Journal of Computer-Aided Molecular Design **17:** 673–686, 2003.

Sampling and convergence in free energy calculations of protein-ligand interactions: The binding of triphenoxypyridine derivatives to factor Xa and trypsin

Alessandra Villa, Ronen Zangi, Gilles Pieffet & Alan E. Mark
The Groningen Biomolecular Sciences and Biotechnology Institute, Department of Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

Received 21 July 2003; Accepted in revised form 26 September 2003

Summary

The binding of a set of 10 triphenoxypyridine derivatives to two serine proteases, factor Xa and trypsin, has been used to analyze factors related to sampling and convergence in free energy calculations based on molecular dynamics simulation techniques. The inhibitors investigated were initially proposed as part of the Critical Assessment of Techniques for Free Energy Evaluation (CATFEE) project for which no experimental results nor any assessment of the predictions submitted by various groups have ever been published. The inhibitors studied represent a severe challenge for explicit free energy calculations. The mutations from one compound to another involve up to 19 atoms, the creation and annihilation of net charge and several alternate binding modes. Nevertheless, we demonstrate that it is possible to obtain highly converged results ($\pm$ 5–10 kJ/mol) even for such complex multi-atom mutations by simulating on a nanosecond time scale. This is achieved by using soft-core potentials to facilitate the creation and deletion of atoms and by a careful choice of mutation pathway. The results show that given modest computational resources, explicit free energy calculations can be successfully applied to realistic problems in drug design.

Introduction

Estimating differences in free energy is central to the process of rational molecular design. This is because all equilibrium properties of a system such as phase behavior, association-dissociation constants, solubilities, adsorption coefficients and conformational equilibria depend on differences in free energy between alternative states. Free energy differences are essentially related to the relative probability of finding a system in a given microscopic state. Many empirical approaches have been developed to estimate interaction or binding free energies between proteins and ligands. However, only by using an approach that samples an appropriate thermodynamic ensemble of states, such as Molecular Dynamics (MD) and Monte Carlo (MC) simulation techniques, from which it is possible to get thermal averages over microscopic configurations at an atomic level, can differences in free energy between two states of a system be estimated directly [1–7]. The difficulty is that the computational cost of obtaining sufficient sampling and converged results has made the routine application of free energy calculations for estimating binding free energies impractical. This situation is, however, rapidly changing. The use of modified intermediate potentials has been shown to improve sampling dramatically and the rapid advance of computer power means that the utility of free energy calculation in molecular design must be constantly re-evaluated.

In this paper, the relative binding affinities of a set of 10 inhibitors to two serine proteases, factor Xa and trypsin, that share sequence and structural homology, are evaluated. Factor Xa activates thrombin and plays a regulatory role in blood coagulation. Thus, it has been a target for the design of anti-thrombotics agents. However, because the active sites of many serine proteases are very similar, it is important that inhibitors

*To whom correspondence should be addressed. E-mail: A.E.Mark@chem.rug.nl

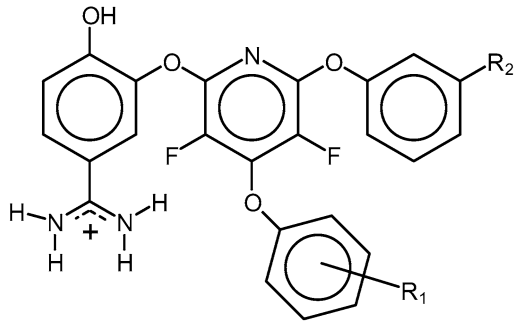

Figure 1. The generic chemical structure of the inhibitors, 2,4,6-triphenoxypyridine. The substituents R_1 and R_2 for each inhibitor are specified in Table 1.

selectively bind to factor Xa and not to other serine proteases, such as trypsin or thrombin.

The set of inhibitors studied in this work were initially proposed as part of the Critical Assessment of Techniques for Free Energy Evaluation (CATFEE) project[1]. This competition, scheduled for summer 2001, was intended as a blind test of approaches to estimate free energies. However, although predictions were submitted by various groups around the world, neither the corresponding experimental data nor any objective assessment of the predictions were ever published by the organizers.

The inhibitors analyzed in this study are based on a triphenoxypyridine template (Figure 1) but differ in the number and in the type of the substituents on two of the three phenyl groups (Table 1). The inhibitors differ significantly from each other and the transformation from one target into another involves mutating many more sites compared with cases normally dealt with in free energy calculations [3, 4, 8–10]. Therefore, the mutations are highly challenging and problems of sampling and convergence are of major concern in the calculations.

Since no experimental data has been made available, discussion of the absolute errors in the calculation is not possible. Instead, in this manuscript we only consider issues related to sampling and convergence. We primarily looked at self consistency, exploiting the fact that the free energy is a state function so the difference in the free energy between two states is independent of the path chosen. In addition, by publishing the ranking of the binding affinities of the 10 inhibitors that we obtained we hope that others can better assess the validity of their own CATFEE results.

[1] URL: http://uqbar.ncifcrf.gov/∼catfee

Methods

In this study complexes of factor Xa and trypsin with a set of 10 inhibitors were investigated. All the inhibitors are derivatives of 2,4,6-triphenoxypyridine displayed in Figure 1. The substituents R_1 and R_2 for each inhibitor are given in Table 1. No structural data on the complexes of factor Xa and trypsin with these inhibitors was available at the time the work was undertaken. For this reason the crystallographic structures of factor Xa and trypsin complexed with 2,6-diphenoxypyridine (PDB reference 1FJS for factor Xa [11] and 1QB1 for trypsin [12]) were used as templates to obtain initial structures. The structure of factor Xa is comprised of two polypeptide chains. However, only the primary chain, which contains the binding site, was included in the calculations. This was done in order to minimize the size of the system and because the minor chains are positioned well away from the inhibitor binding site.

The initial structures of factor Xa and trypsin complexed with inhibitor 1 (I1) were constructed by superimposing the nitrogen atoms of the amino(imino)methyl group and the nitrogen atoms of the imidazolyl group of the inhibitor onto the corresponding atoms in the appropriate template structure. The conformation of I1 used for this procedure was taken from a simulation of the isolated inhibitor in water. The structures of factor Xa and trypsin complexed with inhibitors I2–I10 were derived from the I1 complexes by mutating R_1 and R_2 substituents during the free energy calculations.

Mutations

The inhibitors (see Table 1) can be divided into two groups depending on the nature of the substituents. Inhibitors I1–I6 have the same R_2 substituent while inhibitors I2 and I7–I10 have the same R_1 substituent. Inhibitor I2 belongs to both groups and was used as a reference. The mutations were chosen in order to maximize the number of possible cycle closures while minimizing the size of the mutation itself (i.e. by only performing mutations within a given group). As the free energy for any closed cycle is zero by definition, cycle closure can provide an important check on the degree of convergence within the calculations. Figures 2 and 3 illustrate the pathways that were used to convert one inhibitor into another.

Table 1. The set of 10 inhibitors used in this study to estimate binding free energies to factor Xa and trypsin. R_1 and R_2 are the substituents of 2,4,6-triphenoxypyridine shown in Figure 1. Note that the substituent R_2 is common to inhibitors I1–I6 while the substituent R_1 is common to inhibitors I2 and I7–I10.

