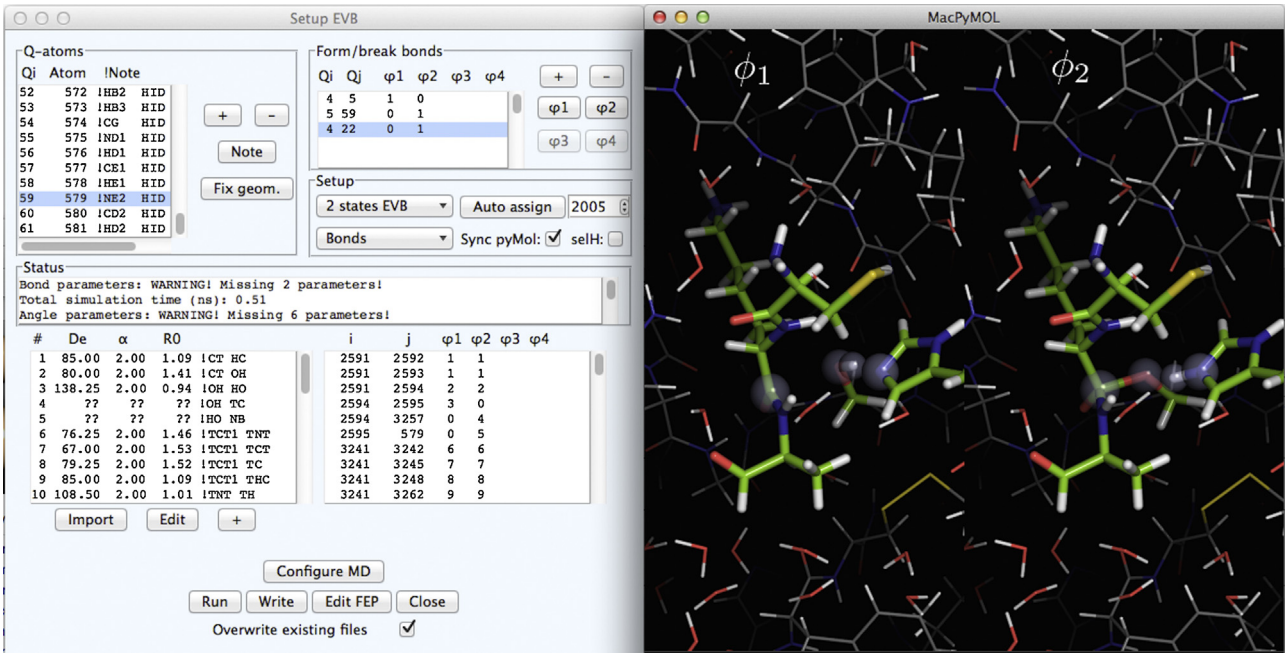

Fig. 5. EVB setup illustrated with sync PyMOL for the tetrahedral intermediate formation in trypsin. The reacting region (Q atoms) is automatically highlighted in green color.

