Improving the Efficiency of Protein-Ligand Binding Free-Energy Calculations by System Truncation

Genheden, Samuel; Ryde, Ulf

Published in:
Journal of Chemical Theory and Computation

DOI:
10.1021/ct200853g

2012

Improving efficiency of protein–ligand binding free-energy calculations by system truncation

Samuel Genheden, Ulf Ryde*

Department of Theoretical Chemistry, Lund University, Chemical Centre,
P. O. Box 124, SE-221 00 Lund, Sweden

*Correspondence to Ulf Ryde, E-mail: Ulf.Ryde@teokem.lu.se,
Tel: +46 – 46 2224502, Fax: +46 – 46 2228648

2013-01-29

Abstract
We have studied whether the efficiency of alchemical free-energy calculations with the Bennett acceptance ratio method of protein–ligand binding energies can be improved by simulating only a part of the protein. To this end, we solvated the full protein in a spherical droplet with a radius of 46 Å, surrounded by vacuum. Then, we systematically reduced the size of the droplet and at the same time ignored protein residues that were outside the droplet. Radii of 40 to 15 Å were tested. Ten inhibitors of the blood clotting factor Xa were studied and the results were compared to an earlier study in which the protein was solvated in a periodic box, showing complete agreement between the two set of calculations within statistical uncertainty. We then show that the simulated system can be truncated down to 15 Å, without changing the calculated affinities by more than 0.5 kJ/mol on average (maximum difference 1.4 kJ/mol). Moreover, we show that reducing the number of intermediate states in the calculations from eleven to three gave deviations that on average were only 0.5 kJ/mol (maximum 1.4 kJ/mol). Together this shows that truncation is an appropriate way to improve efficiency of free-energy calculations for small mutations that preserve the net charge of the ligand. In fact, each calculation of a relative binding affinity requires only 6 simulations, each of which takes ~15 CPU hours of computation on a single processor.

Keywords: free-energy perturbation, Bennett acceptance ratio, ligand-binding affinities, periodic boundary conditions, system truncation, long-range electrostatics.

Introduction

Accurate estimation of protein–ligand binding affinities is a major challenge in computational chemistry. Although formally correct relative free energies can be obtained by alchemical free-energy techniques such a free energy perturbation (FEP) and thermodynamic integration (TI), they have found little use outside academia.[1,2] The main reason for this is that such methods are computationally demanding, because they require simulations of unphysical intermediate states involving extensive sampling of the phase space.[3]

More approximate methods to estimate binding affinities exists, which do not require simulations of intermediate states.[4] They are usually referred to as end-points methods because they sample only the complex, and possibly the free protein and the free ligand.[5] A popular method in this class is MM/GBSA (molecular mechanics with generalised Born and surface-area solvation).[6,7] Although it only requires a simulation of the complex, we have shown that it can actually be computationally more expensive than TI because it is intrinsically imprecise and requires averaging over many independent simulations to reach a precision comparable to that of FEP or TI.[8] In addition, the accuracy of some of the terms in MM/GBSA have been questioned and the method often fails to give a useful accuracy of the predicted affinities.[9,10,11] Another popular end-point method is the linear interaction energy (LIE).[12] We have shown that it is slightly more effective than MM/GBSA, although it also suffers from a poor precision and a varying accuracy.[13,14]

Therefore, alchemical free-energy calculations seem to be the method of choice, at least for relative binding affinities, and the challenge is then to make the method more efficient. Previously, we have analysed how many unphysical intermediate states and how long simulation are needed to obtain accurate results.[8] We showed that for our test case, rather few intermediate states (3 to 5) were enough and the simulation time should be ~1 ns for the protein–ligand simulations and 2 ns for the free ligand simulations.

Apart from improvements in the simulation protocol, the system itself can be changed in a way that reduces the computer requirements. One approach that has been used by some research groups for the calculation of protein–ligand affinities is to simulate only a part of the protein, immersed in a droplet of explicit water molecules. In some approaches, the water droplet is surrounded by vacuum and therefore special care has to be taken to ensure that the water molecules in the droplet have bulk-like behaviour.[15] One approach is to use stochastic boundary conditions, in which the outermost region is simulated using Langevin dynamics and thereby impose friction on the inner region.[16,17]

Another approach is to impose restraints on the water molecules. A radial potential has to be added to prevent the water from evaporating and hence keeping the number density constant through the droplet. This has been obtained with various kinds of potentials.[15,18,19] Moreover, the polarisation orientation of the water molecules is heavily affected by the presence of vacuum. In the SCAAS (surface-constrained all-atom solvent) model,[18] King and Warshel solved this problem by imposing a uniform distribution for the angle between the water dipole vector and the displacement vector from the origin. Essex and Jorgensen have developed a similar method[15] and they found it necessary to restrain also the vector perpendicular to the plane of the water molecule.

Alternatively, the excluded part of the simulated system can be replaced by continuum electrostatics. Roux and co-workers have introduced a technique called generalised solvent boundary potential (GSBP),[20,21] in which the effect of the excluded atoms are modelled using a solvent-shielded static field and a solvent-induced reaction field. The reaction field is expanded in a basis set representing the inner-region charge distribution. Both the static field and the basis set coefficients are pre-calculated by solving the Poisson–Boltzmann equation. Simonson et al. introduced an approach that is a combination of vacuum simulations and continuum corrections.[22] The vacuum simulations are performed using stochastic boundary

conditions and some charges are reduced to mimic solvent screening. After the simulation, full charges are re-introduced and the corresponding free energy is calculated. Finally, the solvation free energy of the system is estimated using Poisson–Boltzmann calculations.

Although these methods have been shown to work well in several applications, to our knowledge, there has been no systematic study on the effect of truncation. How large portion can be truncated without loosing accuracy? In this paper we will report such an investigation. We study the binding of a series of inhibitors of the blood clotting enzyme factor Xa (fXa), which was also used it in our previous study with the full protein.[8] This allows us to investigate the effect of going from an octahedral system with periodic boundary conditions to a full protein in a spherical solvent droplet and then to several truncated spherical system of different sizes. The results show that system truncation is an excellent approach to reduce the computational cost of protein–ligand free-energy calculations without loosing accuracy.

Methods

System preparation. The ten 3-amidinobenzyl-1*H*-indole-2-carboxamide inhibitors considered in this study are shown in Figure 1. They are named after their numbers in the original study.[23] The preparation of most of these ligands has been described before[8] and the new ligands (**5** and **51**) were prepared in an analogous way. All calculations were started from the crystal structure of fXa in complex with ligand **125** (PDB code 1lpk).[23] The crystal structure shows two conformations for one of the amidino groups of the ligand, but only a single conformation of the ligand was studied here (the A conformation) because our previous study did not show any difference between the affinities of the two alternative conformations.[8]

The preparation of the protein has also been described before:[24] All Arg and Lys residues were considered to have a positive charge and the Glu and Asp residues were considered to have a negative charge. His57 and 83 were protonated on the $N^{\delta 1}$ atom, His91, 145, and 199 on the $N^{\varepsilon 2}$ atom, and His13 on both atoms. To be comparable to our previous study, the protein was described by the Amber 99 force field[25] and the ligands with the general Amber force field,[26] with charges derived by the restrained electrostatic potential (RESP) method [27] using potentials calculated at the Hartree–Fock 6-31G* level and sampled with the Merz–Kollman scheme.[28] Parameters for the ligands are provided as supplementary material.

