

Cite This: J. Chem. Theory Comput. 2018, 14, 3218–3227

Estimating Atomic Contributions to Hydration and Binding Using Free Energy Perturbation

Benedict W. J. Irwin^{*,†} and David J. Huggins^{†,‡}

[†]Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, 19 J J Thomson Avenue, Cambridge CB3 0HE, United Kingdom

[‡]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

S Supporting Information

ABSTRACT: We present a general method called atom-wise free energy perturbation (AFEP), which extends a conventional molecular dynamics free energy perturbation (FEP) simulation to give the contribution to a free energy change from each atom. AFEP is derived from an expansion of the Zwanzig equation used in the exponential averaging method by defining that the system total energy can be partitioned into contributions from each atom. A partitioning method is assumed and used to group terms in the expansion to correspond to individual atoms. AFEP is applied to six example free energy changes to demonstrate the method. Firstly, the hydration free energies of methane, methanol, methylamine, methanethiol, and caffeine in water. AFEP highlights the atoms



in the molecules that interact favorably or unfavorably with water. Finally AFEP is applied to the binding free energy of human immunodeficiency virus type 1 protease to lopinavir, and AFEP reveals the contribution of each atom to the binding free energy, indicating candidate areas of the molecule to improve to produce a more strongly binding inhibitor. FEP gives a single value for the free energy change and is already a very useful method. AFEP gives a free energy change for each "part" of the system being simulated, where part can mean individual atoms, chemical groups, amino acids, or larger partitions depending on what the user is trying to measure. This method should have various applications in molecular dynamics studies of physical, chemical, or biochemical phenomena, specifically in the field of computational drug discovery.

INTRODUCTION

Free energy methods refer to an existing set of methods for estimating free energy differences, for example traditional Free Energy Perturbation (FEP) summed using either exponential averaging (EXP) or the Bennett Acceptance Ratio (BAR) and thermodynamic integration (TI). These methods have become a cornerstone of accurate binding free energy calculations,^{1,2} and many reviews have scrutinized the details of such methods, concluding that they are a promising addition to the set of tools used in the drug discovery industry.³⁻⁸ Free energy methods are implemented in many commonly used biological simulation packages, for example NAMD,⁹⁻¹¹ Desmond,¹² GROMACS,¹³ BOSS,¹⁴ AMBER,¹⁵ and others. These methods are path based and rely on a set of intermediate states, named λ -windows, in which the interaction parameter λ is altered from 0 to 1. When $\lambda = 0$, the simulation represents the initial state of a physical system; this might be the unbound state of a protein and drug complex in a binding free energy calculation or a simulation cell of water for a solvation free energy calculation. When $\lambda = 1$, the simulation represents the final state of a physical system. This might be the bound state for the protein-drug complex or the solvated molecule for the solvation free energy calculation. Then a set of intermediate states is taken to gradually measure the free energy change as λ varies from 0 to 1. FEP can be used to extract hydration free energies,¹⁶⁻¹⁸ free energies associated with mutating one molecule into another or one protein into another,¹⁹ and in the context of drug discovery, binding free energies of inhibitors to proteins^{12,20,2}

The method described in this work is called atom-wise free energy perturbation (AFEP) and is an extension to the EXP method. It is applied to energy and trajectory data from a conventional Molecular Dynamics (MD) simulation with multiple lambda windows. AFEP estimates the contributions from each of the atoms in a system to a general free energy change, given that the total energy of the system is defined to be a sum over atomic energies which are calculated from the MD trajectories. AFEP relies on the same set of MD simulations as conventional FEP analysis does, running a simulation at each of the intermediate λ -values. It then calculates quantitatively the contributions to the free energy change from each "part" of the system by approximating the decomposition of free energy in a simple and intuitive way. The definition of "part" is flexible and could be single atoms or

Received: January 10, 2018 Published: April 30, 2018

clusters of atoms grouped together (e.g., chemical groups, amino acids, or the solvent) depending on the desired application. The decomposition provided by AFEP cannot be additive between different systems as shown by Mark et al.,²² and similar chemical motifs cannot be expected to have the same free energy contribution in different molecules. However, the results can be interpreted within a given molecule and used as an empirical tool to highlight atoms which contribute relatively favorably or unfavorably to a free energy change.

In this work the atom-wise decomposition is calculated for two types of free energy change: a few hydration free energies and a binding free energy. The extension to FEP discussed in this study is called Atom-wise Free Energy Perturbation because it processes the results of an FEP simulation to give the atom-wise breakdown of the free energies. Many extensions have previously been developed for free energy methods, ^{1,2,23} and methods for statistically optimizing the results have been developed¹¹ including BAR^{24,25} which is used in this study to calculate total free energy changes. To the authors' knowledge, none of these methods offer the atom-wise distribution results that AFEP provides.

In general, free energy methods output a single number, a free energy change, that can help predict *whether* a drug will bind, and this has the potential to improve the computational drug design workflow.¹² With the correct collaboration there is hope that free energy methods will become an important part of the drug design and discovery process.²⁶ With AFEP one can attempt to determine *why* a drug binds and which parts play which a role in the binding. In this work we give all of the AFEP formulas in the main text. The mathematical derivation of the AFEP method is included in the Supporting Information.

The results are shown for AFEP directly applied to simulations of methane, methanol, methylamine, methanethiol, and caffeine solvated in water to measure a hydration free energy and a system of human immunodeficiency virus type 1 protease (HIV-1pr) with bound and unbound lopinavir drug molecules to estimate the atom-wise distribution of the binding free energy (see Figure 1).