Inhibitor	R_1	R_2
I1	2-OH-4-COO^-	1-methyl-2(2H)-imidazoline
I2	2,6-OCH_3-4-COO^-	1-methyl-2(2H)-imidazoline
I3	2-Cl-4-$COOCH_2CH_3$	1-methyl-2(2H)-imidazoline
I4	2,6-CH_3-4-$COOCH_2CH_3$	1-methyl-2(2H)-imidazoline
I5	2-$OCH_2CH_2CH_2N(CH_3)_2$-5-COO^-	1-methyl-2(1H)-imidazoline
I6	2-Cl	1-methyl-2(1H)-imidazoline
I7	2,6-OCH_3-4-COO^-	$NHC(NH)NH_2$
I8	2,6-OCH_3-4-COO^-	(pyrrolidin-1-yl)(imino)methylamino
I9	2,6-OCH_3-4-COO^-	(1H-imidazolin-2-yl)amino
I10	2,6-OCH_3-4-COO^-	(1-methyl-1H-imidazolin-2-yl)amino

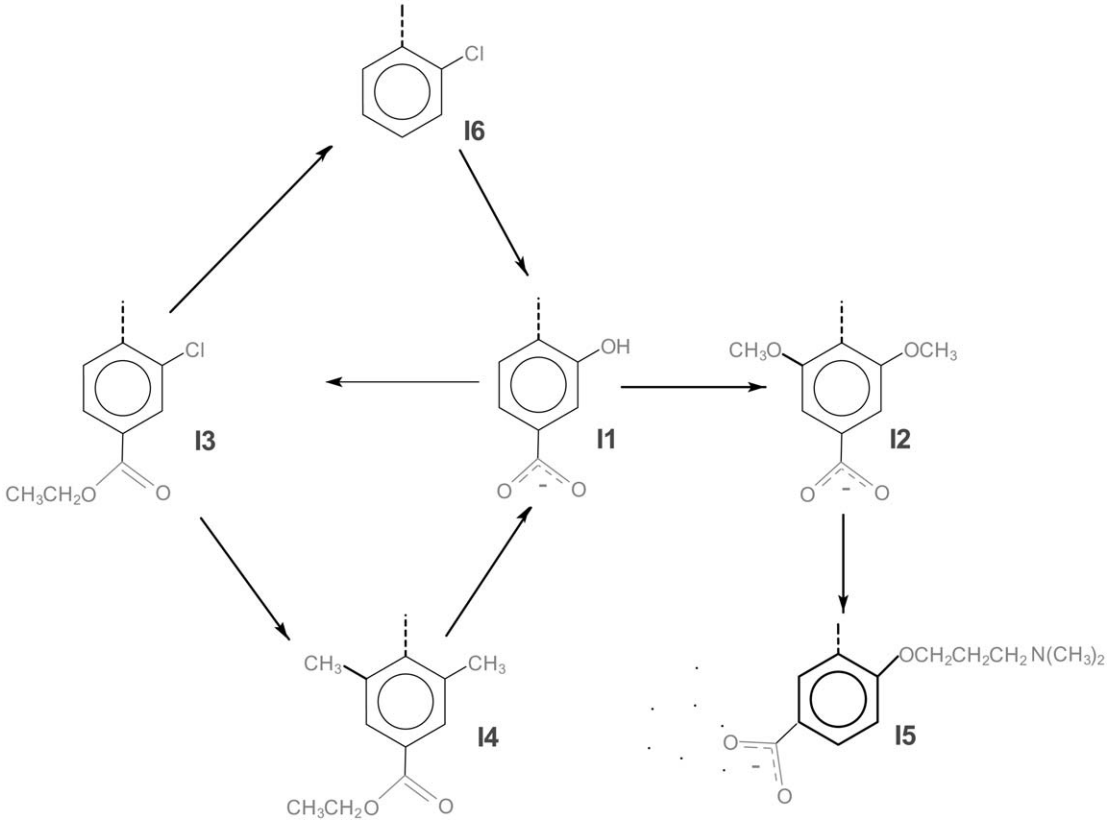

Figure 2. Mutations performed to transform inhibitors I1–I6 into each other. These inhibitors differ in the aromatic substituents, R_1, of the benzene ring in the para position (see Figure 1).

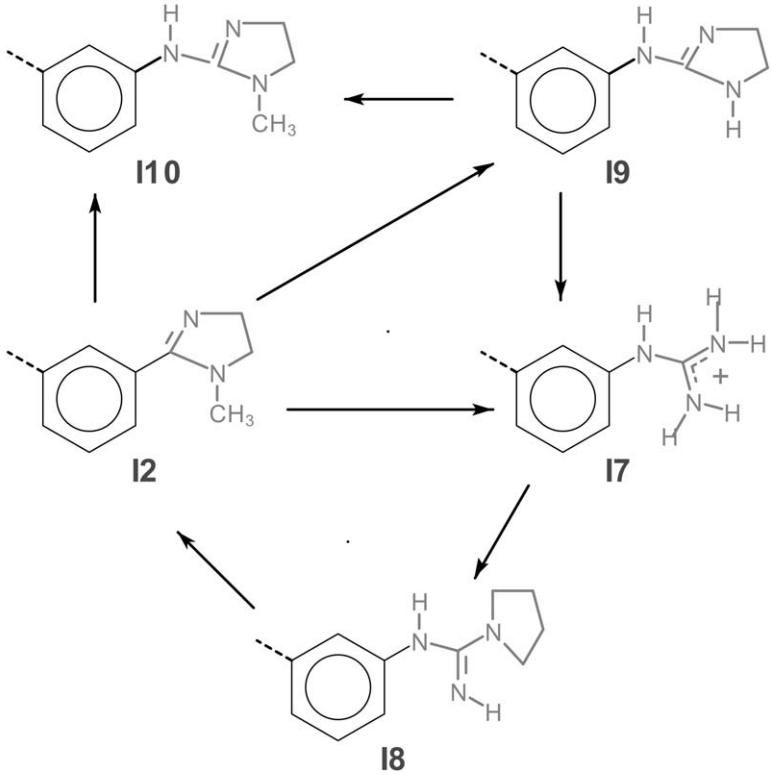

Figure 3. Mutations performed to transform inhibitors I2 and I7–I10 into each other. The inhibitors differ in the aromatic meta-substituent, R_2, of the benzene ring in the ortho position.

Force field

The GROMOS96 (43a2) force field was used to describe both the protein and the inhibitors where possible [13, 14]. When no parameters were available in the standard force field to describe certain atomic interactions, parameters were obtained by fitting to *ab initio* calculations. The primary aim in developing additional parameters was to maintain compatibility with the rest of the force field. The parameters that were not available included the torsional potential between the phenoxy ring and the pyridine, and some atomic charges. The calculations were performed at the Restricted Hartree-Fock (RHF) level with the Gaussian94 [15] program using the 6-31G* basis set. The torsional potential defined by an aromatic carbon and an ether oxygen as central atoms was determined by fitting an empirical potential to the *ab initio* potential energy surface of the model system, *p*-phenylpiridine. Atomic charges were derived by scaling the atomic charges obtained by fitting the RHF/6-31G* molecular electrostatic potential of the corresponding small organic molecules using the method of Merz and Kollman [16]. The additional parameters derived for use in the calculations are presented in Table 2 and Figure 4.

