Single-Step Perturbations to Calculate Free Energy Differences from Unphysical Reference States: Limits on Size, Flexibility, and Character

CHRIS OOSTENBRINK, WILFRED F. VAN GUNSTEREN
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETH Hönggerberg, CH 8093 Zürich, Switzerland

Received 23 January 2003; Accepted 21 March 2003

Abstract: Relative free energies for a series of not too different compounds can be estimated accurately from a single simulation of an unphysical reference state that encompasses the characteristic molecular features of the compounds. Previously, this method has been applied to the calculation of free energies of solvation and of ligand binding for small molecules. In the present study we investigate the limits to the accuracy of the method by applying it to a realistic model of the binding of a set of rather large ligands to the protein factor Xa, a key protein in current efforts to design anticoagulation drugs. The evaluation of the binding free energies and conformations of nine derivatives of a biphenylamidino inhibitor leads to insights regarding the effect of the size, flexibility, and character of the unphysical part of the ligand in the reference state on the accuracy of the predicted binding free energies.

 J Comput Chem 24: 1730–1739, 2003

Key words: molecular dynamics simulation; single-step perturbation; soft-core interaction; factor Xa; free-energy calculation

Introduction

Relative free energies for a series of physically meaningful states can be calculated efficiently from a simulation of a single, unphysical, reference state, using the free energy perturbation formula.[1] In recent years this method has been successfully applied to the calculation of solvation free energies of small nonpolar[2] and polar[3] compounds, as well as to the calculation of relative binding free energies for series of ligands to a common receptor.[4,5] This approach tackles the sampling problem in free energy calculations by simulating an unphysical reference state, designed to sample a broad ensemble of conformations. If the ensemble is broad enough, one can calculate from it the free energy difference between the unphysical reference state and a physically meaningful, realistic end state. By designing the reference state such that in a simulation of the reference state, conformations relevant to several end states are sampled, one can obtain free energy differences for several end states from the single ensemble, and thus calculate relative free energies between multiple end states. As an unphysical reference state we use molecules containing atoms with a soft-core Lennard-Jones interaction.[6] For the calculation of free energies of solvation, one typically uses a single "soft atom" in solution,[2,3] while for the calculation of relative free energies of binding ligands to proteins the ligand in the reference state is a combination of real atoms, forming a molecular framework, and soft atoms at specific sites, allowing for enhanced sampling of the surroundings around these sites.[4,5]

The single-step perturbation method combined with an unphysical reference state has two major advantages over other methods to compute free energy differences. Because only one (or two in case of ligand binding) simulation is required to obtain the free energy differences with a great many other states, and only the terms in which their Hamiltonians differ need to be evaluated, it can easily be a factor of 100 to 1000 times more efficient than thermodynamic integration for a set of, say, 10 compounds. Secondly, the single-step method allows one to identify the particular conformations of a compound that contribute significantly to lower its free energy. Due to the unphysical character of the reference state, a wide variety of conformations is checked with respect to their ability to lower the free energy of the compound. When studying ligand binding, this relation between a binding or free conformation of the ligand and its contribution to its binding free energy offers insight into the relevant binding modes of ligands. As with all techniques, it is of interest to explore the limits of the technique: how far can the differences in terms of size, flexibility, and atomic character between the various compounds be enlarged

Correspondence to: W. F. van Gunsteren; e-mail: wfvgn@igc.phys.chem.ethz.ch

without loosing too much accuracy of the computed free energy differences?

Here we apply this method to the calculation of relative binding free energies for a series of bisphenylamidine carboxylate inhibitors of the enzyme factor Xa (fXa). fXa is an important target protein for the inhibition of blood coagulation. It is a trypsin-like serine protease and is responsible for the proteolysis of prothrombin to produce thrombin. One of the functions of thrombin, which is again a serine protease, is to convert fibrinogen into fibrin, which is responsible for blood clotting. Direct inhibition of fXa is therefore expected to be more efficient than the inhibition of thrombin to prevent blood coagulation. From the many reported inhibitors of fXa[7,8–12] we have selected a group of bisphenylamidine carboxylate analogues[13] to test the single-step perturbation approach on yet another realistic example. This allows us to explore the range of application of the one-step perturbation technique regarding the size of the soft interaction site and the dynamical behavior of the soft atoms.