The four methane-like molecules are chosen as similar simple examples of common chemical side groups. Caffeine is a small molecule that blocks the adenosine receptor in the human body. Although quite small compared to most medicinal molecules, it is an example of a molecule that might be considered in the field of drug discovery. Caffeine dissolves in



Figure 1. Two molecules investigated in this study. Left: The molecule caffeine. Right: The molecule lopinavir, an inhibitor of HIV-1pr.

water at room temperature; it is then expected to have a negative free energy of hydration. Caffeine was chosen as it is a small and familiar molecule that mixes polar and nonpolar groups and can be used to test the predictions of the AFEP method against chemical intuition.

HIV-1pr is one of many constituent proteins in HIV-1. This particular protein is responsible for cutting and cleaving parts of the host, after which the virus will go on to reproduce inside the host using the host's cell based machinery.²⁷ Specifically the virus uses HIV-1pr to cleave Gag and Gag-Pol: two proteins that are essential for the virus to hijack to synthesize a functional and intact viral particle. If HIV-1pr is successfully blocked with a strongly binding inhibitor, that is, an inhibitor with a highly negative binding free energy, the cutting process will be blocked. This will make it impossible for the virus to reproduce, and the infection can subsequently be eliminated by the human body. Lopinavir is a widely used inhibitor of HIV-1pr that is often combined with ritonavir, another drug molecule.²⁸ This application was picked as lopinavir binds strongly to the well studied HIV-1pr structure (see Figure 2).²⁹



Figure 2. HIV-1pr with lopinavir in the bound state. Lopinavir has been colored with green carbon atoms. The water and ions in the simulation cell are not shown.

THEORY

AFEP is a methodology that splits the total free energy change of the system into a sum of atom-wise contributions. The method is based on the Zwanzig equation³⁰ used in the EXP free energy method to calculate the Helmholtz free energy difference ΔF_{AB} between two thermodynamic states A and B

$$\Delta F_{AB} = F_B - F_A = -\beta^{-1} \log \langle e^{-\beta (U_B - U_A)} \rangle_A \tag{1}$$

with $\beta = (k_B T)^{-1}$, where k_B is Boltzmann's constant, U_X is the total energy of system X, and $\langle \cdot \rangle_A$ represents a state average over system A. From hereon, all state averages are performed over system A without loss of generality, and we will omit the subscript A from equations. Eq 1 is a statement that for any free energy differences between systems only the difference in system *energy* is needed to calculate the free energy difference. In practice the free energy difference must be reasonably small because of limited sampling overlap from the start and end states, which would lead to poor convergence. The Zwanzig equation is the starting point for the EXP method, and this equation would normally be applied between each pair of subsequent intermediate λ -values to help smoothly measure a larger system change in smaller steps. For the AFEP method we

define the expansion of the difference in total system energy, ΔU , in terms of the *N* constituent atoms in those systems

$$\Delta U = U_B - U_A = \sum_{k=1}^{N} \Delta u_k \tag{2}$$

$$\Delta u_k = u_{Bk} - u_{Ak} \tag{3}$$

where u_{Xk} is the potential energy associated with atom k in system X. The exact representation of u_{Xk} depends on the force field used to simulate the system, and an expression is given later in the text. The goal of AFEP is to write the free energy change as

$$\Delta F_{AB} = \sum_{k=1}^{N} \Delta \mathcal{F}_{AB}(k) \tag{4}$$

where $\Delta \mathcal{F}(k)$ is the contribution to the free energy from atom k, and the sum ranges over all of the N atoms in the system. It is worth stating that the free energy cannot usually be separated into a simple sum in this way, because the entropic contributions are formed from correlations of the multibody density of the system which are not necessarily separable over atoms.³¹ The results should not be overinterpreted nor expected to sum together to conserve free energy in an additive way.²² The special stylization of $\Delta \mathcal{F}(k)$ then represents that this is a 'contribution' from an atom k but not a true free energy per se. The goal of this work is to derive a mathematical expression for such a free energy contribution and assess if such an estimate has any meaningful predictive power as an empirical tool.

To achieve this atom-wise decomposition we perform the following stages:

• Replace ΔU in the Zwanzig equation (eq 1) with the sum of energies (eq 2). These energies serve as a means of distinguishing which contribution belongs to which atom.

• Expand the exponential term as a Taylor series.

• Perform a multinomial expansion on each power term of this Taylor series.

• Observe that the resulting series is a sum of products of differences in atom-wise energy
$$\Delta u_k$$
. The Δu_k in the products are raised to integer exponents.

• Weight the terms by their exponents and group them into contributions toward each atom in the system. For example, a term that looks like $\Delta u_1^2 \Delta u_2 \Delta u_3$ half belongs to atom 1 and one-quarter to atoms 2 and 3.

• Write the logarithm in the Zwanzig equation as a Taylor series.

• Perform a multinomial expansion on each term in the Taylor series of the logarithm.

• Observe that this series is also a sum of products of differences in the previous terms. These previous terms are also raised to integer powers.

• Apply the same grouping technique to all products of terms to get individual contributions for atoms

• Find the closed form for these individual groupings that correspond to the $\Delta \mathcal{F}_{AB}(k)$ in eq 4.