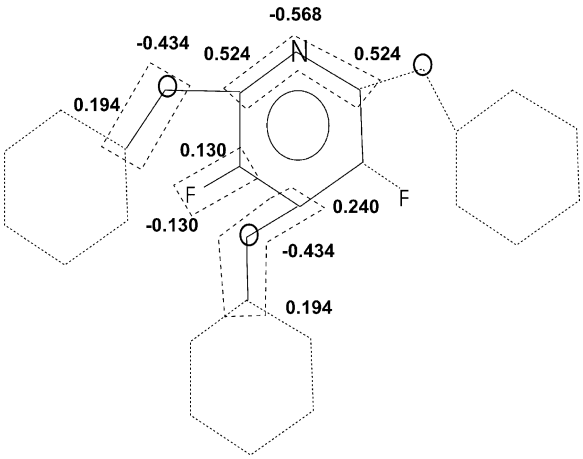

Figure 4. Atomic charges of atoms common to all inhibitors. The dashed polygons indicate charge groups.

Table 2. Bond and torsional parameters as well as atomic charges not available in the GROMOS96 force field [13]. They are calculated by the Restricted Hartree-Fock (RHF) method using the Gaussian94 software package [15].

Bond	Force constant K_b (10^6 kJ mol^{-1} nm^{-4})	Ideal bond length b_0 (nm)
C-Cl	10.2	0.174

Torsion	Force constant K_δ (kJ mol^{-1})	Phase shift cos(δ)	Multiplicity m
* - C - OA - *	14.14	−1	2

Charge group	(*e*-)	(*e*-)	(*e*-)	(*e*-)
C-Cl	C	CL		
	0.15	−0.15		
$COOCH_2$	C	O	OA	CH2
	0.53	−0.38	−0.397	0.247
$CH_2N(CH_3)_2$	CH2	N	CH3	CH3
	0.00	−0.2	0.1	0.1

Computational details

The Molecular Dynamics simulations were performed using the GROMACS package version 3.0 [17–19] in explicit solvent and under periodic boundary conditions. The protein-inhibitor complexes were placed in a truncated octahedron box containing approximately 6300 water molecules. Simulations of the free inhibitor in water were performed in boxes containing approximately 850 water molecules. The Simple Point Charge (SPC) model was used to describe the water molecules [20]. The non-bonded interactions were evaluated using a twin range cutoff of 0.9 and 1.4 nm. Interactions within the shorter range cutoff were evaluated every step while interactions within the longer range cutoff were evaluated every 5 steps. To correct for the neglect of electrostatic interactions beyond the longer range cutoff, a reaction field (RF) correction with $\epsilon_{RF} = 78.0$ was used. To maintain the temperature of the system at a constant value of 300 K, a Berendsen thermostat [21] was applied. The protein, the inhibitor and the solvent were each independently coupled to a temperature bath with a coupling time of 0.1 ps. The pressure was maintained by weak coupling to a reference pressure of 1 bar, with a coupling time of 1.0 ps and an isothermal compressibility of $4.6{\cdot}10^{-5}$ bar^{-1} [21]. The time step used for integrating the equations of motion was 0.002 ps. The bond lengths and angle of the water molecules were constrained using the SETTLE algorithm [22] while the bond lengths within the protein were constrained using the LINCS algorithm [23].

The protein-I1 complex was equilibrated for 2 ns before the free energy calculations were performed. For the case of the free inhibitor in water the equilibration time was 200 ps.

Free energy calculations

The free energy difference between two states of a given system can be obtained using the coupling parameter approach [24]. The Hamiltonian of the system, H, is expressed as a function of a coupling parameter, λ, which describes the path taken from the initial to the final state such that when $\lambda = \lambda_A$ the system corresponds to state A and when $\lambda = \lambda_B$ the system corresponds to state B. The difference in the free energy between the two states, ΔG_{AB}, is then given by the Thermodynamic Integration (TI) equation,

$$\begin{aligned}\Delta G_{AB} &= G(\lambda_B) - G(\lambda_A) = \int_{\lambda_A}^{\lambda_B} \left(\frac{\partial G}{\partial \lambda}\right) d\lambda \\ &= \int_{\lambda_A}^{\lambda_B} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda\end{aligned} \tag{1}$$

In Equation 1, $< \ldots >_\lambda$ denotes an ensemble average at a given value of λ. The integral in Equation 1 may be evaluated numerically using a number of discrete λ-points.

Mutating one atom type into another with multiple steps requires the interpolation of the Lennard-Jones (LJ) and the Coulomb interactions between state A and state B. This procedure generates instabilities at points where atoms are created or annihilated. To circumvent this problem a soft-core potential may be applied where the singularity at the origin is substituted by a core of finite height [25–28]. The interpolated interaction between atom i and atom j is described by,

$$V(r_{ij}, \lambda) = (1 - \lambda)V_A(r_A) + \lambda V_B(r_B), \tag{2}$$

where the modified distances r_A and r_B are,

$$r_A = \left(\alpha\sigma_A{}^6\lambda^2 + r^6\right)^{1/6} \tag{3}$$

$$r_B = \left(\alpha\sigma_B{}^6(1 - \lambda)^2 + r^6\right)^{1/6}, \tag{4}$$

the soft-core parameter, α, controls the height of the potential around $r = 0$ and σ has its normal meaning as in LJ potential function.

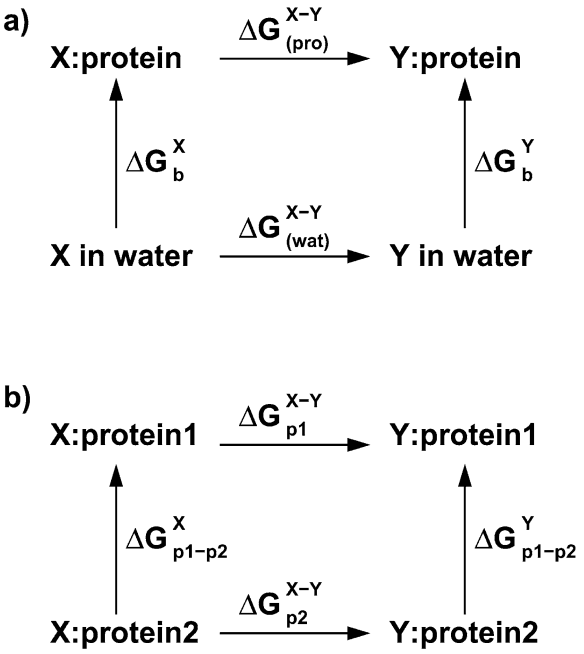

Figure 5. Thermodynamic cycles used to determine the difference in binding free energy between inhibitor X and inhibitor Y relative to (a) the unbound inhibitors in water, and (b) the inhibitors bound to another protein.

The binding free energy, ΔG_b, is the work required to transfer the inhibitor from being free in solution to being bound to the protein. The relative binding free energy, $\Delta\Delta G_b^{X-Y}$, represents the difference in the binding free energy between inhibitor X and inhibitor Y and was evaluated using the thermodynamic cycle shown in Figure 5a,

$$\Delta\Delta G_b^{X-Y} = \Delta G_b^Y - \Delta G_b^X = \Delta G^{X-Y}(\text{pro}) - \Delta G^{X-Y}(\text{wat}) \tag{5}$$

where $\Delta G^{X-Y}(\text{pro})$ and $\Delta G^{X-Y}(\text{wat})$ are the work required to mutate inhibitor X to inhibitor Y in the protein and in water, respectively. A similar expression is obtained for calculating the difference in the binding free energy between inhibitor X and inhibitor Y with respect to two proteins, $p1$ and $p2$, as shown in Figure 5b,

$$\Delta\Delta G_{p1-p2}^{X-Y} = \Delta G_{p1-p2}^Y - \Delta G_{p1-p2}^X = \Delta G_{p1}^{X-Y} - \Delta G_{p2}^{X-Y}. \tag{6}$$

In the calculations of ΔG^{X-Y} one inhibitor was gradually mutated into another inhibitor. In cases where there was no direct correspondence between the atoms in the two molecules 'dummy atoms' were used. A dummy atom is an atom for which the non-bonded interactions with all other atoms are zero. Only bonded and non-bonded interactions were mutated during the calculations. The masses of the atoms were not altered. The non-bonded interactions between the initial and final states were interpolated using a soft-core potential [25] as implemented in the GROMACS simulation package [18, 29].