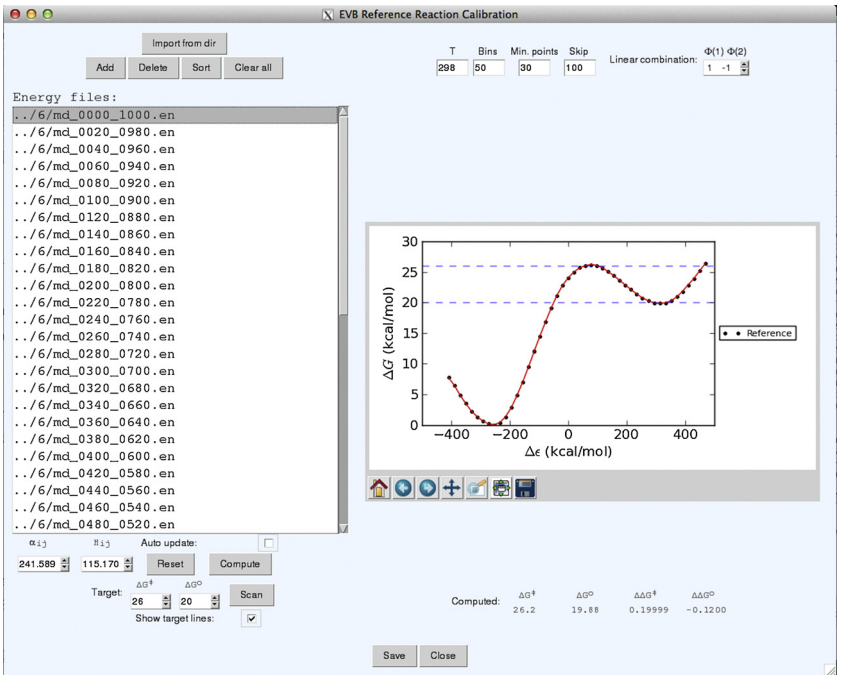

Fig. 6. EVB reference reaction calibration (RRC) window. The blue dotted lines represent the target activation and reaction energies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

is, however, a tedious and time-consuming task when manually calibrating each of the simulations. To cope with this, Qgui can automatically calibrate 2-state EVB reference reactions through the EVB reference reaction calibration (RRC) tool (Fig. 6) available from the Analyze menu. It is possible to calibrate single runs individually or several runs merged together. In principle, the only required input here is the target activation and reaction free energies. In addition, the user may edit other settings such as the linear combination of states, the number of bins and the minimum number of points in each bin.

3.6. *High-throughput analysis of comprehensive EVB simulations*

One of Qgui's significantly new contributions is the ability to efficiently compute and analyze reaction free energy profiles, potential activation energies and thermodynamic activation parameters from the EVB method. Qgui is designed to recognize typical directory patterns for multiple temperature simulations with several runs at each temperature. The interface thus makes it very easy to analyze and compare reaction profiles, different EVB potential activation energy terms, and thermodynamic activation parameters between individual runs, but also between projects/species.

3.7. *EVB application: Rate-limiting step of trypsin*

Atomic coordinates for atlantic salmon trypsin (AST) with PDB entry 1BZX [31] was imported directly into Qgui from the protein data bank. The bovine trypsin pancreatic trypsin inhibitor (BPTI) was chopped to only include the central tripeptide CYS-LYS-ALA (CKA) with the prepare PDB tool. In addition HIS57 and SER195 were chopped at the Cβ position in a separate round to generate coordinates for the reference reaction. PyMOL was used to add hydrogen atoms to the C- and N-terminal of CKA and to the Cβ of the chopped HIS57 and SER195. Parameter and library files for CKA and the chopped residues were then generated with the Prepare Parameters with the OPLS-2005 [27] force field selected. Topologies for the complex and the reference reaction were prepared with the Prepare Topology tool using a 35 Å simulation

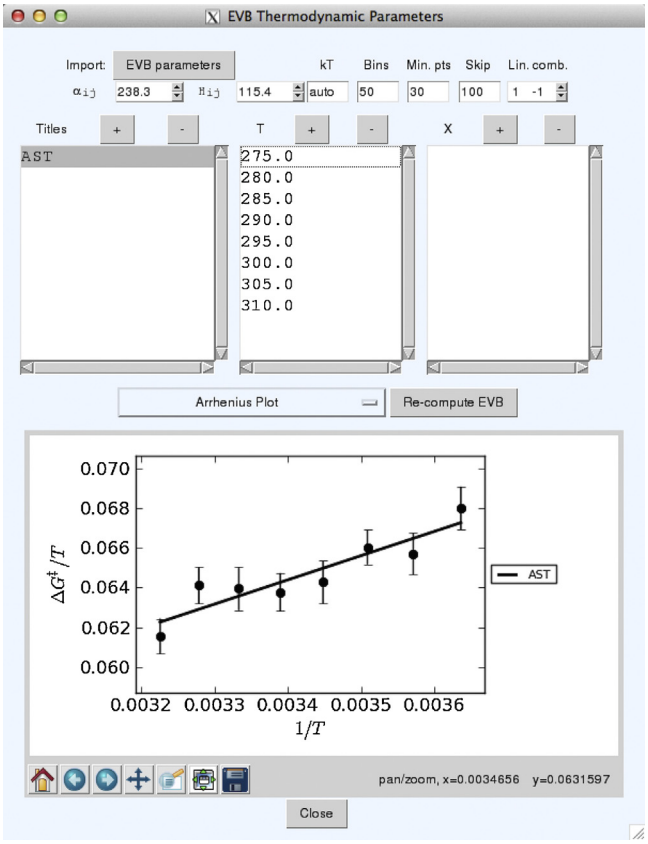

Fig. 7. Arrhenius plot generated using the EVB Thermodynamic Parameters tool in Qgui for Atlantic salmon trypsin with $R^2 = 0.8$.