The discrepancies found between simulated and experimental ligand-binding free energies when exploring a variety of ligands can be due to inaccuracies of the biomolecular force field used or to inaccuracies inherent to the one-step perturbation approach, that is, the choice of unphysical reference state, as applied here. The discrepancies found here can all be traced to limitations of the chosen reference state, and suggest improvements of this choice.

Method

The free energy difference between a state A of a system and a reference state, R, can be calculated from a simulation of state R, using the free energy perturbation (FEP) formula:

$$\Delta G_{\mathrm{AR}} = -k_B T \ln\langle e^{-(E_{\mathrm{A}}-E_{\mathrm{R}})/k_B T}\rangle_R \qquad (1)$$

The brackets indicate an ensemble average over the simulation of state R, k_B is the Boltzmann constant, T is the temperature, and E_{A} and E_{R} are the energies of configurations of the system in states A and R, respectively. This formula only yields a reasonable estimate of the free energy difference between states A and R if the ensemble that is generated for state R contains sufficient configurations that are relevant for state A as well. In cases where the different states in which one is interested are defined by different molecules, this will generally not be the case. Standard multiple-step perturbation approaches divide a pathway between states A and R into a number of unphysical intermediate states and calculate the free energy differences between these intermediate states using eq. (1).[14,15] Similarly, in the thermodynamic integration (TI) approach the system is simulated at several intermediate states in order to be able to integrate the derivative of the Hamiltonian with respect to the coupling parameter λ that connects the two states. The single-step approach we employ is based on the choice of an unphysical reference state that contains soft atoms and so generates an ensemble of configurations that is broad enough to contain relevant configurations not only for one, but for several physically meaningful end states.

So far this method has been applied to a number of reference compounds that contain one or several relatively small (one or two atoms) unphysical groups of soft atoms and have been observed to yield promising relative free energies.[4,5] In the current study we have designed a reference state that contains a single group of eight soft atoms, atoms 41–48 in Figure 1. Also represented in Figure 1 are the different real ligands that we are interested in, as groups of atoms that replace these soft atoms. The force field parameters that were used to describe the interactions of the reference ligand and the real ligands were taken from the GROMOS96 force field (45A3).[16,17] Partial charges can be found in Figure 1, and other force field parameters for the reference ligand in Table 1. Table 2 lists the differences in parameters between the reference ligand and the real ligands. The GROMOS96 atom type 19 is a dummy atom, which has no van der Waals interaction.

For the calculation of relative free energies of binding to fXa, two simulations are required: one of the reference ligand in solution and one where the reference ligand is bound to the protein in solution.[4] The first simulation was carried out by immersing a modeled structure of the reference ligand into a rectangular periodic box containing 1419 simple point charge (SPC)[18] water molecules and two chlorine ions to ensure an overall neutral system. For the protein simulation we started from the crystal structure of fXa complexed to the inhibitor rpr128515 (PDB code 1EZQ)[19] and replaced this (similar) inhibitor with our reference ligand. In this model, the phenylamidino groups occupy the S1 and S4 pockets of the protein, while the soft atoms of the reference ligand point in the direction of Arg143 and Gln192, as described in ref. 13. The protein-inhibitor complex was then placed in a truncated octahedral box, containing 8842 SPC water molecules. Protonation states of the protein were selected to correspond to a pH of 7. The protonation states of the five histidines were assigned after visual inspection of the structure: His57, His83, His145, His199 were expected to be protonated at Nδ1, whereas His91 was expected to be protonated at Nε2. From the protonation states of the protein and the ligand and the calcium ion that was bound to fXa, a net charge of +5 e was calculated. To make the complete system neutral, five water molecules that had the most favorable electrostatic potential for replacement by a negative ion were replaced by five chlorine ions.