All mutations indicated in Figures 2 and 3 were performed for factor Xa and trypsin complexes and for the inhibitors free in water. The free energy was evaluated using 18 λ-points. The number of λ-points was increased for cases where the mutation generated a large perturbation in the system in regions where the integrand in Equation 1 as a function of λ exhibited a sharp discontinuity. For example we display in Figure 6 the value of $\partial H(\lambda)/\partial\lambda$ of two mutations. For the case of I3→I4 mutation, the curve has sharp discontinuity (Figure 6a), while in the case of I3→I6 mutation the curve is smooth (Figure 6b). At each value of λ the system was equilibrated for 50 ps. The derivative $\partial H(\lambda)/\partial\lambda$ was then averaged over 250 ps in the case of the protein-inhibitors complexes and over 150 ps in the case of the inhibitors in water. The average value of the derivatives was calculated at each λ point and the resulting free energy profile was then integrated using the trapezoidal method to obtain ΔG^{X-Y}. The error in $\langle\partial H(\lambda)/\partial\lambda\rangle_\lambda$ was estimated using a block averaging procedure [30, 31] at each λ-point. The individual errors were then integrated to yield an estimate of the error in ΔG^{X-Y}. All calculations were performed on a Linux cluster of 1.7 GHz Pentium-IV based machines. Each λ-point for the complex (300 ps) required approximately 12 CPU hours.

Results

Mutations in water

The free energy values of mutating one inhibitor X into another Y, ΔG^{X-Y}, in water via the pathways shown in Figures 2 and 3 are reported in Table 3 and Figure 8. To check the dependency of the results on the direction of the mutation the free energies of the reverse mutations, ΔG^{Y-X}, were also calculated. Except for I1 → I3 and for I2 → I5, the average discrepancy in the free energy between the forward and the reverse transformation is only 2.7 kJ/mol, just above the average estimated error. Thus, in most cases the results are insensitive to the direction of the

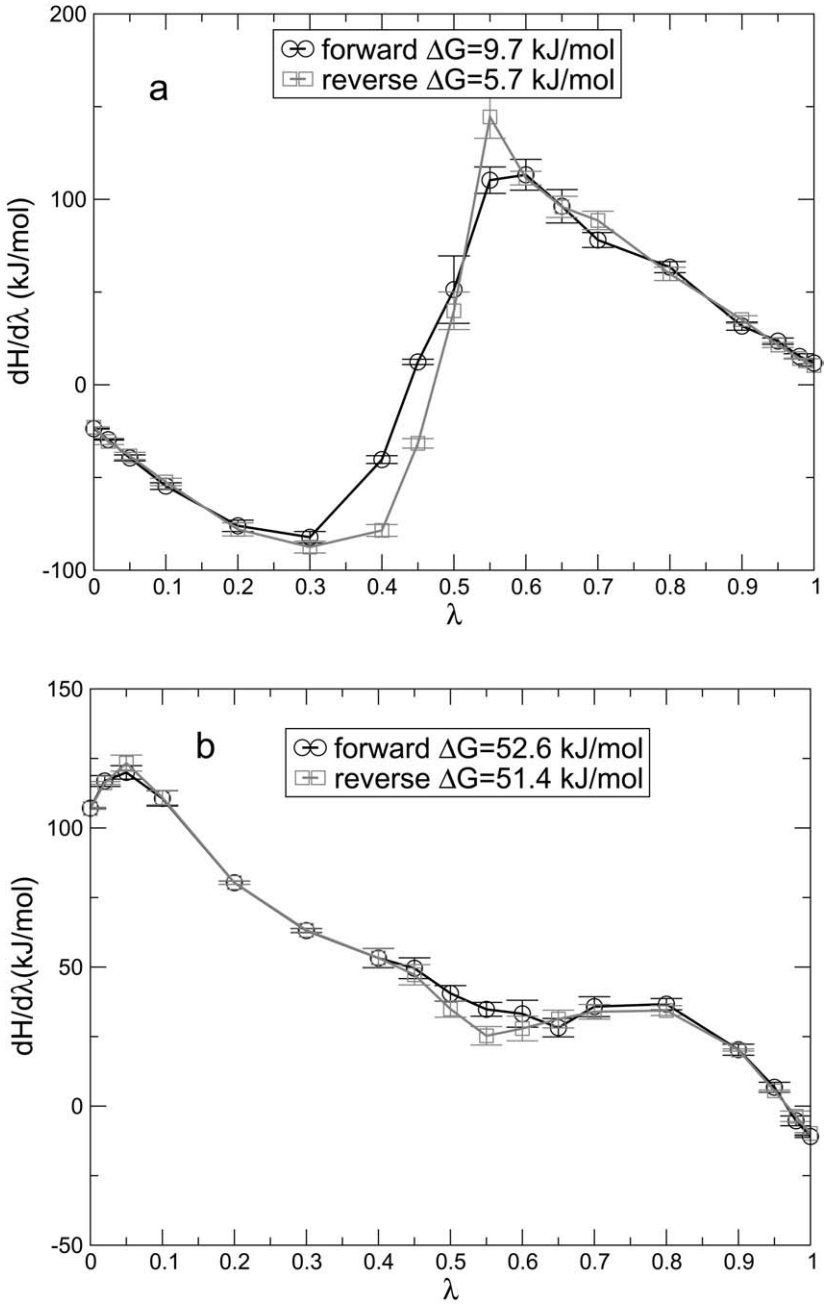

Figure 6. The value of $\partial H(\lambda)/\partial\lambda$ as a function of λ for the forward (black) and backward (gray) transformations of the two mutations: (a) I3 → I4; (b) I3→I6. Error bars are also shown. Note that the curves in (a) show a large discontinuity and a large discrepancy in the value of the integral (δG), while the curves in (b) are smooth and the value of the integral is similar.

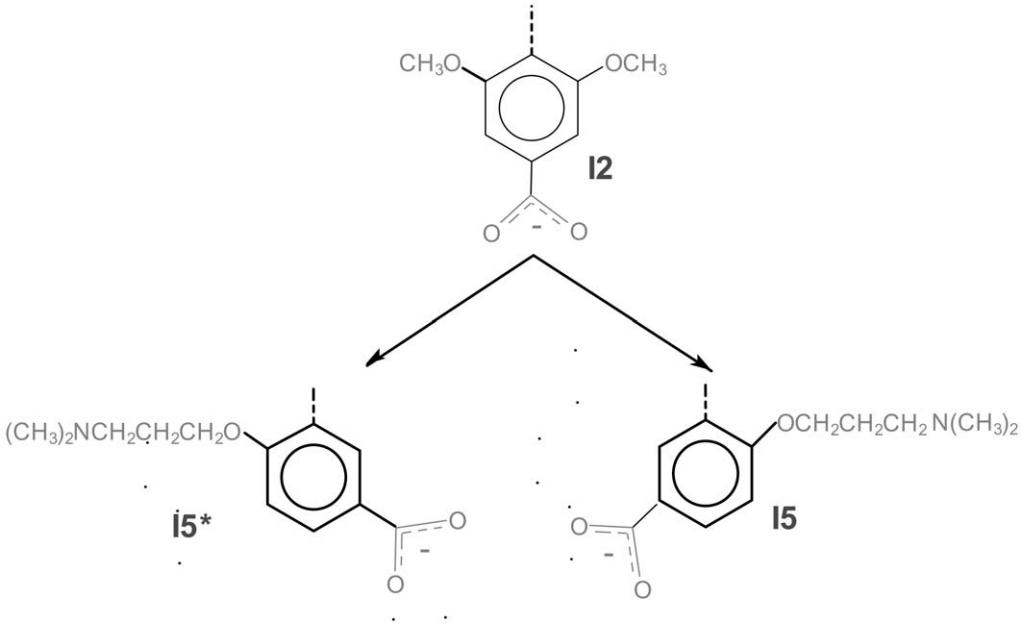

Figure 7. The I2 → I5 mutation.