This procedure is covered in Sections 1–5 in the Supporting Information in full mathematical detail. The resulting closed form expression for $\Delta \mathcal{F}_{AB}(k)$ is given by

$$\Delta \mathcal{F}_{AB}(a) = -\beta^{-1} \frac{\left\langle \frac{\Delta u_a}{\Delta U} (e^{-\beta \Delta U} - 1) \right\rangle}{\langle e^{-\beta \Delta U} - 1 \rangle} \log(1 + \langle e^{-\beta \Delta U} - 1 \rangle)$$
(5)

and by inspection of eqs 1 and 5, we have

$$-\beta^{-1}\log(1+\langle e^{-\beta\Delta U}-1\rangle)=-\beta^{-1}\log(\langle e^{-\beta\Delta U}\rangle)=\Delta F_{AB} \quad (6)$$

so in effect we have defined a set of weights w_k such that

$$\Delta \mathcal{F}_{AB}(a) = w_a \Delta F_{AB} \tag{7}$$

by eq 4 then

$$\sum_{k=1}^{N} w_k = 1 \tag{8}$$

and the weight for atom k is defined by

$$w_{k} = \frac{\left\langle \frac{\Delta u_{k}}{\Delta U} (e^{-\beta \Delta U} - 1) \right\rangle}{\langle e^{-\beta \Delta U} - 1 \rangle}$$
(9)

The calculation of these weights will allow the calculation of the free energy contribution for each atom in the system. Because the weights are independent of the value of the total free energy change ΔF_{AB} any method desired can be used to calculate ΔF_{AB} , for example a standard BAR FEP simulation,²⁴ which is relatively easy using modern methods.¹¹ All that is needed are the energies associated with each atom in the system, u_{Xa} , which will come from the force field parameters used to simulate the system. The full AFEP expression for the estimate of the free energy contribution from atom *a* is then given by

$$\Delta \mathcal{F}_{AB}(a) = \frac{\left\langle \frac{\Delta u_a}{\Delta U} (e^{-\beta \Delta U} - 1) \right\rangle}{\langle e^{-\beta \Delta U} - 1 \rangle} \Delta F_{AB}$$
(10)

One could define other partitioning schemes that give $\Delta \mathcal{F}_{AB}(a) = f(\Delta U, \Delta u_a) \Delta F_{AB}$, for some weight function $f(\Delta U, \Delta u_a)$. A simple example that sums to 1 for all *a* is $f(\Delta U, \Delta u_a) = \langle \Delta u_a / \Delta U \rangle$. This weight function only considers the energy of each atom and would struggle to predict entropic effects; it is also not rooted to a derivation like the AFEP weights.

SIMULATION DETAILS

In the MD simulations, the total system energy comes from electrostatic, Lennard-Jones, bonded, Urey–Bradley, angle, dihedral, and improper terms in the force field, which is a set of parameters for each type of atom in the system. The CHARMM36 force field was used in these simulations.³² This includes the parametrization of each amino acid in the protein and the explicit water molecules around the protein in the simulation. Fixed bonds and bond angles were used, meaning certain terms are not included in the energy calculations.

To get the atom-wise energies u_{Xa} needed to construct the AFEP weights (eq 9) from the MD simulation we must define the connection between the energy and the force field being used. This is given by

$$u_{Xa} = \frac{1}{2} (u_{\text{electrostatic}} + u_{\text{Lennard-Jones}} + u_{\text{bonded}} + u_{\text{Urey-Bradley}}) + \frac{1}{3} u_{\text{angle}} + \frac{1}{4} (u_{\text{dihedral}} + u_{\text{improper}})$$
(11)

where each term is the sum of all appropriate interactions containing atom *a*. This energy is a shared contribution from all types of energy measured with Hamiltonian X (the Hamiltonian will change as the FEP variable, λ , changes). For

simulations with fixed bond lengths and fixed angles, which are common in simulations of biological molecules, we have $u_{bonded} = u_{angle} = 0$. Eq 11 is a statement that for all possible types of interaction, if a number of atoms are involved in an interaction, the energy of that interaction is shared equally between all of the atoms involved. This is an approximation, and variations to this scheme could potentially be made, but to share equally is the least assuming scheme. This work does not include intramolecular terms between the nearest and next nearest bonded neighbors (1-2 and 1-3 terms).

Methane-like Molecules and Caffeine Simulations Details. Caffeine and the four methane-like molecules were simulated with explicit TIP3P³³ water with a lambda schedule that starts with the molecule completely noninteracting with the water ($\lambda = 0$) and ends with full interaction between the molecule and the water ($\lambda = 1$). The resulting free energy change is then the free energy of hydration for the molecule. For the methane-like molecules these free energies are shown in Table 2 along with experimental values.^{34,35} There is

Table 1. Input Parameters for the Calculation in Eq 12^{a}

parameter	value	units
K_r	0.5	kcal/mol/Å ²
$K_{ heta}$	0.005	kcal/mol/rad ²
K_{ϕ}	0.005	kcal/mol
V	1660.5	Å ³
<i>r_{aA,0}</i>	3.318	Å
$ heta_{A,0}$	2.07	rad
T	298	K

^{*a*}The equilibrium parameters are found from the natural binding pose in experimental data.

Table 2. FEP Calculations for Methane, Methanol,Methylamine, and Methanethiol Compared to ExperimentalResults^a

molecule	exptl	ΔF BAR FEP \pm statistical error
methane	2.00 ³⁴	2.45 ± 0.03
methanol	-5.10^{35}	-4.49 ± 0.05
methylamine	-4.57^{35}	-3.41 ± 0.05
methanethiol	-1.20^{34}	-0.06 ± 0.04

^{*a*}Units for all energies are in kcal/mol. References for experimental values are given.

experimental data for the caffeine free energy measurement taken from the FreeSolv database,¹⁶ and a similar computational measurement was performed by Mobley et al.¹⁶ These results make up Table 4, which compare experimental and computational results. The force field parameters for all molecules were generated with the ParamChem CGenFF software.36,37 This software picks the best atom types and charges to represent a given molecule according to its bonding and connectivity similarity with a test set of molecules. The original chemical structure and topology files for caffeine are taken from the ZINC database entry ZINC00001084.38 The systems were first equilibrated under NVT conditions and then NpT conditions to find a suitable density and subsequently run under NpT conditions for the main trajectory data collection. 32 λ -windows were used, and an MD trajectory was collected for each value of λ . Further details of the MD simulations are given in the Supporting Information.