Molecular dynamics simulations of both systems were started assigning initial velocities randomly from a Maxwell-Boltzmann distribution at 50 K and harmonically restraining the positions of the solute atoms to their initial positions with a force constant of 2.5×10^4 kJ mol^{-1} nm^{-2}. Over a period of 40 ps the temperature was increased stepwise, while simultaneously the position restraints were gradually removed. After an equilibration period of about 250 ps, production simulations (2 ns for the solvent simulation, 2×1 ns for the protein simulation, see below) were carried out at a constant temperature of 300 K and a constant pressure of 1 atm. The Berendsen weak coupling scheme[20] was used to maintain both the temperature (coupling time 0.1 ps) and the pressure [coupling time 0.5 ps; isothermal compressibility of 45.75×10^{-5} (kJ mol^{-1} nm^{-3})$^{-1}$].[16] Long-range interactions were calculated using the triple range cut-off scheme. All interactions within 0.8 nm were calculated every time step from a charge-group based pairlist that was generated every five time steps. Interactions between pairs that were between 0.8 and 1.4 nm apart were calculated every fifth time step and kept constant between pairlist updates. A reaction field force[21] was applied to account for the interactions with the medium outside the largest cutoff sphere, with a relative permittivity of 61.[22] The SHAKE algorithm[23] was applied to constrain all bond lengths to their ideal values, which