Table 3. The free energies (in kJ/mol) for mutating inhibitor X into inhibitor Y in water, ΔG^{X-Y}, and for the corresponding reversed mutation, ΔG^{Y-X}. $\mathcal{D}$ is the absolute value of the sum between ΔG^{X-Y} and ΔG^{Y-X}. The descriptions of the chemical transformations are shown in Figures 2 and 3.

Mutation	ΔG^{X-Y}	ΔG^{Y-X}	$\mathcal{D}$	Mutation	ΔG^{X-Y}	ΔG^{Y-X}	$\mathcal{D}$
I1 → I2	77.3 ± 1.3	−73.0 ± 1.6	4.3	I2 → I5	−39.5 ± 3.0	56.5 ± 4.8	17.0
I1* → I2	− − − −	−76.6 ± 1.4	–	I2 → I5*	−51.9 ± 3.1	48.8 ± 3.2	3.1
I1 → I3	271.5 ± 1.6	−258.1 ± 7.4	13.4	I2 → I7	−97.9 ± 1.3	98.7 ± 1.4	0.8
I1* → I3*	270.5 ± 1.7	− − − −	–	I7 → I8	−141.9 ± 1.8	146.6 ± 2.4	4.7
I3 → I6	52.6 ± 0.6	−51.4 ± 0.6	1.2	I8 → I2	238.8 ± 2.1	−239.0± 2.4	0.2
I6 → I1	−321.5 ± 1.5	322.9 ± 1.7	1.4	I9 → I7	−28.6 ± 1.4	32.9 ± 1.4	4.3
I6* → I1*	− − − −	321.0 ± 1.9	–	I2 → I9	−66.0 ± 1.2	65.9 ± 2.4	0.1
I3 → I4	9.7 ± 1.3	−5.7 ± 1.1	4.0	I9 → I10	2.8 ± 0.3	−4.0 ± 0.3	1.2
I4 → I1	−277.8 ± 1.7	271.6 ± 1.9	6.2	I2 → I10	−61.0 ± 1.1	65.0 ± 1.0	4.0

mutation, i.e. are reversible. The two cases where reversibility was not observed are discussed below.

In principle the aromatic rings in the inhibitor are free to rotate. However, such rotations may not necessarily be observed on the time scale of the simulations. For cases where such rotation would yield a different state (i.e. for cases not axially symmetric) the results may be dependent on the starting conformation. This could occur for inhibitors I1, I3, I5 and I6 (see Figure 2). Therefore, the mutations for both orientations were performed. The alternative orientation is indicated by a star, '*'. In water, free rotation of the aromatic rings is expected. The results should be independent of the initial conformation of the inhibitor. This is indeed the case, as can be seen from the values reported in Table 3 and Figure 8. The difference in the free energy starting from the two different conformers is within the estimated error (0.3–3.2 kJ/mol). There is only one exception, the I2 → I5 transformation, which also showed a large discrepancy between the forward and backward directions. This mutation involves many atoms as well as the annihilation and creation of a charge. The pathways for mutating I2 to I5 or I5* are shown in Figure 7. Simulations were conducted to examine the dependency of $\Delta G^{2\to5}$(wat) and of $\Delta G^{2\to5*}$(wat) on the sampling time. The res-

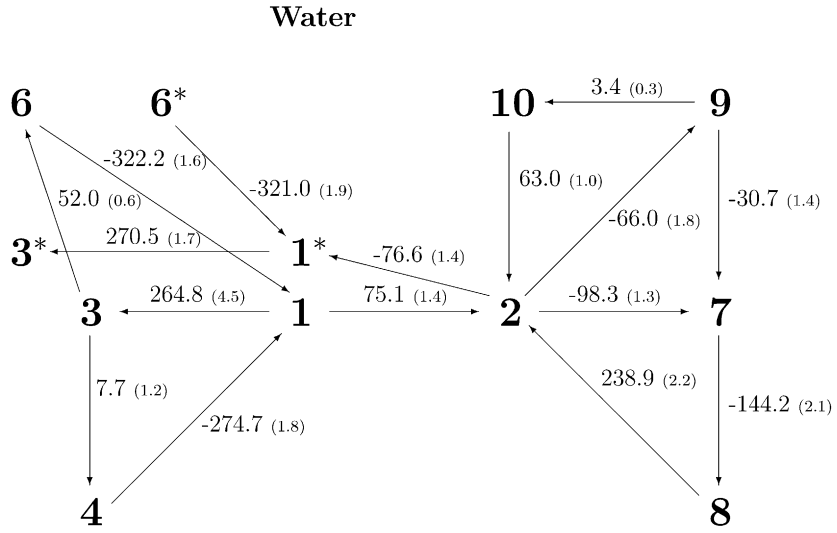

Figure 8. Schematic diagram of the mutations between the inhibitors in water. The values shown are the free energy changes (kJ/mol). The numbers in parentheses are the estimated errors.

ults are shown in Table 4. As the sampling time is increased, the difference in the free energy for the two conformations of I5 decreases from 12.4 kJ/mol when the sampling time is 150 ps to 1 kJ/mol (within the estimated error) when the sampling time is 6 ns. In this case, 150 ps is insufficient to sample the rotation of the phenoxy ring. This is because the amine group in I5 can form strong intramolecular interactions. For inhibitors I1, I3 and I6 similar intramolecular interactions are not possible.

In Table 5 the free energy values for the closed paths shown schematically in Figures 2 and 3 are reported. The magnitude of the error in evaluating the free energy along circular paths is comparable in the forward and reverse directions. The values reported in Table 5 correspond to the average of the forward and the backward transformations. The deviations from zero are in the range 0.4–5.4 kJ/mol which is smaller than the sum of the estimated errors of the individual steps, indicating there is some cancellation of error within the cycles.

In the I1 $\rightarrow$ I3 mutation, the derivative, $\partial H(\lambda)/\partial\lambda$, varied sharply as a function of λ in the region $\lambda \leq 0.4$. Therefore, in an attempt to improve the convergence of the calculation for this mutation, 6 additional λ-points were added: $\lambda = 0.04$, 0.06, 0.08, 0.15, 0.25 and 0.35. However, the difference in the free energy compared to the 18 λ-point calculation was only 2.2 kJ/mol. The free energy profiles of the forward and backward mutation in the region $\lambda \geq 0.4$ exhibited some degree of hysteresis. Therefore, the sampling time for all $\lambda \geq 0.4$ was extended to 1 ns in order to increase convergence. Again, no significant improvement was observed. The difference in the free energy compared to the 150 ps sampling time was 0.3 kJ/mol. The transformation from I1 to I3 involves an annihilation of charge of the para-carboxy group. It is not completely clear to us if this could be an explanation for the large discrepancy observed in this mutation.