sphere solvated with TIP3P [32] water molecules. A 2-state EVB simulation describing the tetrahedral intermediate formation step was prepared using the Setup EVB tool (Fig. 5). The entire CKA was selected as EVB atoms in addition to the modified HIS and SER

residues. The 'Sync PyMol' function was used to quickly define the proton transfer from SER195 to HIS57 together with the concerted nucleophilic attack by SER195 to the carbonyl carbon of LYS in CKA as illustrated in Fig. 5. Parameters for the EVB atoms in state 1 and 2 were then auto-assigned with the built in tool communicating with the ffld_server (Schrödinger, LCC, New York). Additional MD/EVB settings were assign as described in ref. [17]. A total of 10 reference reactions were calibrated so that the free energy surface coincided with $\Delta G^{\ddagger} = 26$ kcal/mol and $\Delta G = 20$ kcal/mol (see ref. [17]) with the Qgui EVB RRC tool (Fig. 6). The resulting average α_2 and H_{12} from these runs are 238.3 ± 0.7 kcal/mol and 115.4 ± 0.5 kcal/mol, respectively. A total of 8 temperatures in the range 275–310 were submitted with 50 runs at each temperature resulting in a total simulation time of 204 ns for the complex.

At 300 K an average activation free energy of 19.2 ± 0.3 kcal/mol was computed in the enzyme. Even though we here included ~2 times as many EVB atoms, these computations are in excellent agreement with our previous results [17]. From the computed Arrhenius plot illustrated in Fig. 7, a $\Delta H^{\ddagger}$ and $T\Delta S^{\ddagger}$ of 12.2 ± 0.8 kcal/mol and -6.7 kcal/mol, respectively, was obtained at 300 K. Utilizing the Analyze EVB Reaction Energies tool (Fig. S5), $\Delta U^{\ddagger}_{rr+rs}$ is calculated to 7.0 ± 1.9 kcal/mol, which is in very good agreement to our previous studies [17].

Acknowledgements

This work was supported by the Research Council of Norway through a Centre of Excellence Grant (Grant No. 179568/V30) and the Swedish Research Council (VR). Christoffer Lind, Laura Liikanen and Jonas Nordhaug Myhre are acknowledged for contributing with inputs and scientific discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jmgm.2015.05.007