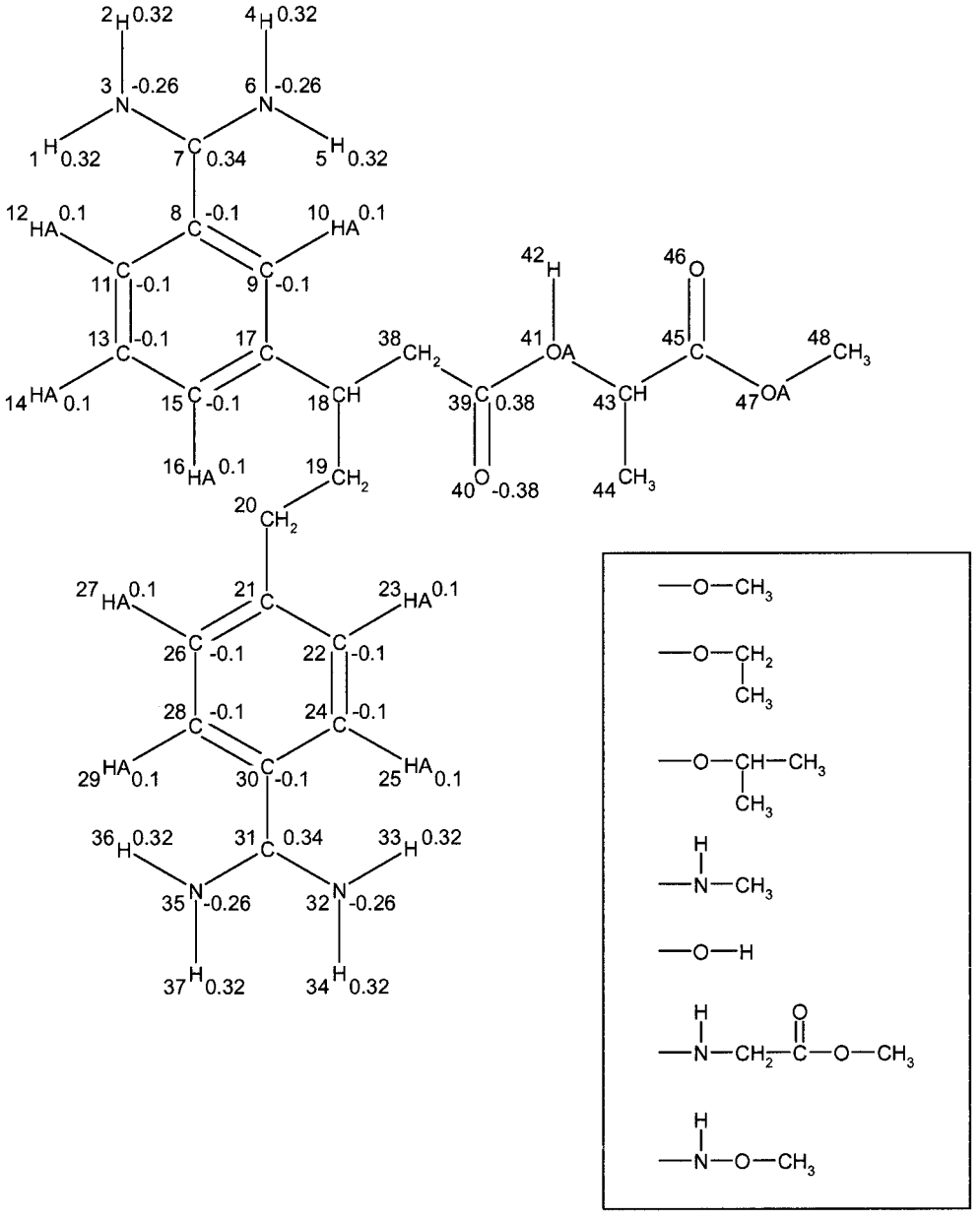

Figure 1. Reference ligand and the seven real ligands of the protein factor Xa considered in this study. Atom numbers (left label) and partial charges in *e* (right label). If no charge is quoted, it is zero. In the reference state, atoms 41–48 are treated as soft atoms. For the real ligands, the drawn fragments replace these atoms. Force field parameters for the reference ligand are listed in Table 1, those for the real ligands in Table 2.

allowed for a time step of 2 fs. Atomic coordinates were stored every 0.1 ps for later analysis and application of eq. (1).

After a total of approximately 1.5 ns of the first protein simulation, the dihedral angle defined by atoms 18, 38, 39, and 40 in the reference ligand (see Fig. 1) was rotated to a value of 180° by applying a dihedral angle restraint during 20 ps of simulation, after which the torsional barrier was increased to 33.5 kJ/mol to prevent the dihedral from returning to *cis*. Another 1 ns of simulation was carried out for the reference ligand with this dihedral angle barrier for reasons that will become apparent in the Discussion. The two different orientations of the reference ligand are shown in Figure 2.

Results

All analyses were carried out over simulations of 2 ns of the reference ligand in solvent and 2 × 1 ns of the reference ligand bound to the protein, as outlined in the previous section. Figure 3

Table 1. Force Field Parameters for the Reference Ligand Shown in Figure 1.

Atom name	H	HA	C	CH	CH2	CH3	N	O	OA
van der Waals type	18	17	11	12	13	14	9	1	3

Bond	Type
OA—H, N—H	2
C—HA	3
C═O	4
C—N	10
C—OA, CH—OA, CH_3—OA	12
C—C	15
C—CH, C—CH2, CH—CH2, CH—CH3, CH2—CH2	26

Bond angle	Type
C—OA—CH3	9
C—CH—CH2, C—CH—CH3, CH2—CH—CH2, OA—CH—C, OA—CH—CH3, C—CH2—CH, C—CH2—CH2, CH—CH2—CH2	14
CH—OA—H	17
CH—C—OA, CH2—C—OA	18
C—N—H	22
H—N—H	23
C—C—HA	24
C—C—C, C—C—CH, C—C—CH2	26
N—C—N, C—C—N	27
CH—C—O, CH2—C—O	29
C—OA—CH	30
C—OA—H	31
O—C—OA	32