The protein–ligand complexes and the free ligands were solvated in a sphere of TIP3P[29] water molecules on a grid using a combination of the Q program (version 5),[19] the Amber 10 suite of programs,[30] as well as in-house scripts. First, the complex was fully immersed in a water sphere that extended at least 10 Å outside the protein. Second, water molecules outside of a certain radius from the nitrogen atom of the indole ring of the ligand were deleted (cf. Figure 1). Radii of 40, 35, 30, 25, 20, and 15 Å were used for these simulations. Protein residues with all atoms outside the simulation sphere were kept in the simulation, but were excluded from the calculations of non-bonded interactions. We also tested to cut away the protein atoms outside the sphere, but this did not improve the efficiency (and of course did not change the results). For the free ligand, solvent spheres with radii of 25, 20, or 15 Å was used.

Free-energy calculations. We have calculated the relative free energy of eight inhibitor transformations, as is described in Figure 1, using a thermodynamic cycle that has been described previously.[8,31] The free energies of the transformations were calculated using the Bennett acceptance ratio (BAR)[32]

$$\Delta G_i = RT\left(\ln\frac{\langle f(-\Delta U+C)\rangle_{\lambda_{i+1}}}{\langle f(\Delta U-C)\rangle_{\lambda_i}}\right)+C \qquad (1)$$

with

$$f(x) = 1/(1+\exp(x/RT)) \qquad (2)$$

and

$$C = \Delta G_i + RT\ln\frac{N_i}{N_{i+1}} \qquad (3)$$

where R and T are the gas constant and the absolute temperature, respectively, ΔU is the difference in energy between the system at λ_i and the system at λ_{i+1} and $U(\lambda) = (1 - \lambda)U_0 + \lambda U_1$, where U_0 and U_1 are the potentials of two physical end states. N_i is the number of samples when sampling at λ_i. The sampling was always performed for the system at λ_i as indicated in Eqn 1 and simulations were performed at λ = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1.0. The total free energy, ΔG, was obtained by summing over all λ values.

The transformation at each λ value was divided into an electrostatic and a van der Waals part, and the charges were mutated before the Lennard-Jones parameters were changed. A single-topology protocol was used[1] and dummy atoms were introduced for vanishing atoms. In the van der Waals transformation, a soft-core potential, as implemented in the Q package was used:[19,33]

$$V_{\mathrm{vdw}} = \frac{A_{ij}}{(r_{ij}^6+\alpha_{\mathrm{vdW}})^2} - \frac{B_{ij}}{r_{ij}^6+\alpha_{\mathrm{vdW}}} \qquad (4)$$

where i and j are two atoms, r_{ij} is the distance between them, A_{ij} and B_{ij} are the Lennard-Jones parameters in the force field, and α_{vdW} is a soft-core parameter, set to 10 Å^6 in all calculations. The soft-core potential was used only for atoms that were changed to a different atom type in the perturbations.

A soft-core version of the Coulomb potential was implemented into the Q package in an analogous fashion so that the electrostatic and van der Waals transformations could be calculated in a single simulation:

$$V_{\mathrm{ele}} = \frac{q_i q_j}{4\pi\,\epsilon_0\sqrt{r_{ij}^2+(1-\lambda)\alpha_{\mathrm{el}}}} \qquad (5)$$

where q_i and q_j are the atomic charges, ε_0 is the vacuum permittivity, and α_{el} is a soft-core parameter, set to 10 Å^2 in all calculations. The equation involves a λ dependence, which ensures that it coincidences with a normal Coulomb potential for λ = 1 (and V_{ele} = 0 for disappearing atom, which always were considered to be the λ = 0 state, because the charge vanishes in that state).

The potential in Eqn 4 does not contain any λ dependence. Therefore, the soft-core potential will be active also in the end state for non-disappearing atoms. Strictly, a free-energy calculation going from the soft-core potential to the normal Lennard-Jones potential is needed to reach the correct end state. However, test calculations showed that the free-energy change of such perturbations was negligible, e.g. 0.004 kJ/mol for the **49**→**53** transformation. The

same applies for non-disappearing perturbed atoms with the soft-core Coulomb potential: The free-energy difference between the soft-core and normal Coulomb potential for the **49**→**53** transformation was only 0.06 kJ/mol.

The approach using separate electrostatic and van der Waals transformations will be called the two-transformations approach (TTA) and it was used in all transformations unless otherwise stated. The approach in which electrostatics and van der Waals interactions are changed simultaneously, using the electrostatic soft-core potential in Eqn 5, will be called the single-transformation approach (STA).

MM/GBSA calculations. We have also carried out MM/GBSA calculations (molecular mechanics with generalised Born and surface-area solvation) for the same fXa–ligand complexes.[6,7] In this approximate method, the free energy of binding is calculated as the difference in free energy between the complex (PL), the protein (P), and the ligand (L), viz., $\Delta G_{\text{bind}} = G(\text{PL}) - G(\text{P}) - G(\text{L})$. Each of these free energies are estimated according to

$$G = \langle E_{\text{ele}} + E_{\text{vdw}} + G_{\text{pol}} + G_{\text{np}} - TS \rangle \qquad (6)$$

where E_{ele} and E_{vdw} are the molecular mechanics electrostatic and van der Waals energies, respectively, evaluated using the same force field as in the simulation but with no cut-off. G_{pol} is the polar solvation free energy evaluated using the generalised Born method of Onufriev, Bash and Case (model I, with $\alpha = 0.8$, $\beta = 0$, and $\gamma = 2.91$).[34] G_{np} is the non-polar solvation free energy calculated from the solvent-accessible surface area (SASA), according to $G_{\text{np}} = \gamma\text{SASA} + b$, where $\gamma = 0.0227$ kJ/mol/Å^2 and $b = 3.85$ kJ/mol.[35] Finally, S is an entropy estimate taken as a sum of translational, rotational and vibrational contributions. The translational and rotational entropies were estimated by statistical mechanical formulas of gas-phase molecules.[6,7] The vibrational entropy was estimated from a normal-mode analysis of a truncated and buffered system (8 + 4 Å).[24] The averages in Eqn 6, were evaluated at snapshots from molecular dynamics (MD) simulation of the complex, as is typical in MM/GBSA calculations.[6,7] The averages were calculated over 40 snapshots from 40 independent simulations, i.e., in total 1600 snapshots. All MM/GBSA calculations were performed with the Amber 10 suite of programs.[30]

Error estimates. All reported uncertainties are standard errors of the mean (standard deviations divided by the square root of the number of samples). The uncertainty of the BAR free energies calculated at each λ value was estimated by bootstrapping and the total uncertainty was taken as the square root of the sum of the squares of the individual uncertainties. For MM/GBSA, the reported standard error is the standard deviation of the mean over the 40 independent simulations (ignoring the standard deviation among the 40 snapshots in each simulation).

The performance of the free-energy estimates was quantified by the mean unsigned error (MUE), the correlation coefficient (r^2), and Kendall's rank correlation coefficient (τ) compared to experimental data. The latter was calculated only for the eight transformations that were explicitly studied, not for all combinations that can be formed from these transformations. The standard deviation of these quality measures was obtain by a simple simulation approach:[36] Each transformation was assigned a random number from a Gaussian distribution with the mean and standard deviation of the mean obtained from the BAR or MM/GBSA calculations. The quality measures (MUE, r^2, and τ) where then calculated and the procedure was repeated 1000 times. The standard error of these estimates is reported as the uncertainty.