Table 4. The free energies (kJ/mol) for the mutations I2 $\rightarrow$ I5 and I2 $\rightarrow$ I5* in water obtained using different sampling times.

Sampling time (ps)	$\Delta G^{2\rightarrow 5}$(wat)	$\Delta G^{2\rightarrow 5*}$(wat)	$\Delta\Delta G$
150	-39.5 ± 3.0	-51.9 ± 3.1	12.4
2000	-52.5 ± 2.6	-45.3 ± 9.6	7.2
4000	-51.5 ± 1.9	-47.1 ± 6.1	4.4
6000	-46.7 ± 4.9	-47.7 ± 4.5	1.0

Ligand-protein complexes

A prerequisite to obtain the correct relative free energy for a protein-inhibitor complex is that the inhibitor forms the proper interactions inside the binding site. Figures 9 and 10 show the structures after 2 ns of simulation of inhibitor I1 bound to factor Xa and to trypsin, respectively. Note that in both cases the

Table 5. Free energies (kJ/mol) of mutating one inhibitor into another along circular paths as described schematically in Figures 2 and 3 in water, in factor Xa and in trypsin.

Circular path	Water	Factor Xa	Trypsin
1-3-4-1	-2.2 ± 7.5	-12.1 ± 5.9	13.2 ± 6.6
1-3-6-1	-5.4 ± 6.7	6.0 ± 4.7	9.3 ± 6.2
2-9-7-2	1.6 ± 4.5	5.5 ± 6.2	-0.7 ± 6.0
2-9-10-2	0.4 ± 3.1	3.2 ± 3.1	-3.6 ± 5.2
2-7-8-2	-3.6 ± 5.6	5.3 ± 6.5	-4.9 ± 6.6

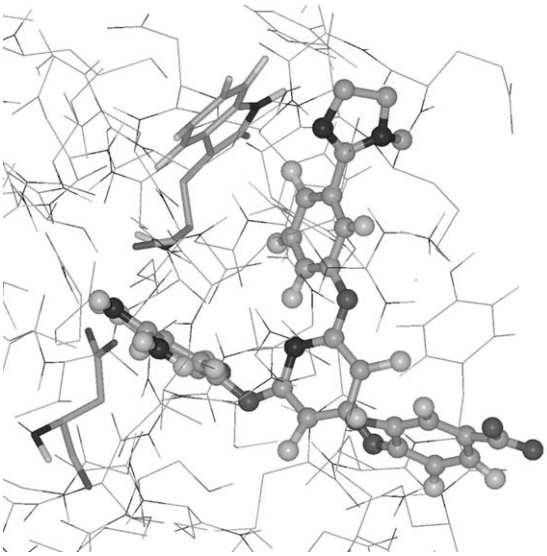

Figure 9. The binding site of factor Xa bound to I1. The inhibitor and the residues Asp189 and Trp215, that form hydrogen bonds with the benzylamidine ring of I1, are emphasized.

numbering of the residues is the same as the one adopted by the X-ray PDB files. The benzylamidine ring resides in the S1 pocket making hydrogen bonds with Asp189. The average distance between the amidine group of the inhibitor and the carbonyl oxygens of this residue is 2.1 Å for both factor Xa and trypsin. The hydroxyl group in the para position on the benzylamidine ring forms a hydrogen bond with Ser195. The average bond distances in factor Xa and in trypsin are 2.3 Å and 2.1 Å, respectively. The interactions in the S1 binding site in the simulations are consistent with results obtained from X-ray crystallographic studies of trypsin [12] and factor Xa [11] complexed with bisphenoxypyridine.

The phenyl ring of inhibitor I1, carrying the substituent 1-methyl-2-(2H)-imidazoline, lies in the vicinity of Trp215 in the S4 pocket. The average distance between the center of mass of these two groups is 6.0 Å in factor Xa and 5.0 Å in trypsin. In the X-ray structures of the bisphenoxypyridine complexes, the phenoxy group fits snugly into the S4 pocket and lies across the indole ring of Trp215 (4.0 Å).

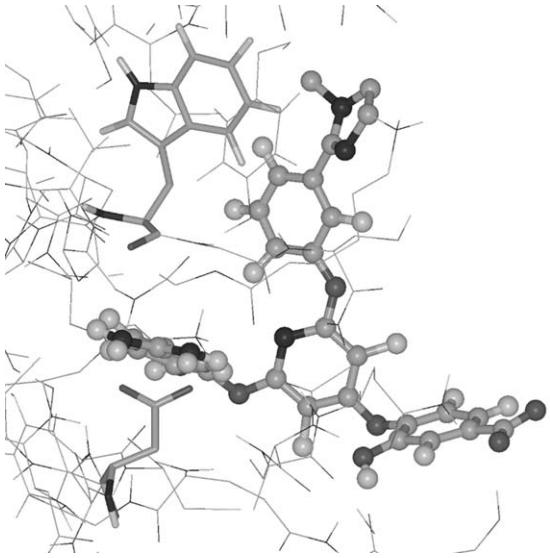

Figure 10. The binding site of trypsin bound to I1. The inhibitor and the residues Asp189 and Trp215, that form hydrogen bonds with the benzylamidine ring of I1, are emphasized.

Mutations in factor Xa and trypsin

Starting from the equilibrated structure of the protein-I1 complex, all mutations shown in Figures 2 and 3 were performed, except for the transformation I2 $\rightarrow$ I5. For this mutation, the results in water (Table 4) had already indicated that a minimum sampling time at each λ value in the order of nanoseconds is required. As in the complex even longer sampling times might be necessary to achieve convergence and simulating the protein-inhibitor system is computationally very demanding, this mutation was not performed in factor Xa or trypsin.

As shown earlier, the rotation of the aromatic ring in water is, in general, sampled adequately within 150 ps. However, inside the protein sampling is restricted and the aromatic rings may not adopt all possible orientations spontaneously. Therefore, for cases where rotation around the aromatic ring would yield different states, the calculations were performed in both orientations. The results of the mutations in factor Xa and in trypsin are reported in Figures 11 and 12, respectively. The difference between the free energy of the same mutation but with a different start-

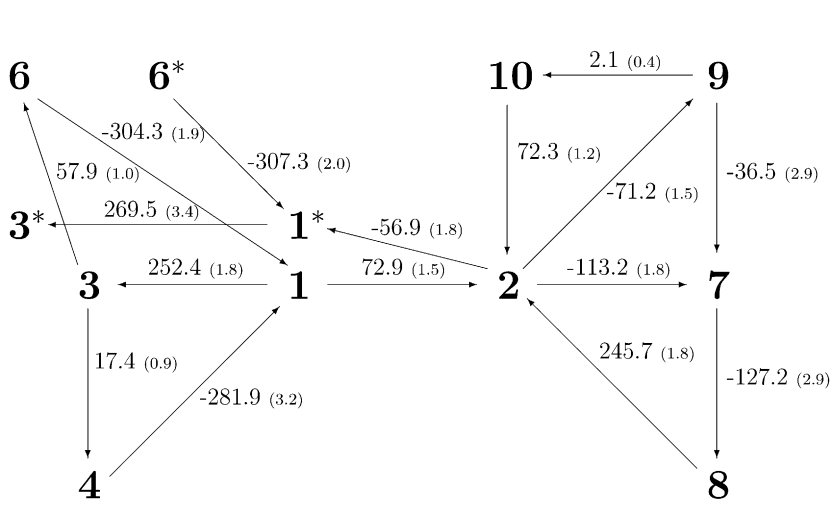

Figure 11. Schematic diagram of the mutations between the inhibitors in factor Xa. The values shown are the free energy changes (kJ/mol). The numbers in parentheses are the estimated errors.

ing conformation of the inhibitor ranged from 2.8 to 17.1 kJ/mol. Thus, 250 ps at each λ point is not sufficient to sample an equilibrium distribution of orientations of the benzene rings inside the binding pocket. In fact, it is probable that due to steric hindrance such rotations are not possible while the inhibitor is inside the binding pocket and that treating the conformations as independent would be required irrespective of the sampling time. Thus, from now on both orientations are considered separately and the preferred orientation is assumed to be the one that corresponds to the lower free energy.