Torsional dihedral	Type
43—45—47—48	3
1—3—7—8, 4—6—7—8, 3—7—8—9, 24—30—31—32, 30—31—32—33, 30—31—35—36, 40—39—41—43	4
17—18—19—20, 18—19—20—21, 17—18—38—39	17
9—17—18—19, 19—20—21—22, 18—38—39—40, 39—41—43—45, 41—43—45—47	20

Improper dihedral	Type
Centered on N, centered on C, aromatic C—C—C—C	1
18—19—17—38	2
Centered on atom 43	No improper

All types correspond to those of the GROMOS96 force field.[16,17] Atoms are listed by types or by their numbers according to Figure 1. For the real ligands (inset in Figure 1) see Table II.

shows the atom-positional root-mean-square displacement of the protein backbone atoms for both protein simulations with respect to the initial X-ray structure. The deviation from the X-ray structure fluctuates steadily around a value of 0.155 nm, a reasonable value indicating a structurally stable simulation. The calcium ion remains complexed to Asp70, Asn72, Glu75, and Glu80 throughout the simulations.

Figure 4 shows the distribution of the dihedral angle defined by atoms 18, 38, 39, and 40 (see Fig. 1) for all three simulations. It is clear that the higher torsional barrier in the second protein simulation strongly limits the flexibility around this dihedral angle (dashed line). Interestingly, use of the original size of the torsional barrier induces almost opposite behavior for the reference ligand in solvent (solid line) and when bound to the protein (dotted line). A

Table 2. Force Field Parameters for the Perturbation that Transforms the Reference Ligand into the Seven Real Ligands Shown in the Inset of Figure 1.

Atom number	Reference		OCH3		OC2H5		O-i-C3H7		NHCH3		OH		NHC3O2H5		NHOCH3	
	IAC	q	IAC	q	IAC	q	IAC	q	IAC	q	IAC	q	IAC	q	IAC	q
39	11	0.38	11	0.56	11	0.56	11	0.56	11	0.38	11	0.53	11	0.38	11	0.38
40	1	−0.38	1	−0.38	1	−0.38	1	−0.38	1	−0.38	1	−0.38	1	−0.38	1	−0.38
41	3[a]	0.0	3	−0.36	3	−0.36	3	−0.36	5	−0.28	3	−0.548	5	−0.28	5	−0.1
42	18[a]	0.0	19	0.0	19	0.0	19	0.0	18	0.28	18	0.398	18	0.28	18	0.28
43	12[a]	0.0	14	0.18	13	0.18	12	0.18	14	0.0	19	0.0	13	0.0	3	−0.36
44	14[a]	0.0	19	0.0	14	0.0	14	0.0	19	0.0	19	0.0	19	0.0	14	0.18
45	11[a]	0.0	19	0.0	19	0.0	14	0.0	19	0.0	19	0.0	11	0.56	19	0.0
46	1[a]	0.0	19	0.0	19	0.0	19	0.0	19	0.0	19	0.0	1	−0.38	19	0.0
47	3[a]	0.0	19	0.0	19	0.0	19	0.0	19	0.0	19	0.0	3	−0.36	19	0.0
48	14[a]	0.0	19	0.0	19	0.0	19	0.0	19	0.0	19	0.0	14	0.18	19	0.0

Bonds	Reference	OCH3	OC2H5	O-i-C3H7	NHCH3	OH	NHC3O2H5	NHOCH3
39—41	12	12	12	12	20	12	20	20
41—43	12	12	12	12	20	12	20	12

Bond angles	Reference	OCH3	OC2H5	O-i-C3H7	NHCH3	OH	NHC3O2H5	NHOCH3
39—41—43	30	11	11	11	30	30	30	30
39—41—42	31	31	31	31	31	11	31	31
41—43—44	14	14	14	14	14	14	14	11

IAC: integer atom code, GROMOS96 nonbonded interaction atom type; q: partial charge in e. For bonds and bond angles GROMOS96 types are listed.
[a]"Soft" interaction with $\alpha_{LJ} = 0.151$.

simulation starting with this dihedral angle at 180° but with the original dihedral angle torsional barrier returns to a value between 270 and 360° within 10 ps (data not shown). No significant (<1.5% of simulation time) hydrogen bonding involving atom 42 of the reference ligand was observed in either of the protein simulations.