Molecular dynamics simulations. All MD simulations were performed with the Q suite of programs.[19] Water molecules were subjected to both radial and polarisation restraints as implemented in Q. The former is a combination of half-harmonic and Morse potentials, and the latter is an implementation of the SCAAS model.[18] When simulating the truncated protein, solute atoms outside the simulation sphere were kept fixed at their initial positions using a strong harmonic restraint (837 kJ/mol/Å^2), and solute atoms in the outermost 2 Å shell were weakly restrained (84 kJ/mol/Å^2). When simulating the free ligand, the centre of mass of the ligand was weakly restrained (22 kJ/mol/Å^2) to the centre of the simulation sphere. In all simulations, the non-bonded cut-off was set to 10 Å, except between the ligand and the surroundings, for which no cut-off was applied. Long-range electrostatics were treated using a local reaction field (LRF) algorithm.[37] The non-bonded pair list was updated every 25th time step. The temperature was kept at 300 K using a Berendsen thermostat[38] with a 1 ps coupling time. The SHAKE algorithm[39] was used to constrain bonds involving hydrogen atoms and a 2 fs time step was used.

The simulations for the BAR calculations were performed as following: The system at λ = 1 was equilibrated, first using a 20 ps simulation in which all hydrogen atoms and water molecules were allowed to move, although they were restrained towards their starting position with a harmonic restraints of 105 kJ/mol/Å^2, and then by a 30 ps unrestrained simulation. Thereafter, the perturbation simulations were started. They consisted of 20 ps restrained equilibration, 200 ps unrestrained equilibration, and 1 ns production. Energy differences were sampled every tenth picosecond.

The MM/GBSA simulations were performed as following: 40 independent simulations were initiated by assigning different starting velocities. Each of these simulations consisted of a 20 ps simulation using the same restraints as described above, a 1 ns unrestrained equilibration, and a 200 ps production simulation. Snapshots for energy analysis were collected every 5 ps, and hence $40\times40=1600$ snapshots were used in the energy evaluation.[36]

Results and Discussion

Free energy estimates using full protein. We have carried out alchemical free-energy calculations to obtain the relative binding free energy of eight inhibitor pairs to fXa. Initially, the entire protein–complex as well as the free ligands were immersed in spherical water droplets with radii of 46 and 25 Å, respectively. The results from these calculations are shown in Table 1. It can be seen that the statistical precision of the calculations is excellent: The standard error is less than 0.3 kJ/mol for all of the eight transformations.

Four of the eight calculated free energy differences are within 1.4 kJ/mol of the experimental value and these also give the correct sign of the energy difference. Three of the remaining free energies are 4–5 kJ/mol from the experimental results, whereas the last one, the **63**→**39** transformation, gives an error of 14 kJ/mol. The calculated energy differences for all these four transformation also have the wrong sign, although the calculated result for the **47**→**5** transformation is not significantly different from zero. Consequently, the correlation coefficient (r^2) for all eight transformations is mediocre, 0.5, and τ is poor, 0.00. The MUE is only 3.9 kJ/mol, but this is actually slightly larger than the results of the null hypothesis that all transformations give a zero energy difference (3.5 kJ/mol).

Six of the transformations were included in our previous study, which was performed using periodic boundary conditions (PBC) with long-range electrostatics treated with Ewald summation. The free energies were calculated by TI with a dual topology, but the simulations were performed with the same force field as in the current study. Therefore, it is of interest to

compare the previous calculations with the current ones. From the results in Table 1, it can be seen that the new calculations have a much better precision (0.1–0.3 kJ/mol) than the PBC calculations (0.6–1.2 kJ/mol). There are at least two reasons for this. First, BAR gives a better precision than TI: For example, if we instead calculate the free-energy difference for the **125**→**53** transformation with the spherical simulations with the TI method, we obtain a standard error or 0.27 kJ/mol, instead of 0.06 kJ/mol with BAR. The remaining difference probably comes from the use of a dual topology in the PBC calculations, which is known to give a poorer precision than a single topology.[1] Naturally, an improved precision is desirable in free-energy estimations, provided that it not only an effect of a more restricted phase space sampling. However, in the present case, there is no indication that the spherical simulations sample a smaller phase space.

In particular, the results in Table 1 clearly show that the two simulation techniques give result that are very similar: All the calculated free-energy difference agree within 1 kJ/mol for the two simulations, except for the problematic **63**→**39** transformation, which gives a difference of 2 kJ/mol. In fact, none of the differences are statistically significant at the 95% level. This good agreement shows that the two simulation protocols give identical results. Moreover, it indicates that the results are reasonably converged and that the differences to experiments may be caused by deficiencies in the force field or by the uncertainty in the experimental data (unfortunately the experimental uncertainty was not reported,[23] but it is typically 2–4 kJ/mol).[40,41]

Effect of truncation. Although the calculations do not reproduce the experimental energy differences satisfactorily in several cases, this is a secondary issue for this study. The main objective is to investigate the effect of truncating the simulated system. How small can we make the system and still reproduce the results of the full simulation? Hence, our reference will not be the experimental data, but the simulations using full protein. Therefore, we made the simulated protein–ligand complexes systematically smaller by simulating systems with radii of 40, 35, 30, 25, 20, or 15 Å (the full system has a radius of 46 Å). The results of these simulations are shown in Table 2.

It can be seen that the truncations led to changes in the relative binding energies of the eight transformations of no more than 1.8 kJ/mol. In fact, in 79% of the cases, the difference is 0.4 kJ/mol or less. Looking at the individual electrostatics and van der Waals transformations (data not shown), this is the case for 79% of the transformations, and the maximum change is 1.2 kJ/mol for the electrostatics and 2.2 kJ/mol for the van der Waals transformations. There is a clear indication that the error increases slightly when the radius is decreased: The mean absolute deviation (MAD) increases from 0.2 to 0.5 kJ/mol going from a radius of 40 Å to 15 Å. Concomitantly, the standard deviation of the errors increases from 0.3 to 0.6 kJ/mol. Even if the errors are small, they are often statistically significant, owing to the high precision of the calculations: At a 40 Å radius, one of the transformations have statistically significant differences at the 98% level and this number increases up to five for the smallest radius.

Next, the size of the free-ligand simulations was reduced in a similar way to 20 or 15 Å. The results of these simulations are also shown in Table 2 and they show similar trends although the differences are smaller. The maximum difference is 0.4 kJ/mol and none of the differences are statistically significant at 98% confidence at any radius.

It should be noted that even if the total relative binding free energies are small for all transformations (0 to –4 kJ/mol; cf. Table 1), this is the sum of four individual terms (electrostatics and van der Waals transformation for the free and bound ligand), which are appreciably larger, –164 to 61 kJ/mol for the electrostatics terms and –4 to 9 kJ/mol for the van der Waals terms as is shown in the last four columns in Table 1. Therefore, the small

effects of the truncations are not caused only by the fact that the energy terms are small but rather because the simulations of the ligands in water and in the protein give similar results. Consequently, we can conclude that if average errors of 0.5 kJ/mol and maximum errors of 1.4 kJ/mol are acceptable, the simulated system can be truncated to a radius of 15 Å for both the protein and the free ligand, leading to a reduction of the number of atoms from 38844 to 1480 for the protein simulations.

Unfortunately, it is not possible to make the simulated systems smaller for two reasons. First, the potential that prevents the molecules from evaporating is not parametrised for spheres smaller than 12 Å.[19] Second, the considered inhibitors are rather large, with ~18 Å between the most distant atoms. This means that with a 15 Å radius, there are only two layers of water molecules outside the molecule, one of which is strongly affected by the SCAAS restraints. Therefore, it is not reasonable to make the sphere smaller.

Effect of the number of λ values. In our previous study of fXa, we found that the efficiency of the calculations can be considerably improved by simulating at fewer intermediate λ values.[8] Therefore, we calculated the free-energy differences also in this study with six (λ = 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0), five (λ = 0.1, 0.3, 0.5, 0.7, 0.9), three (λ = 0.0, 0.5, and 1.0), or two (λ = 0.0 and 1.0) λ values. The results of these calculations are collected in Table 3.