References

[1] J. Åqvist, C. Lind, J. Sund, G. Wallin, Bridging the gap between ribosome structure and biochemistry by mechanistic computations, Curr. Opin. Struct. Biol. 22 (2012) 815–823.
[2] L.V. Bock, C. Blau, G.F. Schröder, I.I. Davydov, N. Fischer, H. Stark, et al., Energy barriers and driving forces in tRNA translocation through the ribosome, Nat. Struct. Mol. Biol. 20 (2013) 1390–1396.
[3] G. Zhao, J.R. Perilla, E.L. Yufenyuy, X. Meng, B. Chen, J. Ning, et al., Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics, Nature 497 (2013) 643–646.
[4] D.S.D. Larsson, L. Liljas, D. van der Spoel, Virus capsid dissolution studied by microsecond molecular dynamics simulations, PLoS Comput. Biol. 8 (2012), http://dx.doi.org/10.1371/journal.pcbi.1002502
[5] E.R. May, Recent developments in molecular simulation approaches to study spherical virus capsids, Mol. Simul. 40 (2014) 878–888.
[6] B.O. Brandsdal, F. Osterberg, M. Almlöf, I. Feierberg, V.B. Luzhkov, J. Åqvist, Free energy calculations and ligand binding, Adv. Prot. Chem. 66 (2003) 123–158.
[7] J. Åqvist, V.B. Luzhkov, B.O. Brandsdal, Ligand binding affinities from MD simulations, Acc. Chem. Res. 35 (2002) 358–365.
[8] C.J. Woods, J.W. Essex, M.A. King, The Development of Replica-Exchange-Based Free-Energy Methods, J Phys. Chem. B 107 (2003) 13703–13710.
[9] C.H. Bennett, Efficient estimation of free-energy differences from Monte-Carlo data, J. Comput. Phys. 22 (1976) 245–268.
[10] W. Jiang, B. Roux, Free energy perturbation hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations, J. Chem. Theory Comput. 6 (2010) 2559–2565.
[11] S. Kumar, D. Bouzida, R.H. Swendsen, P.A. Kollman, J.M. Rosenberg, The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The Method, J. Comput. Chem. 13 (1992) 1011–1021.
[12] J.W. Pitera, W.F. van Gunsteren, A comparison of non-bonded scaling approaches for free energy calculations, Mol. Simul. 28 (2002) 45–65.
[13] T. Steinbrecher, D.L. Mobley, D.A. Case, Nonlinear scaling schemes for Lennard–Jones interactions in free energy calculations, J. Chem. Phys. 127 (2007) 214108.
[14] A. Warshel, Computer Modeling of Chemical Reactions in Enzymes and Solutions, Wiley-Interscience, New York, 1991.
[15] J. Åqvist, A. Warshel, Simulation of enzyme-reactions using valence-bond force-fields and other hybrid quantum-classical approaches, Chem. Rev. 93 (1993) 2523–2544.
[16] S. Bjelic, B.O. Brandsdal, J. Åqvist, Cold adaptation of enzyme reaction rates, Biochemistry 47 (2008) 10049–10057.
[17] G.V. Isaksen, J. Åqvist, B.O. Brandsdal, Protein surface softness is the origin of enzyme cold-adaptation of trypsin, PLoS Comput. Biol. 10 (2014), http://dx.doi.org/10.1371/journal.pcbi.1003813
[18] J. Marelius, K. Kolmodin, I. Feierberg, J. Åqvist, Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems, J. Mol. Graph. Model. 16 (1998) 213–225.
[19] J. Åqvist, C. Medina, J.E. Samuelsson, A new method for predicting binding affinity in computer-aided drug design, Protein Eng. 7 (1994) 385–391.
[20] P.A. Kollman, Free-energy calculations—applications to chemical and biochemical phenomena, Chem. Rev. 93 (1993) 2395–2417.
[21] J. Sund, C. Lind, J. Åqvist, Binding site preorganization and ligand discrimination in the purine riboswitch, J. Phys. Chem. B 119 (2015) 773–782.
[22] M.P. Allen, D.J. Tildesley, Computer Simulation of Liquids, Oxford University Press, New York, 1987.
[23] V. Luzhkov, J. Åqvist, Free-energy perturbation calculations of binding and transition-state energies: hydrolysis of phenyl esters by β-cyclodextrin, Chem. Phys. Lett. 302 (1999) 267–272.
[24] R.W. Zwanzig, High-temperature equation of state by a perturbation method. 1. Nonpolar gases, J. Chem. Phys. 22 (1954) 1420–1426.
[25] S. Bjelic, J. Åqvist, Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site, Biochemistry 43 (2004) 14521–14528.
[26] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (1996) 11225–11236.
[27] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, J. Phys. Chem. B 105 (2001) 6474–6487.
[28] J. Åqvist, Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions, J. Comput. Chem. 17 (1996) 1587–1597.
[29] B.O. Brandsdal, A.O. Smalås, J. Åqvist, Electrostatic effects play a central role in cold adaptation of trypsin, FEBS Lett. 499 (2001) 171–175.
[30] B.H. Norman, T.I. Richardson, J.A. Dodge, L.A. Pfeifer, G.L. Durst, Y. Wang, et al., Benzopyrans as selective estrogen receptor β agonists (SERBAs). Part 4: functionalization of the benzopyran A-ring, Bioorg. Med. Chem. Lett. 17 (2007) 5082–5085.
[31] R. Helland, I. Leiros, G.I. Berglund, N.P. Willassen, A.O. Smalås, The crystal structure of anionic salmon trypsin in complex with bovine pancreatic trypsin inhibitor, Eur. J. Biochem. 256 (1998) 317–324.
[32] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.