As a test of convergence the results of the free energies along circular paths are reported in Table 5. The average deviation from zero is 6.4 kJ/mol, which is approximately equal to the accumulated error.

The binding free energies are obtained by subtracting the free energies of the mutations in water from the corresponding free energies in the protein. The relative free energy of binding of the two inhibitors inside the two proteins, $\Delta\Delta G^{X\to Y}_{p1-p2}$, was also obtained by subtracting the free energies of the mutations in one protein from the corresponding free energies in the other protein. It is convenient to have a single reference state for the inhibitor mutations, thus, in Table 6 the relative binding free energies are given with respect to I2, $\Delta\Delta G^{2\to Y}$. Where there are two or more pathways by which the free energy could be calculated the value corresponding to the pathway with the smallest accumulated error is given. The values provide an indication of the relative binding affinity of the inhibitors. Based on the results in Table 6 the ranking of the inhibitors between trypsin and factor Xa was determined and is shown in Figure 13. A distinction between the inhibitors is only made if the difference in the binding free energies is larger than the associated error. The average error, estimated from the closure of the thermodynamic cycle, of $\Delta\Delta G^{2\to Y}_{p1-p2}$ for the case of trypsin and factor Xa is 5.4 kJ/mol. Therefore, the inhibitors were divided into groups that differ by more than 10 kJ/mol. According to this ranking, I3 has the highest affinity for trypsin with respect to factor Xa, while inhibitors I7, I9 and I10 have the highest affinity for factor Xa with respect to trypsin.

Table 6. The difference of the average free energy (kJ/mol) of the mutation between the inhibitors referenced to inhibitor I2, $\Delta\Delta G^{2\to Y}$, in water, factor Xa and trypsin.

	FactorXa–water	Trypsin–water	Trypsin–factorXa
I1	2.2 ± 2.9	-0.7 ± 3.5	-20.4 ± 3.9
I2	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
I3	-10.7 ± 5.2	-10.9 ± 7.2	-29.6 ± 9.3
I4	-1.0 ± 5.6	-0.8 ± 5.8	0.2 ± 5.6
I6	-4.8 ± 5.5	2.5 ± 7.0	-3.5 ± 7.5
I7	-15.9 ± 3.4	0.6 ± 3.3	16.4 ± 3.8
I8	1.1 ± 4.6	-6.6 ± 4.7	-7.8 ± 5.4
I9	-10.1 ± 3.4	5.1 ± 3.3	15.1 ± 3.8
I10	-11.4 ± 3.5	6.0 ± 3.6	17.4 ± 4.1

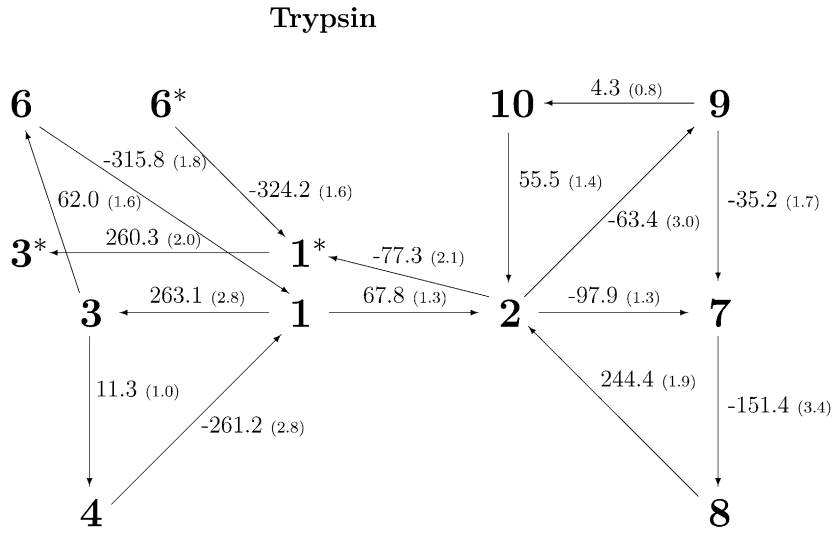

Figure 12. Schematic diagram of the mutations between the inhibitors in trypsin. The values shown are the free energy changes (kJ/mol). The numbers in parentheses are the estimated errors.

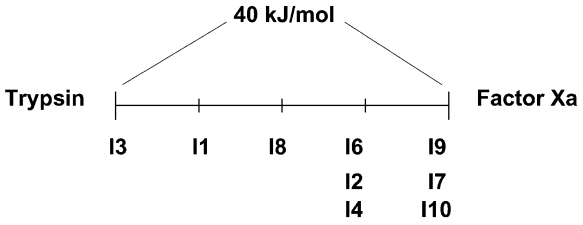

Figure 13. Schematic illustration of the inhibitors classified according to their relative affinities to trypsin and factor Xa. The resolution of the binding affinities between the different groups is approximately 10 kJ/mol.

Discussion and conclusions

To our knowledge, no experimental data on the binding constants of the 10 inhibitors to either of the two proteins was publicly available at the time this paper was written. Such information is the ultimate test of the success of any free energy calculation. The accuracy of free energy calculations depends on two factors. First, the accuracy of the force field. This can only be confirmed by comparison to experimental results. Second, free energy calculations depend on the degree of sampling and convergence. This is in principle independent of whether we know the 'true' answer. In any case to quote a recent review by Chipot and Pearlman [7], "it is clear that in some cases the amazingly good agreement between theory and experiment ... must have been fortuitous". For cases where the transformation involves a simple mutation (i.e. involves a small number of atoms and no change of charge) and the location of the compound inside the binding site is known experimentally, good agreement between the calculated binding affinities and experimental values is generally found. However, for large mutations that involve the annihilation and creation of adjacent groups of atoms, difficulties in sampling led to large errors when determining binding affinities. Note, that even converged results can still deviate from experimental results due to simulation conditions which are not appropriate for the system investigated. For example the treatment of the electrostatic interactions such as the RF approach is not ideal for simulation of large proteins.

The calculation of the forward and the backwards mutations in water shows that reasonable convergence was obtained during the 150 ps sampling time at each λ point for all the transformations studied, except for I2 $\rightarrow$ I5. Slight improvements upon increasing the sampling time could be achieved as the average difference between the forward and the backward mutation was still slightly higher than the average estimated error. Another indication that the results had converged is that the mutations in water were insensitive to the initial or final conformation of the inhibitor. In particular, sufficient averaging over the different rotational states of the benzene rings was obtained. For I2 $\rightarrow$ I5, a mutation characterized by a long chain tertiary amine that can form favorable intra-molecular interactions, approximately 6 ns at each λ value was needed to achieve convergence. Depending on the type of mutations performed, the free energy along closed cycles

in water (that encompass 3 individual transformations) is in the range 0.4–5.4 kJ/mol. This range of error is smaller than the sum of the estimated errors of the individual transformations. This cycle closure is obtained when using the average values from the forward and the backward mutations, suggesting that a further increase in sampling time would certainly decrease the error.