Considering the simulations with a 20 Å sphere first, the difference between six and eleven λ values is less than 1 kJ/mol for both the protein–ligand and the free-ligand simulations, as well as for the total binding free energy. The MAD for ΔG_{bind} over the eight studied transformations is only 0.1 kJ/mol.

With five λ values, the difference in ΔG_{bind} increases to 0.3 kJ/mol on average, with a maximum of 1.3 kJ/mol for the **53**→**47** transformation. In variance to the other number of λ values, the calculations with five λ values involves the extrapolation of the results at λ = 0.1 and λ = 0.9 to that of λ = 0 and λ = 1. We tested various extrapolation schemes, but a simple linear extrapolation involving two points worked best and is therefore used for the results in Table 3.

With three λ values, MAD for ΔG_{bind} increases to 0.5 kJ/mol and the maximum difference increases to 1.4 kJ/mol. However, the individual differences for the free and bound simulations are even larger, up to 2.0 kJ/mol for the **53**→**47** transformation.

If only the two end points are simulated (λ = 0.0 and 1.0), the results deteriorate significantly: The MAD increases to 0.9 kJ/mol and five of the simulations give errors that are larger than 1 kJ/mol, with a maximum of 3.1 kJ/mol for the **47**→**5** transformation. These deviations are probably too large to be acceptable in most applications. Therefore, we tend to recommend calculations with three λ values.

The results with a 15 Å sphere are analogous with MADs for ΔG_{bind} of 0.2, 0.4, 0.2, and 1.0 kJ/mol for six, five, three, and two λ values, respectively. There is no indication that the smaller sphere gives worse results. This clearly shows that the efficiency of the BAR calculations can be increased by using three λ values.

The single-transformation approach (STA). All results up to now were obtained with the two-transformation approach (TTA), in which the charges are first transformed in one set of simulations and then the van der Waals parameters are transformed in a second set of simulations. However, both transformations can be done in a single set of simulations, provided that soft-core Coulomb potentials are employed. We have previously shown that such an approach can improve both the precision and accuracy of TI calculations of binding affinities.[8] Therefore, a soft-core Coulomb potential was implemented into the Q software in analogy with the soft-core Lennard-Jones potential. We assume that the conclusions drawn above hold true also for this STA approach and therefore present results only with 20 or 15 Å

simulation spheres, and with eleven, six, five, three, and two λ values. These results are collected in Table 4.

The STA results are very similar to those of obtained with TTA, as expected. With eleven λ values, the MAD between the TTA and STA calculations is only 0.8 and 0.5 kJ/mol for the 20 and 15 Å spheres, respectively. The largest deviation, –1.9 or –1.3 kJ/mol, is found for the problematic **63→39** transformation, with the STA results being slightly closer to the experiments. The precision of the STA is similar to of the TTA, with average standard errors of 0.19 kJ/mol (0.13 kJ/mol without the **63→39** transformation which gives a twice as large standard error than the other transformations, 0.6 kJ/mol).

The calculations with fewer λ values give similar results to those with eleven λ values also with the STA approach. The largest error (3.1 kJ/mol) is found with two λ values. With three λ values, the maximum error is 1.8 kJ/mol and the MAD is 0.5 kJ/mol.

MM/GBSA calculations. It is of interest to compare the results of the rigorous BAR calculations with those of an approximate method such as MM/GBSA. We assumed that it is sufficient to use a small simulation sphere and therefore performed the MD simulations using a 20 Å protein sphere and post-processed them to obtain MM/GBSA estimates (however, note that all protein residues were included in the calculations, although those outside the 20 Å radius were kept fixed at the crystal structure in the MD simulations). The results are collected in Table 5.

The primary product of MM/GBSA is the absolute affinities of the ten inhibitors involved in the eight transformations in Figure 1, so these are shown in the first part of Table 5. Comparing with experiments, the MM/GBSA method gives estimates that are too negative by 27 kJ/mol on average. This has been observed several times before and the shift depends on the details of the calculations, in particular the continuum-solvation method.[9] However, the τ is rather good (0.49). Moreover, the MAD after removal of the systematic error is only 4 kJ/mol, although the null hypothesis that all inhibitors have the same affinity gives the same result. The correlation coefficient r^2 is 0.35.

Compared to MM/GBSA calculations with PBC and the full system,[8] the present calculations give slightly more negative affinities, with differences of 2–15 kJ/mol (average 10 kJ/mol). The correlation coefficient (r^2) between the two sets of calculations is 0.54. Looking at the individual terms in Eqn 6 (not shown), the electrostatic energy shows the largest difference with a mean signed difference (MSD) of –38 kJ/mol. The van der Waals energy and polar solvation show intermediate MSDs with the opposite sign, 19 and 18 kJ/mol, respectively. The other two terms show only minor differences. As the difference between the two simulations is mainly a shift in the absolute affinities, they reproduce the experimental results equally well (there are no statistically significant difference in the three quality measures, MUE, r^2, and τ). The new calculations give slightly higher standard errors (1.1–2.0 kJ/mol) than the PBC calculations (0.8–1.6 kJ/mol).

Next, we computed the binding-affinity differences for the transformations in Figure 1, i.e. those calculated by BAR. Most of the predications are similar to the experimental difference. The correlation coefficient ($r^2 = 0.07 \pm 0.12$) is significantly worse than for the BAR calculations (0.50 ± 0.18) at the 95% level. However, the MUEs (4 kJ/mol) of the two methods are similar and τ of MM/GBSA is actually better because the sign is correct for five of the transformations, but the difference is not statistically significant. The most conspicuous difference between the BAR and MM/GBSA results is the much worse precision of the MM/GBSA results (1.8–2.8 kJ/mol, compared to 0.06–0.36 kJ/mol). This indicates that 60–900 times more simulations are needed to reach the same precision of the MM/GBSA results, showing that BAR is a more effective method (i.e. much less computational effort is needed to reach the same precision of the calculated affinities).

There is a fair correlation (r^2 = 0.47) between the MM/GBSA relative energies calculated with a spherical system and the previous PBC simulations. However, the individual free-energy differences differ by up to 6 kJ/mol, which is larger than expected from the estimated standard errors (three of the six transformations show statistically significant differences at the 95% level). Thus, even if the differences between the two sets of simulations is mainly a constant shift, there are also some systematic differences that make the results significantly different, although the difference is not larger than 6 kJ/mol.

Timings. The 1.2 ns simulations (0.2 ns equilibration and 1 ns production) for the protein–ligand complexes take ~915, ~36 and ~14.5 CPU hours on a single processor (Intel Xeon 2.26 GHz) using the 46, 20 and 15 Å simulation spheres, respectively. Thus, the calculation of one binding free energy difference takes 40260 CPU hours for the full TTA reference calculations with eleven λ values (44 simulations needed), 432 or 174 CPU hours with TTA and three λ values (12 simulations), and 216 or 87 CPU hours with STA (6 simulations), if we employ a 20 or 15 Å radius, respectively. Of course all the individual simulations can be run in parallel.

Comparing with our previous computations using periodic boundary conditions,[8] the most optimal protocol (using STA) took 337 CPU hours. Thus we can conclude that the truncation increased the efficiency by a factor of ~1.6 or ~3.8 if we use a 20 or 15 Å sphere, respectively. The relative modest increase in efficiency is caused by the fact that the PBC calculations employed particle-mesh Ewald calculations for the long-range electrostatics, which allows for the use of a smaller cut-off radius for these time-consuming interactions. Compared to the full calculations with a spherical system and eleven λ values, we gain a speed-up of 460 when using only three λ values and a 15 Å radius.