HIV-1pr with Lopinavir Simulations. The structures used for the simulation of lopinavir in the binding pose of HIV-1pr are from The Protein Data Bank, reference 2Q5K.^{39,40} For convenience of processing, the entire free energy calculation was performed in one cell. In this cell there are two copies of the drug lopinavir: one is in the solution and is fully interacting at the beginning of the simulation, and the other is restrained in the natural binding pose in the protein from the PDB data³⁹ and is noninteracting in the beginning of the simulation. The copy in the solvent was placed 42 Å from the protein which was deemed to be suitably far to stop the majority of interactions with the protein. The simulations had three stages of lambda schedule, as is common for such binding free energy calculations.²⁰ The first stage is to turn on the Lennard-Jones interactions of lopinavir in the binding site; this creates a cavity in the protein, and the interactions are turned on very slowly at first. This is to prevent a so-called "end point catastrophe" in an FEP simulation, which is where very weakly interacting atoms in the MD simulation can overlap and the divergent form of their interaction potential leads to very large energies and slow convergence.¹¹ The second stage is to turn the Coulombic charge-based interactions of the solvent based copy of lopinavir off and to turn the charge-based interactions of the binding site ligand on; the third stage is to turn the Lennard-Jones interactions of the solvent copy of lopinavir off. This leaves the end point of the simulation representing the fully interacting and bound inhibitor in the protein and the copy of the inhibitor in the solution fully uninteracting. To keep the inhibitor in the binding site throughout the simulation, a tethering force is used between three atoms in the protein and one atom in the drug molecule. These restraints can be seen in Figure 3. There is a



Figure 3. Tethering of lopinavir (green carbons) to the protein (gray carbons). There is a strong hydrogen bond between H and O310 (dashed pink). Four atoms were chosen, O310 the ligand (pink oxygen) and C42, C47, and C44 in HIV-1pr (orange carbons). There is one dihedral restraint through all four, one angular restraint through O310, C47, and C44, and one distance restraint between O310 and C47.

free energy term that must be considered from this tethering which can be decomposed into two parts: the unnatural energy associated with the unphysical tethering force and the unnatural entropic term associated with prohibiting the ligand from accessing the full volume of the simulation cell. The first contribution is measured using a thermodynamic integration

methane	$\Delta F_{ m AFEP}$	methanol	$\Delta F_{ m AFEP}$	methylamine	$\Delta F_{ m AFEP}$	methanethiol	$\Delta F_{ m AFEP}$
С	2.21	С	-0.07	С	-0.04	С	-0.44
H1	-0.26	0	-4.79	H1	0.34	H1	0.29
H2	-0.24	H1	0.38	H2	0.34	H2	0.35
H3	-0.25	H2	0.60	H3	0.40	H3	0.32
H4	-0.23	H3	0.52	Ν	-4.30	S	-0.82
		H4 (O)	1.11	H4 (N)	1.05	H4 (S)	0.28
				H5 (N)	0.51		
^{<i>a</i>} The units of all	free energy char	nges are in kcal/mo	ol.				

Table 3. AFEP Free Energy Contributions for Each Atom in Methane, Methanol, Methylamine, and Methanethiol⁴

(TI)^{41,42} of the system by varying the strength of each degree of freedom in the tethering forces. The strength of these contributions can then be calculated and corrected for. The latter contribution is analytically calculated using a derived equation; the expression for this correction is given by eq 12. Further details of the MD simulation of HIV-1pr with lopinavir are given in the Supporting Information.

Restraints Correction Factor. A three term restraint across four atoms is used to hold lopinavir in the binding pose in the binding site of HIV-1pr. Three atoms from the binding site and one atom in the ligand are connected with a dihedral term across all four atoms, an angular term across two of the protein atoms and the ligand atom and a separation term from one of the atoms in the protein to the atom in the ligand. This allows the ligand to adopt different orientational poses during the simulation, but keeps it fixed to the protein such that it does not wander through the partially interacting protein when the ligand and protein are only weakly interacting. An analytic correction term ΔA_r can be derived to compensate for the restricted environment the ligand resides in as there is an entropic cost of the ligand not being free to wander around the simulation cell. The strength of the restraints is chosen to be suitably weak as not to affect the AFEP results.

Following the method used in the paper by Boresch et al.²⁰ who analytically calculate this entropic cost for a six point restraint of the ligand in the binding site of the protein, a similar expression was derived for a three point restraint; this gives the free energy correction term associated with constraining the ligand when interactions are turned off as

$$\Delta A_{r} = -k_{\rm B}T \log \left[\frac{8\pi^{2}V \sqrt{K_{r}K_{\theta}K_{\phi}}}{r_{\rm aA,0}^{2} \sin \theta_{A,0} (2\pi k_{\rm B}T)^{3/2}} \right]$$
(12)

where V is the standard system volume for 1 molar concentration, K_r is the strength constant of the distance constraint, K_{θ} is the strength constant of the angular constraint, and K_{ϕ} is the strength constant of the dihedral constraint. $r_{aA,0}$ is the equilibrium distance between the drug atom and the protein atom. $\theta_{A,0}$ is the equilibrium angle between the atoms with the angular constraint. In addition to this entropic term, there is the energetic cost of the restraints themselves. This should be made as small as necessary to not perturb the system unnaturally when interactions are at the normal level. The restraint must also be strong enough to hold the ligand in the natural binding pose when interactions with the rest of the system are switched off. A set of TI measurements was taken out to ensure the contribution from the restraints when interactions were turned on was close to 0 kcal/mol, which would indicate that it is not interfering with the dynamics of the fully interacting system. The total free energy associated with the restraint was found to be 0.6113 kcal/mol. This value is

relatively small and is deemed to be acceptable for the purposes of this simulation.