The results of the free energy calculations involving the protein-ligand complexes show that due to steric hindrance full sampling of the rotational states of the aromatic rings does not occur on a time scale of a few hundred picoseconds. Thus, in order to evaluate the binding free energy in such cases, it is necessary to consider each of the possible orientations separately. For the three cases that we examined the difference between the free energies of the transformations with two different orientations ranged from ~ 3 to ~ 17 kJ/mol. Such differences are sufficient to distinguish between the two orientations of the inhibitor inside the binding pocket of the receptor.

The type and complexity of the mutations that were performed in this study are very challenging. The number of mutated sites, their proximity in space and the creation or annihilation of the charges all contribute to the difficulty of obtaining converged results. Nevertheless, it has been possible to obtain closure of thermodynamic cycles containing 3 individual steps to within ± 5 kJ/mol. This is a very high degree of convergence considering the mutations involved and should give confidence that, given an appropriate force field, nanosecond simulations can yield accurate estimates of the binding free energies in cases relevant to drug design.

Finally, despite the CATFEE competition fiasco there is a clear need to objectively test methods to predict interaction free energies. The necessary data is already available in the proprietary databases of various pharmaceutical companies. We can only hope that a small proportion of such data can be released so that true 'blind' tests can be performed and evaluated by the research community. This is certainly in the interest of everybody.

Acknowledgements

This research has been supported by a Marie Curie Fellowship of the European Community, the Fifth Framework Programme, under contract numbers MCFI-1999-00161, HPMF-CT-1999-00-381 and by the NWO contract number OC-98-007.

References

1. van Gunsteren, W.F. and Berendsen, H. J.C., Angew. Chem. Int. Ed. Engl., 29 (1990) 992.
2. Kollman, P.A., Chem. Rev., 93 (1993) 2395.
3. Mark, A.E., van Helden, S.P., Smith, P.E., Janssen, L.H.M. and van Gunsteren, W.F., J. Am. Chem. Soc., 116 (1994) 6293.
4. Mark, A.E., Free energy perturbation calculations. In Encyclopedia of Computational Chemistry, van RagueSchleyer, P., editor, Vol. 2, 1070–1083. John Wiley & Sons, New York, 1998.
5. Dixit, S.B. and Chipot, C., J. Phys. Chem. A, 105 (2001) 9795.
6. van Gunsteren, W.F., Daura, X. and Mark, A.E., Helv. Chim. Acta, 85 (2002) 3113.
7. Chipot, C. and Pearlman, D.A., Mol. Sim., 28 (2002) 1.
8. Radmer, R.J. and Kollman, P.A., J. Comput.-Aided Mol. Des., 12 (1998) 215.
9. Essex, J.W., Severance, D.L., Tirado-Rives, J. and Jorgensen, W.L., J. Phys. Chem. B, 101 (1997) 9663.
10. Blas, J.R., Marquez, M., Sessier, J.L., Luque, F.J. and Orozco, M., J. Am. Chem. Soc., 124 (2002) 12796.
11. Adler, M., Davey, D.D., Phillips, G.B., Kim, S.H., Jancarik, J., Rumennik, G., Light, D.R. and Whitlow, M., Biochemistry, 39 (2000) 12534.
12. Whitlow, M., Arnaiz, D.O., Buckman, B.O., Davey, D.D., Griedel, B., Guilford, W.J., Koovakkat, S.K., Liang, A., Mohan, R., Phillips, G.B., Seto, M., Haw, K.J.S., Xu, W., Zhao, Z., Light, D.R. and Morrissey, M.M., Act. Cryst. D., D55 (1999) 1395.
13. van Gunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W. R.P. and Tironi, I.G., Biomolecular Simulation: GROMOS96 Manual and User Guide. BIOMOS b.v., Zürich, Groningen, 1996.
14. Schuler, L.D. and van Gunsteren, W.F., Mol. Sim., 25 (2000) 301.
15. Frisch, M.J., Trucks, G.W., Schlegel, H.B., Gill, P. M.W., Johnson, B.G., Robb, M.A., Cheeseman, J.R., Keith, T.A., Petersson, G.A., Montgomery, J.A., Raghavachari, K., Al-Laham, M.A., Zakrzewski, V.G., Ortiz, J.V., Foresman, J.B., Ciosloswki, J., Stefanof, B.B., Nanayakkara, A., Challacombe, M., Peng, C.Y., Ayala, P.Y., Chen, W., Wong, M.W., Andres, J.L., Replogle, E.S., Gomperts, R., Martin, R.L., Fox, D.J., Binkley, J.S., Defrees, D.J., Baker, J., Stewart, J.P., Head-Gordon, M., Gonzalez, C. and Pople, J.A., Gaussian 94, Revision A.1. Gaussian, Inc., Pittsburgh, PA, 1995.
16. Besler, B.H., Merz Jr. K.M. and Kollman, P.A., J. Comp. Chem., 11 (1990) 431.
17. Berendsen, H. J.C., van der Spoel, D. and van Drunen, R., Comp. Phys. Comm., 91 (1995) 43.
18. Lindahl, E., Hess, B. and van der Spoel, D., J. Mol. Model., 7 (2001) 306.
19. van der Spoel, D., van Buuren, A.R., Apol, E., Meulenhoff, P.J., Tieleman, D.P., Sijbers, A.L.T.M., Hess, B., Feenstra, K.A., Lindahl, E., van Drunen, R. and Berendsen, H.J.C., Gromacs User Manual version 3.0. Nijenborgh 4, 9747 AG Groningen, The Netherlands. Internet: http://www.gromacs.org, 2001.
20. Berendsen, H.J.C., Postma, J. P.M., van Gunsteren, W.F., and Hermans, J., Interaction models for water in relation to protein

hydration. In Intermolecular Forces, Pullman, B. (Ed.) D. Reidel Publishing Company, Dordrecht, The Netherlands, 1981, pp. 331–342.
21. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R., J. Chem. Phys., 81 (1984) 3684.
22. Miyamoto, S. and Kollman, P.A., J. Comp. Chem., 13 (1992) 952.
23. Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M., J. Comp. Chem., 18 (1997) 1463.
24. Kirkwood, J.G., J. Chem. Phys., 3 (1935) 300.
25. Beutler, T.C., Mark, A.E., van Schaik, R.C., Geber, P.R. and van Gunsteren, W.F., Chem. Phys. Lett., 222 (1994) 529.
26. Liu, H., Mark, A.E. and van Gunsteren, W.F., J. Phys. Chem., 100 (1996) 9485.
27. Schaefer, H., van Gunsteren, W. F. and Mark, A.E., J. Comp. Chem., 20 (1999) 1604.
28. Pitera, J.W. and van Gunsteren, W.F., J. Phys. Chem. B, 105 (2001) 11264.
29. Villa, A. and Mark, A.E., J. Comp. Chem., 23 (2002) 548.
30. Bishop, M. and Frinks, S., J. Chem. Phys., 87 (1987) 3675.
31. Allen, M.P. and Tildesley, D.J., Computer Simulations of Liquids. Oxford Science Publications, Oxford, UK, 1987.