How general are the results? A natural question is how general the results in this paper are. All the studied transformations preserve the charge of the ligand (+1 or +2 *e*) and they are small, involving the transformation of a hydrogen atom to a heavy atom, with 0–3 hydrogen atoms, except for the **53**→**125** (O→NH_2) and **63**→**39** (H→OCH_3) transformations. Moreover, all the transformations take place on the surface of the protein with the R_1, R_2, and X sites (Figure 1) pointing mainly out into the solution, whereas sites R_3 and R_4 are still on the surface, but interacting somewhat more with the protein, as can be seen in Figure 2. The second, unperturbed, amidinobenzyl group of the ligand interacts with Asp-189 inside the protein, but the rest of the binding site is also quite polar (Figure 2).

Clearly, the possibility to truncate the simulated system is affected by type of the transformations. The dipole moments (with respect to the centre of the mass) of the studied ligands are 1.0–1.4 D. Consequently, the difference in the Onsager solvation energy caused by the various transformations is quite small and shows a cubic dependence on the radius of the simulated sphere, as can be seen in Figure 3a. At 15 Å, it is negligible, <0.01 kJ/mol. On the other hand, the total dipole moment of the protein–ligand complex (with respect to the centre of the simulated sphere) is appreciably larger, so the Onsager correction at 15 Å is up to 0.4 kJ/mol for the studied transformations (0.2 kJ/mol at 20 Å), as is also shown in Figure 3a. This may explain the variation of the results for the smallest systems, but it is unlikely that a simple Onsager correction will improve the results, because it assumes a uniform dielectric constant both for the removed protein and solvent. Instead, more sophisticated methods[20,21,22] are needed if an accuracy better than 1 kJ/mol is needed or if you want to study even smaller systems.

Dipole–dipole interactions show a similar cubic distance dependence, giving negligible contributions at 15 Å distance (<0.03 kJ/mol). On the other hand, charge–dipole interactions show a quadratic dependence on the distance and are still noticeable at 15 Å, up to 0.5 kJ/mol (Figure 3a). However, for solvent-exposed charges, they are typically scaled

down by solvation and dynamic effects, as is manifested by an effective dielectric constant of 20 or more.[42] The only buried charges in fXa, Asp-189 interacting with the ligand, Asp-102 of the catalytic triad, and Asp-194, forming an ionic pair with the amino terminal, are all close to the ligand and therefore included also in the smallest truncated system. Thus, the distance-dependence of the various electrostatic interactions confirms and explains why the truncations work well in the present systems. Such an investigation can easily be performed for any sets of ligands and proteins to estimate how large truncations are possible. In particular, Figure 3b shows that if the transformations involve changes in the net charge of the ligand, the expected size of the interactions increases by several orders of magnitude, so that the Born solvation term and the charge–charge interactions do not become negligible even with simulated systems of the size of 100 Å. This explains why such transformations are much harder to study with alchemical free-energy methods.[1,8]

The validity of the reduction in the number of λ values can be checked with standard FEP convergence methods, e.g. by the hysteresis of the FEP results, the difference between FEP and BAR estimates, or by more sophisticated overlap measures.[43,44] In a forthcoming publication, we will examine our suggested method for the binding of over 100 ligands to ten different proteins.

Conclusions

In this paper, we study how the binding free-energy difference for eight pairs of 3-amidinobenzyl-1*H*-indole-2-carboxamide inhibitors to blood clotting factor Xa, calculated by alchemical free-energy calculations depends on the size of the simulated system. We have shown that calculations of the entire protein in a spherical water droplet reproduce free energies that were obtained using periodic boundary conditions within statistical precision. In fact, the new calculations, obtained with BAR and a single topology, rather than TI with a dual topology, give a much better precision with the same length of the simulations (0.2 + 1.0 ns) and a similar number of intermediate states (11 or 9), 0.06–0.29 kJ/mol, compared to 0.6–1.2 kJ/mol.

Second, we have systematically truncated the spherical system, by removing water molecules and ignoring interactions with protein residues outside a certain radius. We show that we can reduce the radius of the simulated system from 46 Å down to 15 Å without changing the calculated free-energy differences by more than 0.5 kJ/mol on average (maximum change 1.4 kJ/mol). This is a quite amazing result, showing that the simulated system can be reduced from 38844 to 1480 atoms without changing the calculated free-energy differences by more than 1 kJ/mol. Neither any sophisticated model of the removed parts of the surrounding is needed, nor any account of long-range electrostatic effects. Only some restraints on the water molecules at the surface of the simulated sphere are employed. On the other hand, the results are in good agreement with previous studies showing that only residues within 12–16 Å of the active site of an enzyme need to be considered when studying chemical reactions.[45,46]

Third, we have investigated how many intermediate states have to be included in the BAR calculations. We showed that only a single intermediate state ($\lambda = 0.5$) needs to be simulated, if average and maximum deviations of 0.5 and 1.4 kJ/mol are acceptable.

Fourth, we have implemented soft-core Coulomb potentials into the Q package, which allows us to do the electrostatic and van der Waals perturbations in a single step (STA), rather than in two separate steps (TTA). The results of the two approaches are closely similar (0.5–0.8 kJ/mol average difference), except for the largest transformation (2 kJ/mol difference), for which the STA result is closer to the experiments. Also for STA it is possible to employ only three λ values in the calculations.

Finally, we have compared the BAR results with those obtained with the MM/GBSA

method. The latter results gives a slightly worse correlation coefficient to the experimental results than BAR, but in particular they have a much worse precision, meaning that much more simulations are needed to obtain MM/GBSA results of an equal quality.

Altogether, these results indicate that alchemical free-energy calculations BAR are a valuable tool that could be used in drug design to calculate relative binding affinities of drug candidates with the same scaffold and the same net charge. Estimates with a precision of 0.1–0.6 kJ/mol can be obtained from six simulations that can be run in parallel on a single processor within 15 h for small transformations preserving the net charge of the ligand.

Acknowledgements
We thank Johan Åqvist and coworkers for help with the Q software package. This investigation has been supported by grants from the Swedish research council (project 2010-5025) and from the FLÄK research school in pharmaceutical science at Lund University. It has also been supported by computer resources of Lunarc at Lund University, NSC at Linköping University, C3SE at Chalmers University of Technology, and HPC2N at Umeå University.

Supporting Information Available: Amber topology and parameter files for the ten ligands. This material is available free of charge via the Internet at http://pubs.acs.org.

1 Michel, J.; Essex, J. W. *J. Comput.-Aided Mol. Des*. **2010**, *24*, 639-658.
2 Chipot, C.; Rozanska, X.; Dixit, S. B. *J. Comput.-Aided Mol. Design.* **2005**, *19*, 765-770.
3 Chipot, C.; Pohorille, A. Eds. *Free Energy Calculations*. **2007**, Springer, New York.
4 Gohlke, H.; Klebe, G. *Angew. Chem. Int. Ed.* **2002**, *41*, 2644-2676.
5 Foloppe, N.; Hubbard, R. *Curr. Med. Chem.* **2006**, *13*, 3583-3608.
6 Srinivasan, J.; Cheatham III, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. *J. Am. Chem. Soc.* **1998**, *37*, 9401-9809.
7 Kollman, P., A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham III, T.; E., *Acc. Chem. Res.* **2000**, *33*, 889-897.
8 Genheden, S.; Nilsson, I.; Ryde, U. *J. Chem. Inf. Model.* **2011**, *51*, 947–958.
9 Genheden, S.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Ryde, U. *J. Phys. Chem. B* **2010**, *114*, 8505-8516.
10 Genheden, S.; Kongsted, J.; Söderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2010**, *6*, 3558–3568.
11 Genheden, S.; Mikulskis, P.; Hu, L.; Kongsted, J.; Söderhjelm, P.; Ryde, U. *J. Am. Chem. Soc.* **2011**, *133*, 13081–13092.
12 Åqvist, J.; Medina, C.; Samuelsson, J. E. *Prot. Eng.* **1994**, *7*, 385-391.
13 Genheden, S.; Ryde, U. *J. Chem. Theory Comput.* **2011**, *7*, 3768-3778.
14 Mikulskis, P.; Genheden, S.; Rydberg, P.; Sandberg, L.; Olsen, L.; U. Ryde *J. Comput.-Aided Mol. Design*, **2012,** in press, DOI:10.1007/s10822-011-9524-z.
15 Essex, J. W.; Jorgensen, W. L. *J. Comput. Chem.* **1995**, *16*, 951-972.
16 Berkowitz, M.; McCammon, J. A. *Chem. Phys. Lett.* **1982**, *90*, 215-217.
17 Brünger, A. T.; Brooks III, C. L., Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495-500.
18 King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647–366.
19 Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. *J Mol. Graph. Model.* **1998**, *16*, 213-225.