Figure 3 shows a small section of the system to demonstrate the restraints used. Table 1 shows the values of the constants used in the simulation; these are then used to calculate the correction factor (eq 12).

RESULTS

Methane, Methanol, Methylamine, and Methanethiol. FEP calculations were performed for methane, methanol, methylamine, and methanethiol to demonstrate the AFEP method for similar small molecules with different side chains. Table 2 shows the total hydration free energy change for these four molecules. The calculated values agree with the experimental values reasonably well. Table 3 shows that atom-wise contributions for the molecules as calculated using AFEP.

Caffeine. A free energy calculation was performed for caffeine solvated in water; AFEP was then applied to the trajectory information to produce an atom-wise breakdown of the hydration free energy. Table 4 shows the total free energy calculated from a standard FEP simulation compared with experiment and a similar computation.

 Table 4. Total Free Energy Values Computed by Mobley et al., Experimentally Measured Values, and Values Computed with a BAR FEP Simulation^a

molecule	Mobley et al. ¹⁶	exptl ¹⁶	ΔF BAR FEP		
caffeine	-17.62 ± 0.04	-12.64 ± 0.74	-18.53 ± 0.17		
^{<i>a</i>} The uncertainty quoted in the BAR FEP table is the statistical error					
from the BA	R routine.				

Figure 4 shows the atom-wise breakdown of free energies associated with each atom in the caffeine molecule in units of kcal/mol. All of the water molecules in the simulation have been partitioned into one group because they are indistinguishable. For the caffeine simulation the contribution from the water molecules sums to exactly $\Delta F/2$. This arises because the atom-wise energy is defined to be shared evenly across interactions in eq 11. When partitioned in this way the free energy contribution per water molecule is less meaningful, and specific water molecules of chemical interest should not be partitioned into the bulk. Contributions from water molecules within certain shells from the solute could be analyzed by partitioning them carefully according to their distance from the solute.

Some of the atoms in a caffeine molecule are effectively in the same chemical environment. For example H1, H2, and H3 (as labeled in Figure 4) are in the same environment because C1 could rotate freely about its bond with N1. In the limit of

Journal of Chemical Theory and Computation



Figure 4. Hydration free energy contributions from each atom in the caffeine molecule in kcal/mol. In the right-hand side image, blue, white, and red atoms have unfavorable, neutral, and favorable contributions, respectively.

perfect sampling we would expect these three hydrogen atoms to have the same free energy contribution. In practice it may take a long time in an MD simulation to observe the conformational changes required to sample this. This problem is common in many forms of MD. Obvious symmetries can be input manually by grouping atoms together; however, the user should be careful not to insert fictitious symmetries. In the case of the MD simulations performed in this study, fixing the atoms prevents the rotation, and the weights for these atoms are not expected to be the same. A deeper analysis of the values in Figure 4 is given in the Analysis section.

Figure 5 shows the convergence of some of the atom-wise weights associated with atoms in the caffeine molecule as the



Figure 5. Convergence of the atom-wise weights with considered trajectory frames for some of the atoms in a caffeine molecule solvated in water. Not all atoms are shown to increase the readability of the plot. Points are taken at every 50 frames from 50 to 1800.

number of considered trajectory frames is increased. The convergence is steady and fairly rapid only requiring around 1000 frames to get a reasonable feel for the contribution from each atom. The values in Figure 4 are calculated from 1800 MD trajectory frames and are expected to be converged according to Figure 5.

HIV-1pr with Lopinavir. Table 5 shows the five terms making up the total system free energy, the total free energy, and an experimental comparison. Three terms are from the different stages of simulation, and the remaining two are

Article

Table 5. Five Contributions Towards the Total Binding Free
Energy and Their Total Compared with an Experimental
Value ^d

stage	method	ΔF	statistical error
			a -
nonbonded	BAR FEP	-21.44	0.57
turning off unbound molecule nonbonded	BAR FEP	-5.92	0.28 ^a
switching Coulombic interactions	BAR FEP	-1.77	0.12 ^a
analytic correction term	calculated	5.92	
TI restraints (energetic)	TI	-0.61	0.02 ^b
total binding free energy	all	-23.82	0.65
experimental binding free energy ^{c43,44}		-15.7	

^{*a*}Statistical errors from the BAR FEP routine. ^{*b*}Half the spread of the forward and backward TI calculations. ^{*c*}Sourced from BindingDB.⁴⁴ ^{*d*}The errors are statistical errors and do not take into account systematic errors from simulation parameters.

correction terms from the restraints used to control the simulations.

Figure 6 shows the results of an AFEP calculation on the MD simulation trajectories. Each contribution is given by $w_{\text{bound}}\Delta F_{\text{bound}} + w_{\text{swap}}\Delta F_{\text{swap}} + w_{\text{unbound}}\Delta F_{\text{unbound}}$. In each case the weights are given by the difference between the atom-wise weights from the copy of the inhibitor in the binding site and



Figure 6. A representation of the binding free energy contributions from the lopinavir molecule in the binding site of the protein (not shown). The binding pose has been altered to display the atoms clearly. Blue atoms (C15–O4) are the least favorably contributing. Red atoms (HO4-H2) are strongly favorable with white weakly favorable or neutral. The atom-wise free energy contributions are shown for 12 unfavorable contributors and 12 favorable contributors as labeled in the figure.