The free energy estimate as calculated from eq. (1) can be followed over time, by taking the running average of the exponential. For any real compound A, a favorable conformation will be encountered only occasionally, contributing strongly to this average and inducing a drop in the free energy estimate. This results in the typical saw tooth behavior that is shown in Figure 5 for both protein simulations and the pure solvent simulation. As the simulation progresses and no additional conformations are sampled that significantly lower the free energy estimate, the size of the free energy slowly increases while the size of the successive drops decreases. We note here that the absence of free energy drops does not necessarily mean that no significant conformations are sampled anymore, but merely that the most contributing conformations have been sampled. Additional favorable conformations might not result in an abrupt decrease in the free energy, but do reduce the gradual increase in free energy during the rest of the simulation.

Table 3 lists the final estimates of the free energy values for all three simulations and the relative free energies of binding calculated from the solvent simulation results on the one hand and those of both protein simulations on the other.

Discussion

The free energies in the fifth column of Table 3 that were calculated from the solvent simulation do not match the experimental values for all compounds. The relative free energies for the three aliphatic sidechains (atoms numbered 41–48 in Fig. 1) correspond reasonably well to the experimental values, but for the compounds with a real hydrogen atom at position 42 in the reference ligand too favorable free energies are calculated, except for the largest compound in the set, $NHC_3O_2H_5$. This led us to suspect that the orientation of the sidechain of the inhibitor in the inhibitor-protein complex might be different for the individual inhibitor compounds. As can be seen from Figure 4, the softness of the atoms in the unphysical reference ligand is not sufficient to sample a broad distribution of the dihedral angle defining the orientation in the protein simulation. In order to determine whether the system is caught in a local minimum or the observed dihedral angle distribution corresponds to the true dynamical behavior of the unphysical reference ligand sidechain, we have restrained the dihedral angle to a value of 180° during 20 ps. Within 10 ps of removing the restraint, the dihedral angle moves back to values around 315°,