20 Im, W.; Bernèche, S.; Roux, B. *J. Chem. Phys.* **2001**, *114*, 2924-2937.
21 Banavali, N. K.; Im, W.; Roux, B. *J. Chem. Phys.* **2002**, *117*, 7381-7388.
22 Simonson, T.; Archontis, G.; Karplus, M. *J. Phys. Chem.* **1997**, *101*, 8349-8362.
23 Matter, H.; Defossa, E.; Heinelt, U.; Blohm, P.-M.; Schneider, D.; Müller, A.; Hreok, Si.; Schreuder, H.; Liesum, A.; Brachvogel, V.; Lönze, P.; Walser, A.; Al-Obeidi, F.; Wildgoose, P. *J. Med. Chem.* **2002**, *45*, 2749-2769.
24 Kongsted, J.; Ryde, U. *J. Comput-Aided Mol. Design.* **2009**, *23*, 63-71.
25 Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem Soc.* **1995**, *117*, 5179-5197.
26 Wang, J. M.; Wolf, R. M.; Caldwell, K. W.; Kollman, P. A., Case, D. A. *J. Comput. Chem.*, **2004**, *25*, 1157-1174.
27 Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
28 Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431-439.
29 Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impley, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.
30 Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R., Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K., F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R., Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *Amber 10*, University of California, San Francisco, **2008**.
31 Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J*. **1997**, *72*, 1047-1069.
32 Bennett C. H. J. *Comput. Phys.* **1976**, *22*, 245–268.
33 Zacharias, M.; Straatsma, T. P; McCammon, J. A *J. Chem. Phys*. **1994**, *100*, 9025-9031.
34 Onufriev, A.; Bashford, D.; Case, D. A. *Proteins* **2004**, *55*, 383-394.
35 Kuhn, B.; Kollman, P. A. *J. Med. Chem.* **2000**, *43*, 3786-3791.
36 Genheden, S.; Ryde, U. *J. Comput Chem.* **2010**, *31*, 837-846.
37 Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100–3107.
38 Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
39 Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.*, **1977**, *23*, 327-341.
40 Gilson, M. *Ann. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21-42.
41 Brown, S. P.; Muchmore, S. W.; Hajduk, P. J. *Drug Discov. Today* **2009**, *14*, 420-427.
42 Schutz, C. N.; Warshel, A. *Proteins* **2001**, *44*, 400-417.
43Bhattacharyya, A. *Bull. Cal. Math. Soc.* **1943**, *35*, 99-109.
44 Wu, D.; Kofke, D. A.; *J. Chem. Phys.* **2005**, *123*, 054103.
45 Kaukonen, M.; Söderhjelm, P.; Heimdal, J.; Ryde, U. *J. Chem. Theory Comput.* **2008**, *4*, 985-1001.
46 Hu, L.; Eliasson, J.; Heimdal, J.; Ryde, U. *J. Phys. Chem. A* **2009**, *113*, 11793–11800.

Table 1. Relative binding free energies (in kJ/mol) for the eight pairs of fXa inhibitors in Figure 1, calculated in a spherical water droplet including the full protein–ligand complex and compared to experimental data (Exp),[23] as well as to previous calculation in a periodic octahedral box (PBC).[8] [a]

	Spherical	Exp	PBC	ΔG^{el}_{bound}	ΔG^{el}_{free}	ΔG^{vdW}_{bound}	ΔG^{vdW}_{free}
125→53	-0.15 ±0.06	-1.0	-1.3 ±0.9	-14.71	-14.54	0.49	0.47
53→9	-0.64 ±0.12	-1.9	-0.2 ±0.7	4.39	4.02	3.98	4.99
53→47	-1.14 ±0.22	-2.5	-0.7 ±0.6	-37.41	-37.22	0.13	1.08
53→49	-1.05 ±0.10	2.5	-1.6 ±0.8	-110.14	-109.83	-1.35	-0.61
53→50	-0.43 ±0.11	-1.9	-0.9 ±0.8	-50.61	-50.38	-0.80	-0.59
53→51	-0.56 ±0.12	3.5		-163.49	-163.55	-1.30	-0.68
63→39	-3.78 ±0.18	10.1	-2.0 ±1.2	60.71	60.48	-1.24	2.77
47→5	-0.18 ±0.29	4.9		35.52	34.08	9.07	10.69
MUE	3.94 ±0.06 (0.73)		3.5 ±0.3 (0.8)				
r^2	0.47 ±0.07 (0.17)		0.69 ±0.29 (0.28)				
τ	0.04 ±0.04 (0.19)		0.33 ±0.26 (0.36)				

[a] The four last columns show the results of the four perturbations contributing to the calculated free energies of the spherical system, viz. the electrostatics and van der Waals perturbations, obtained with the ligand bound to the protein or free in solution. The net binding free energy is $\Delta G_{\text{bind}} = \Delta G^{\text{el}}_{\text{bound}} - \Delta G^{\text{el}}_{\text{free}} + \Delta G^{\text{vdW}}_{\text{bound}} - \Delta G^{\text{vdW}}_{\text{free}}$. Standard errors for the quality measures assume that the experimental data are exact, whereas the values in bracket were obtained by assuming a typical precision of 2.4 ($1.7\sqrt{2}$) kJ/mol for the experimental data.[41]

Table 2. Deviations of calculated free-energy differences (kJ/mol) for the various transformations using smaller systems compared to the simulation with the full protein (46 Å) or a 25 Å sphere for the free ligand.[a]

	Protein simulations						Free-ligand simulations	
Radius (Å)	40	35	30	25	20	15	20	15
125→53	0.1	0.2	0.2	0.4	0.2	-0.3	-0.1	0.0
53→9	0.2	0.1	0.3	0.3	0.2	0.4	0.4	0.4
53→47	-0.1	0.1	0.1	0.7	**-1.8**	0.1	0.0	0.2
53→49	-0.2	-0.1	-0.1	-0.1	-0.4	**-1.4**	-0.1	0.1
53→50	0.4	0.1	0.0	0.3	-0.2	0.1	-0.3	-0.1
53→51	0.2	0.2	-0.1	-0.1	-0.3	-0.8	0.0	-0.2
63→39	0.0	0.6	0.1	0.0	0.3	-0.4	0.3	-0.2
47→5	-0.6	-0.9	0.0	-0.7	**-1.3**	-0.7	0.1	0.4
MAD[b]	0.2	0.3	0.1	0.3	0.6	0.5	0.2	0.2
# atoms[c]	27303	18239	11553	6554	3528	1480	3415	1375

[a] If the deviation is negative, the simulations with a smaller sphere give a more positive energy.