Journal of Chemical Theory and Computation

the inhibitor in the solvent. The 12 most contributing and 12 least contributing sites across the lopinavir molecule are labeled. Red atoms are strongly contributing to the binding. Blue sites are those that disfavor the binding. White sites are neutral. Some of the red atoms are points of symmetry; for example, C35, H6, C15, C6, and H2 could all be reflected on the molecular structure by rotations of the side groups. This again implies perfect sampling of all conformational degrees of freedom has not occurred; however, in this case the binding site will restrict the ligand and prohibit these rotations. AFEP proves useful in determining that this asymmetry is present in the results.

ANALYSIS

Methane-like Molecules. The free energy calculations for four methane-like molecules mentioned in Tables 2 and 3 show a reasonable agreement with experimental free energy values. For methane the AFEP decomposition predicts that each of the hydrogens is weakly favorable and of similar magnitude which could be interpreted as interactions with the surrounding water. In methanol the oxygen atom is very favorable, and the central carbon has become slightly polarized by the addition of the oxygen. The least favorable atom is the hydrogen in the hydroxyl group, H4; this may be due to hydrogen atoms from surrounding water molecules which are bonding with O crowding closely to H4 and offsetting the contributions made from hydrogen bonds from water to H4. If H4 and O are combined, then the contribution from the OH group is still very favorable. A similar solvent crowding effect may explain the positive contributions from H4 and H5 in methylamine. Again the central carbon is slightly polarized, and the nitrogen atom is very favorable, presumably from hydrogen bonding with the surrounding water. For methanethiol, the net free energy change is small, but each of the contributions is of similar magnitude to methanol and methylamine. The sulfur atom is relatively favorable. The free energies between different molecules cannot be easily interpreted, which is expected according to Mark et al.²² For example, if methane and methanol are compared, the free energy of mutating H to OH cannot be found by using the total free energy and by splitting methane into $CH_3 + H$. This is because the CH_3 partition of methane and methanol is not equivalent, partly due to the polarization of the central carbon in methanol. For all cases, the results potentially make more sense when chemical groups are summed over.

Caffeine Total Free Energy. Table 4 shows the total hydration free energy for experimental and simulated caffeine. There is a mismatch between the total free energy for computational simulations and experimental measurements; this is likely because the force field does not contain full information about the molecular interaction, and this is a common problem with classical molecular dynamics simulations that is not specific to the AFEP methodology. The agreement between the two computational methods is within 1 kcal/mol; this is a reasonable tolerance for demonstrating the simulation was run correctly, and the results can be used to explore the AFEP method. The simulation by Mobley et al. was performed in the GROMACS molecular dynamics software¹⁶ and has slight differences in the input and parameter files. It is possible that convergence of the total free energy change could be improved by adding more λ -windows to the BAR calculation. However, showing an accurate result for the total free energy change is not the goal of this work; moreover one

of the goals is to show that AFEP produces a valid decomposition independent of the accuracy of the input free energy change.

Caffeine Atom-wise Free Energy. For caffeine the atomwise weights given by eq 9 were well converged. This can be seen in Figure 5, where after around 1000 considered trajectory frames the individual contributions to the free energy do not change considerably. The atom-wise values shown in Figure 6 correspond to 1800 input trajectory frames. Although the results appear well converged, we can tell that perfect sampling has not occurred. Chemically speaking, H1, H2, and H3 should have exactly the same contribution by rotational symmetry about the C1-N1 axis. However, the free energy contribution from H2 is somewhat different than H1 and H3. This highlights that fixed atoms were used in the simulation; if fixed atoms were not used, then such a difference would have indicated a sampling problem. These kinds of sampling problems are not particular to the AFEP methodology and are common in MD simulations. However, for AFEP there is the advantage that prior symmetry information of this kind could be input manually using the atom partitioning scheme. The researcher may still have to be careful before inputting such degeneracies. For example, if the energy of a conformational barrier is particularly high, much greater than the average thermal energy in the system, $k_{\rm B}T$, then one may lose information by partitioning the atoms together. In this case AFEP could even be a way of diagnosing such conformational barriers, by checking sets of atoms that should have the same contribution by symmetry for distinctly different contributions.

If the AFEP contributions are summed over methyl groups, two out of three cancel to almost zero net contribution. The signs of contributions are consistent across the molecule; for example, all three methyl groups show positive hydrogens attached to a negative carbon attached to a negative nitrogen. N2 is the best contributor to the solubility of the molecule. This makes sense as it is the most solvent exposed part of the molecule and the N2 molecule is assigned a large partial charge in the force field parameters. According to the AFEP simulation C6 is actively resisting the solvation of the molecule. Both of the oxygen atoms have quite large contributions to the hydration free energy. This makes sense as they will create hydrogen bonds with the surrounding water. H4 and C2 are both unfavorably contributing to the solvation. This may suggest that replacing H4 with an OH or NH₂ group would increase the solvation free energy of the molecule.

Lopinvair Total Binding Free Energy. The total free energy change of the HIV-1pr and lopinavir complex was made from five distinct contributions; there are the contributions from the three lambda stages, turning the VDW interactions on in the binding site, swapping the charge interactions from the solvent molecule to the bound molecule, and turning the VDW interactions off in the solvent, as well as the analytic correction for the entropic cost of restraining a ligand in free space to the binding site (eq 12), and the free energy associated with the restraints measured using thermodynamic integration. The sum of these five contributions gave the final value of -22.59 kcal/mol. The experimental result for this binding is -15.2 kcal/mol⁴⁵ as found using the BindingDB.⁴⁴

Lopinvair Atom-wise Free Energy. The AFEP contributions were found for atoms in the lopinavir molecule. The unfavorably contributing sites are colored blue, and the favorable sites are colored red in Figure 6. Both the oxygens and the nearest carbons are positively contributing for O1 and O4. This is in contrast to the caffeine molecule, where oxygens are paired with unfavorable carbons. There are a few examples of bonded pairs with a favorable and unfavorable atom, for example HN4 and N4, HN2 and N2, HN3 and N3, and HO4 and O4; the latter pairs the strongest of each type of contribution, and all pairs result in significant cancellation if summed over. There are some obvious side groups on the molecule that have few strongly binding atoms and are quite neutral to the binding process. One example is the ring featuring H6, which could potentially be replaced with a similar sized side group that might display a stronger binding.