[b] Mean absolute deviation compared to a full protein simulation or a simulation sphere of 25 Å for the free ligand.

[c] Number of atoms that are allowed to move in the simulated system (the number of atoms in the full systems are 38844 and 6319 for the protein and free ligand simulations, respectively). The numbers apply for the largest ligand **39**.

Table 3. Effect of using fewer λ values.[a]

# λ	6			5			3			2		
	bound	free	ΔG_{bind}	bound	free	ΔG_{bind}	bound	free	ΔG_{bind}	bound	free	ΔG_{bind}
Radius = 20 Å												
125→53	-0.1	0.0	0.0	0.1	0.1	0.1	-0.2	0.0	-0.2	0.0	0.1	-0.1
53→9	0.6	0.5	-0.1	0.2	-0.6	-0.7	0.7	**1.4**	0.6	0.1	**3.0**	**2.8**
53→47	-0.9	-0.5	-0.4	-0.8	**-1.3**	0.5	**-2.0**	**-1.5**	-0.5	**-3.1**	**-2.1**	-1.0
53→49	0.0	0.1	-0.2	0.0	-0.1	0.1	-0.2	-0.1	-0.1	-0.6	0.3	-0.9
53→50	0.0	0.0	0.0	0.1	0.0	0.0	0.0	-0.2	0.2	0.7	0.4	0.3
53→51	0.0	0.2	-0.1	0.0	-0.1	0.2	0.6	0.1	0.4	0.8	0.6	0.3
63→39	-0.1	-0.3	0.1	-0.4	-0.7	0.3	0.4	-0.5	0.9	**1.5**	0.1	**1.4**
47→5	-0.7	-0.5	-0.2	-0.7	0.0	-0.7	-0.3	**-1.7**	**1.4**	-0.9	**-1.4**	0.5
MAD[b]	0.3	0.3	0.1	0.3	0.4	0.3	0.6	0.7	0.5	1.0	1.0	0.9
Radius = 15 Å												
125→53	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
53→9	0.4	0.2	-0.1	0.4	-0.4	-0.8	0.9	**1.3**	0.5	-0.1	**2.4**	**2.5**
53→47	-0.5	-0.6	0.1	-0.8	**-1.4**	0.6	**-1.7**	**-1.4**	-0.3	**-2.4**	**-2.1**	-0.3
53→49	0.0	-0.1	0.1	0.1	0.2	-0.1	-0.4	0.2	-0.7	-0.5	**1.2**	**-1.6**
53→50	0.1	-0.1	0.2	0.0	0.2	-0.2	0.1	0.0	0.2	0.6	0.3	0.4
53→51	-0.1	-0.1	0.0	0.2	0.1	0.1	-0.1	-0.2	0.0	0.3	0.1	0.2
63→39	-0.4	-0.1	-0.3	0.3	-0.4	0.7	0.6	0.6	0.1	-0.9	1.0	**-1.9**
47→5	**-1.1**	-0.4	-0.7	-0.8	0.1	-0.9	**-1.6**	**-1.8**	0.2	**-1.6**	-0.7	-1.0
MAD[b]	0.3	0.2	0.2	0.3	0.3	0.4	0.7	0.7	0.2	0.8	1.0	1.0

[a] For each radius, the difference (kJ/mol) in the relative binding free energy for the eight transformations between the calculation with 11 and with fewer λ values is presented. Deviations larger than 1 kJ/mol are highlighted in bold face. Results are presented both for the protein–ligand complex (bound) and for the free-ligand (free) simulations, as well as for their difference (ΔG_{bind}). In each case, the difference between 11 and fewer λ values are taken.
[b] Mean absolute deviation compared to calculations using 11 λ values.

Table 4. Calculated relative binding free energies using the single-transformation approach (kJ/mol).[a]

	Calculated		ΔTTA[b]	ΔSTA[c]	ΔSTA[c]	ΔSTA[c]	ΔSTA[c]
# λ	11		11	6	5	3	2
Radius = 20 Å							
125→53	-0.34	±0.04	-0.1	-0.1	0.1	-0.1	0.1
53→9	0.29	±0.15	**-1.2**	0.5	0.4	0.7	0.6
53→47	1.12	±0.26	-0.4	**1.2**	**1.0**	-0.5	**1.8**
53→49	-1.48	±0.09	0.8	0.0	0.0	-0.3	-0.6
53→50	-0.98	±0.08	0.4	-0.1	0.1	-0.2	-0.2
53→51	-0.63	±0.09	0.4	0.2	-0.1	-0.3	0.2
63→39	-1.91	±0.56	**-1.9**	-0.1	0.1	**-1.6**	**-3.1**
47→5	0.11	±0.21	**1.1**	0.3	-0.2	-0.4	-0.4
MUE	4.04	±0.09 (0.70)	0.8	0.3	0.3	0.5	0.9
r^2	0.40	±0.11 (0.16)					
τ	-0.48	±0.06 (0.26)					
Radius = 15 Å							
125→53	0.05	±0.04	0.2	0.0	0.0	-0.1	-0.4
53→9	0.02	±0.14	-0.6	0.0	-0.3	0.3	0.3
53→47	-1.02	±0.23	0.0	0.9	0.1	0.1	0.4
53→49	0.18	±0.10	0.2	-0.2	0.3	-0.7	**-2.6**
53→50	-0.86	±0.07	0.3	-0.1	0.1	0.2	0.2
53→51	-0.28	±0.08	0.3	-0.1	0.1	-0.2	-0.5
63→39	-2.33	±0.63	**-1.3**	-0.1	-0.5	**-1.8**	**-1.0**
47→5	0.10	±0.20	0.8	0.0	-0.1	0.5	0.3
MUE	3.60	±0.09 (0.72)	0.5	0.2	0.2	0.5	0.7
r^2	0.22	±0.10 (0.16)					
τ	0.19	±0.13 (0.18)					

[a] The MUE, r^2, and τ values are calculated with respect to the experimental data.[23] The standard errors of these estimates were obtained assuming that the experimental data are exact, whereas those in brackets were obtained by assuming an uncertainty of 2.4 kJ/mol.[41]
[b] Deviation compared to the TTA calculations (cf. Table 2) using 11 λ values. If the deviation is negative, TTA gives a more negative energy.
[c] Deviation compared to using 11 of λ values with the STA. If the deviation is negative, the simulation with less λ values gives a more positive energy.
.

Table 5. MM/GBSA estimates of the ten inhibitors and the eight transformations in Figure 1.[a]

	Calculated		Exp.	PBC	
5	-72.3	±1.7	-41.9		
9	-79.1	±1.2	-46.2	-65.5	±0.9
39	-63.9	±2.0	-27.3	-49.4	±0.9
47	-73.1	±1.1	-46.8	-58.8	±1.1
49	-61.7	±1.5	-41.9	-59.4	±1.0
50	-67.3	±1.3	-46.2	-57.8	±0.8
51	-65.9	±1.3	-40.9		
53	-70.4	±1.4	-44.3	-62.5	±1.0
63	-63.4	±1.9	-37.4	-52.2	±1.6
125	-71.4	±1.2	-43.4	-63.4	±1.0
MADtr	3.8	±0.4 (0.6)		3.1	±0.4 (0.6)
r^2	0.35	±0.10 (0.12)		0.67	±0.06 (0.10)
τ	0.49	±0.08 (0.12)		0.40	±0.08 (0.13)
125→53	1.0	±1.8	-0.9	0.9	±1.4
53→9	-8.7	±1.8	-1.9	-3.0	±1.4
53→47	-2.7	±1.8	-2.5	3.7	±1.5
53→49	8.6	±2.0	2.4	3.1	±1.4
53→50	3.1	±1.9	-1.9	4.7	±1.4
53→51	4.5	±1.9	3.4		
63→39	-0.6	±2.8	10.1	2.8	±1.6
47→5	0.8	±2.0	4.9		
MUE	4.5	±0.7 (1.0)		3.9	±0.5 (1.0)
r^2	0.07	±0.09 (0.12)		0.03	±0.09 (0.12)
τ	0.25	±0.23 (0.29)		0.00	±0.17 (0.32)

[a] Free energies in kJ/mol were estimated using a 20 Å simulation sphere. The results in the upper part of the table are the MM/GBSA absolute affinities of the ten inhibitors, whereas those in the lower part are the differences for the eight transformations. Experimental[23] and previous results obtained with periodic boundary conditions (PBC)[8] are also included.

Figure 1. Ligands and transformations considered in this study. The three groups in brackets in the upper part of the figure are the R_1, R_2, R_3 groups, whereas the single group in the lower part is the R_4 group.

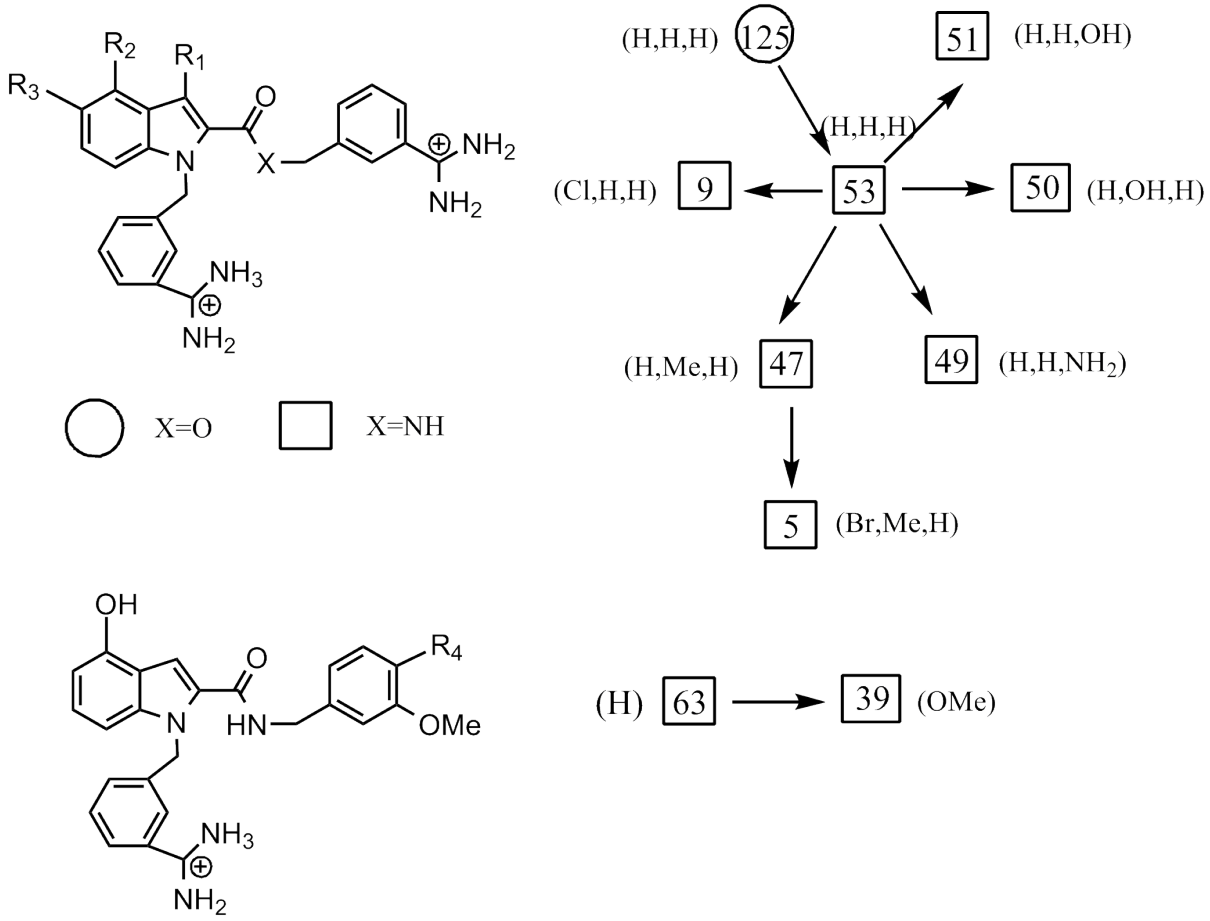

Figure 2. The binding site of factor Xa with ligand **39** bound. The protein is shown as a space-filling model with regions with a negative (red) and positive (blue) electrostatic potential marked. The five perturbed sites of the ligand are marked with balls, from the right R_1, R_2, R_3, X, and R_4 (cf. Figure 1).

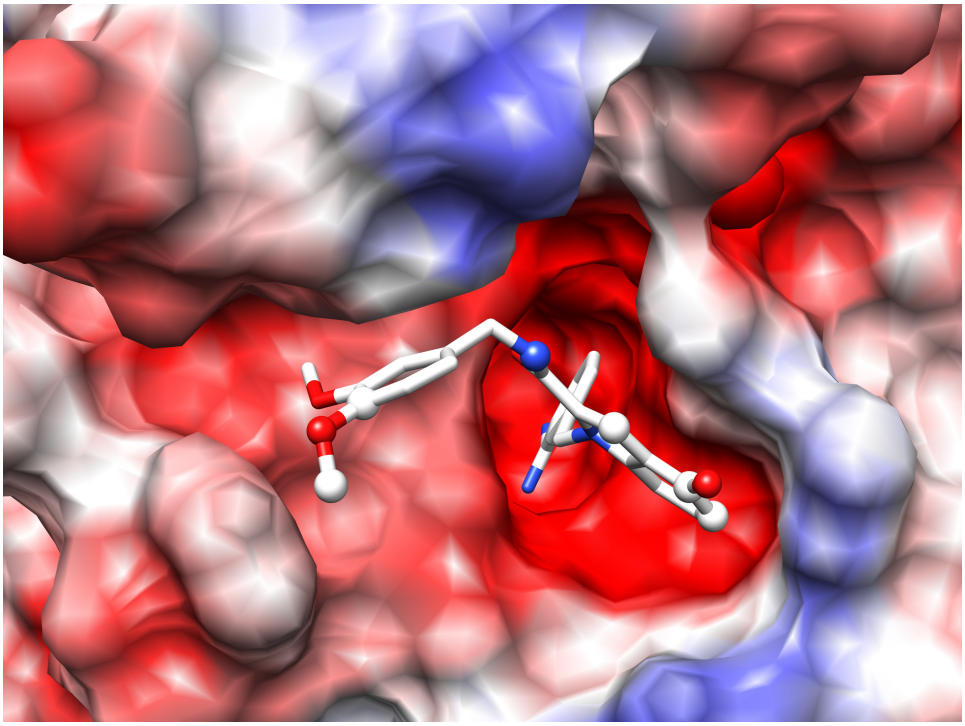

Figure 3. Distance dependence of various interactions: a) The Onsager solvation energy for the bound and free ligand, dipole–dipole, and dipole-charge, as well as the b) charge–charge, interaction and the Born solvation energy. The energies are calculated for a dipole change from 1.4 to 1.0 D of the ligand (36.5 to 25.9 D for the bound ligand), a water dipole of 1.85 D, and a charge of +1 *e*. The dielectric constant was always assumed to be 80. Note the different energy scale in the two figures.