The protein atoms in the binding site can also be considered and also have atom-wise contributions. Visualization of the interactions between the drug and the binding site are hard due to the complicated 3-dimensional nature of the problem. In a drug discovery context only the drug molecule can be altered and not the protein, so these binding site atoms are not shown in this work. However, in investigations looking into the mechanism of specific protein interactions, this additional information is expected to be useful.

CONCLUSIONS

The Atom-wise Free Energy Perturbation (AFEP) method described in this article provides a detailed breakdown of a free energy change across partitions of atoms in a molecular dynamics simulation. The partitions can be selected by the user to capture information at an appropriate length scale. The results shown in this work indicate that a full atom-wise decomposition may be too detailed for certain features, and groups such as amines and hydroxyls should probably be summed over. The main used equations were directly derived from the Zwanzig equation as used in the exponential averaging method (EXP) for free energy changes at fixed volume. During this derivation it was defined that the system energy is decomposable as a sum over the atoms in the system. Two further assumptions were made in the derivation, the first that when decomposing the system energy into individual contributions associated with atoms, the energy is shared equally between groups of atoms interacting with a given force field term. The second that a simple repartitioning scheme can be used during the multinomial expansion stages of the derivation to give a meaningful estimate of the free energy at each atom. Because of the nature of the last approximation, the AFEP method cannot produce an additive decomposition of the free energy, and comparisons between the partitions of different molecules should be made very carefully. However, AFEP can be used as an empirical tool for studying the internal free energy differences of a given molecule.

The AFEP weights used in the decomposition appear to converge relatively quickly. The system may be partitioned as the user chooses under the method because these weights can be linearly combined. This means interchangeable components of the system, namely the fluid molecules that are in an indistinguishable environment, can be combined together. Atoms that are symmetric by rotation of chemical groups can also be partitioned together in this manner, as could entire amino acids if a large protein was being studied.

The potential applications of this information are numerous. Only six specific examples were covered in this work, but AFEP could be applied to simulations relating to various branches of physics, materials science, chemistry, and biochemistry. AFEP was applied to a calculation of the hydration free energy of methane, methanol, methylamine, methanethiol, and caffeine. AFEP highlighted which atoms appear to interact the most with the surrounding water molecules. AFEP makes sensible decompositions for the four methane-like molecules and indicates a particular hydrogen atom in caffeine (H4) is relatively weakly interacting with the surrounding water. This type of analysis could be a useful tool to predict, for example, which molecule similar to caffeine dissolves into water more readily.

AFEP was also used to produce a free energy breakdown of the binding free energy change for a protein—ligand complex, HIV-1pr with lopinavir. The binding free energy was made of three stages, AFEP was applied to each stage, and the results were summed along with two correction factors for the restraints used in the MD simulations. AFEP highlighted some sections of lopinavir that are neutral to the binding and may find use as a tool to suggest improvements to a given ligand to increase binding strength. AFEP also showed which components of the lopinavir molecule are the strongest binding and are likely to be essential in the binding process.

The AFEP method can used to analyze the free energy contributions of atoms in most physical and chemical systems from appropriate molecular dynamics simulations arranged in lambda windows. This technology will have applications in the field of computational drug discovery and may assist in developing ligands for other disease target proteins in less time and cost. There is also a great scope for further biological and chemical uses away from the field of drug discovery, as the methods used are general and unassuming about the underlying physical system.

ASSOCIATED CONTENT

S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00027.

Derivation of the equations used in the AFEP method and proof that they are equivalent to the Zwanzig equation, details of the molecular dynamics simulations performed (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: djh210@cam.ac.uk.

ORCID 🔍

David J. Huggins: 0000-0003-1579-2496

Funding

B.W.J.I. acknowledges the EPSRC Centre for Doctoral Training in Computational Methods for Materials Science for funding under grant number EP/L015552/1. Work in D.J.H.'s lab was supported by the Medical Research Council (MRC) under Grant No. ML/L007266/1. This work was performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (http://www.hpc.cam.ac.uk/), provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and funding from the Science and Technology Facilities Council.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Cappel, D.; Hall, M. L.; Lenselink, E. B.; Beuming, T.; Qi, J.; Bradner, J.; Sherman, W. Relative Binding Free Energy Calculations Applied to Protein Homology Models. *J. Chem. Inf. Model.* **2016**, *56*, 2388–2400.

(2) Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* 2017, *50*, 1625–1632.

(3) Jorgensen, W. L. Free energy calculations: a breakthrough for modeling organic chemistry in solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.

(4) Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Annu. Rep. Comput. Chem. 2007, 3, 41–59.

(5) Michel, J.; Essex, J. W. Prediction of protein-ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 639–658.

(6) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.

(7) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. *J. Comput.-Aided Mol. Des.* **2015**, 29, 397–411.

(8) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. J. Chem. Inf. Model. **201**7, *57*, 2911–2937.

(9) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(10) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. J. Mol. Graphics **1996**, 14, 33–38.

(11) Liu, P.; Dehez, F.; Cai, W.; Chipot, C. A toolkit for the analysis of free-energy perturbation calculations. *J. Chem. Theory Comput.* **2012**, *8*, 2606–2616.

(12) Wang, L.; et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.

(13) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29*, 845–854.

(14) Jorgensen, W. L.; Tirado-Rives, J. Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. J. Comput. Chem. 2005, 26, 1689–1700.

(15) Case, D. A.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D.; Kollman, P. A. *Amber17*; 2017.

(16) Mobley, D. L.; Guthrie, J. P. FreeSolv: a database of experimental and calculated hydration free energies, with input files. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 711–720.

(17) Huggins, D. J.; Payne, M. C. Assessing the accuracy of inhomogeneous fluid solvation theory in predicting hydration free energies of simple solutes. *J. Phys. Chem. B* **2013**, *117*, 8232–8244.

(18) Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. J. Chem. Theory Comput. **2010**, *6*, 1509–1519.

(19) Steinbrecher, T.; Abel, R.; Clark, A.; Friesner, R. Free Energy Perturbation Calculations of the Thermodynamics of Protein Side-Chain Mutations. J. Mol. Biol. 2017, 429, 923–929.

(20) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.

(21) Rocklin, G. J.; Mobley, D. L.; Dill, K. A. Separated topologies-A method for relative binding free energy calculations using orientational restraints. *J. Chem. Phys.* **2013**, *138*, 085104.

(22) Mark, A. E.; van Gunsteren, W. F. Decomposition of the Free Energy of a System in Terms of Specific Interactions. *J. Mol. Biol.* **1994**, 240, 167–176.

(23) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* 2008, 129, 124105.

(24) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. J. Comput. Phys. 1976, 22, 245-268.

(25) Hahn, A. M.; Then, H. Characteristic of Bennett's acceptance ratio method. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* **2009**, *80*, 031111.

(26) Sherborne, B.; Shanmugasundaram, V.; Cheng, A. C.; Christ, C. D.; DesJarlais, R. L.; Duca, J. S.; Lewis, R. A.; Loughney, D. A.; Manas, E. S.; McGaughey, G. B.; Peishoff, C. E.; van Vlijmen, H. Collaborating to improve the use of free-energy and other quantitative methods in drug discovery. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 1139–1141.

(27) Clavel, F.; Mammano, F. Role of Gag in HIV Resistance to Protease Inhibitors. *Viruses* **2010**, *2*, 1411–1426.

(28) Walmsley, S.; Bernstein, B.; King, M.; Arribas, J.; Beall, G.; Ruane, P.; Johnson, M.; Johnson, D.; Lalonde, R.; Japour, A.; Brun, S.; Sun, E. Lopinavir-Ritonavir versus Nelfinavir for the Initial Treatment of HIV Infection. N. Engl. J. Med. 2002, 346, 2039–2046.

(29) Ghosh, A. K.; Osswald, H. L.; Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. J. Med. Chem. 2016, 59, 5172–5208.

(30) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. J. Chem. Phys. **1954**, 22, 1420–1426.

(31) Irwin, B. W. J.; Huggins, D. J. On the accuracy of one-and twoparticle solvation entropies. J. Chem. Phys. 2017, 146, 194111.

(32) Huang, J.; MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.

(33) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(34) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. Thermodynamics of solute transfer from water to hexadecane. *J. Chem. Soc., Perkin Trans.* 2 **1990**, 291.

(35) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. J. Chem. Phys. **1984**, 81, 2016–2027.

(36) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. J. Chem. Inf. Model. 2012, 52, 3144–3154.

(37) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. J. Chem. Inf. Model. 2012, 52, 3155–3168.

(38) Irwin, J. J.; Shoichet, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. J. Chem. Inf. Model. 2005, 45, 177–182.

(39) Reddy, G. S. K. K.; Ali, A.; Nalam, M. N. L.; Anjum, S. G.; Cao, H.; Nathans, R. S.; Schiffer, C. A.; Rana, T. M. Design and Synthesis of HIV-1 Protease Inhibitors Incorporating Oxazolidinones as P2/P2' Ligands in Pseudosymmetric Dipeptide Isosteres. *J. Med. Chem.* **2007**, *50*, 4316–4328.

(40) Berman, H. M. The Protein Data Bank. Nucleic Acids Res. 2000, 28, 235-242.

(41) Straatsma, T. P.; McCammon, J. A. Multiconfiguration thermodynamic integration. J. Chem. Phys. **1991**, 95, 1175–1188.

(42) Kirkwood, J. G. Statistical Mechanics of Fluid Mixtures. J. Chem. Phys. 1935, 3, 300–313.

Journal of Chemical Theory and Computation

(43) Ali, A.; Reddy, G. S. K. K.; Cao, H.; Anjum, S. G.; Nalam, M. N. L.; Schiffer, C. A.; Rana, T. M. Discovery of HIV-1 Protease Inhibitors with Picomolar Affinities Incorporating N -Aryl-oxazolidinone-5-carboxamides as Novel P2 Ligands. *J. Med. Chem.* **2006**, *49*, 7342–7356.

(44) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* 2007, 35, D198–D201.

(45) Clemente, J. C.; Coman, R. M.; Thiaville, M. M.; Janka, L. K.; Jeung, J. A.; Nukoolkarn, S.; Govindasamy, L.; Agbandje-McKenna, M.; McKenna, R.; Leelamanit, W.; Goodenow, M. M.; Dunn, B. M. Analysis of HIV-1 CRF_01 A/E Protease Inhibitor Resistance: Structural Determinants for Maintaining Sensitivity and Developing Resistance to Atazanavir †. *Biochemistry* **2006**, *45*, 5468–5477.