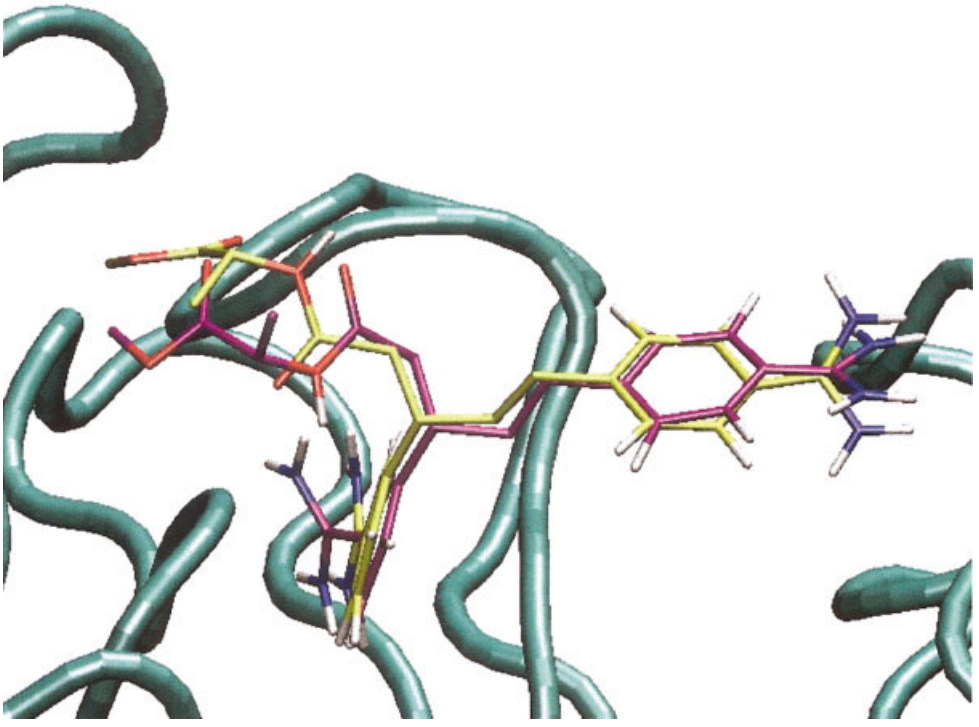

Figure 2. Different orientations of the soft atoms of the ligand due to the 180° rotation around the dihedral angle defined by atoms 18, 38, 39, and 40 of the reference ligand. The backbone of the protein is shown as a solid tube, reference ligand in *trans* conformation in purple, reference ligand in *cis* conformation in yellow.

showing that there are no significant barriers that keep the dihedral angle in a local minimum. However, it is still very well possible that orientations of the unphysical reference ligand sidechain exist that do not correspond to a low-energy region of the reference state, but would have low energy for some of the real ligands. For this reason, the second simulation of the reference ligand bound to the protein was carried out, harmonically restraining the mentioned dihedral angle to 180° during 20 ps and increasing the height of the dihedral angle barrier to 33.5 kJ/mol to prevent it from turning to *cis*. It is clear from Figure 4 that this high barrier height reduces the sampling of the dihedral angle even further, but in a completely different region, so that a comparison between the free energies obtained from each simulation can be made.

The eighth column in Table 3 lists the relative free energies calculated using the second protein simulation. With the exception of the hydroxyl group, the experimental ranking of the binding energies of the ligands is reproduced correctly. Moreover, the relative free energies for the compounds $NHCH_3$ and $NHOCH_3$, when compared to those for the aliphatic groups, are reasonable, especially considering the fact that aminoxy-groups are known to have C—N—O—C dihedral angle values around 90°, which is not observed in the simulations of the reference ligand. The discrepancy observed for the remaining two compounds could be explained by their respective sizes. The $—NHC_3O_2H_5$ group occupies seven of the eight soft atoms in the reference ligand. A favorable conformation of the reference ligand for this compound thus only occurs if these seven atoms do not show any overlap with surrounding atoms. The chances of finding such conformations are small, particularly in the protein surroundings, where there might be only a few conformations of the reference ligand for which this situation is possible. This could, therefore, result in a relatively large relative free energy of binding. From the sixth column in Table 3 we can see that for the last two compounds larger mutation free energies in the protein have indeed been calculated. Apparently, a larger group can only be accommodated in the protein if atom number 42 in the reference ligand does not get a positive charge. In the case of $NHOCH_3$ we observe a similar effect in the solvent simulation, resulting in a reasonable value for the relative free energy. In the solvent simulation, the only charges that are present in the reference state sidechain are on the carbonyl atoms, according to which the water molecules will be oriented. The $NHOCH_3$ group has an inverted dipole with respect to all other groups, resulting in a distinctly different free energy of mutation in this medium, that is, more unfavorable.

This consideration brings about another short coming of the one-step perturbation method for these compounds, namely the fact that the soft atoms are treated as neutral atoms in the reference state. In principle, it would have been possible to put (soft) charges on these atoms as well, but this was not done for two reasons. In the first place, it would not at all have been straightforward to decide which charges should be used. The reference state should be as close as possible to, for example, the OCH_2CH_3 and the $NHOCH_3$ groups, which have completely different charge distributions. Similarly, there are reasonably large partial charges on the

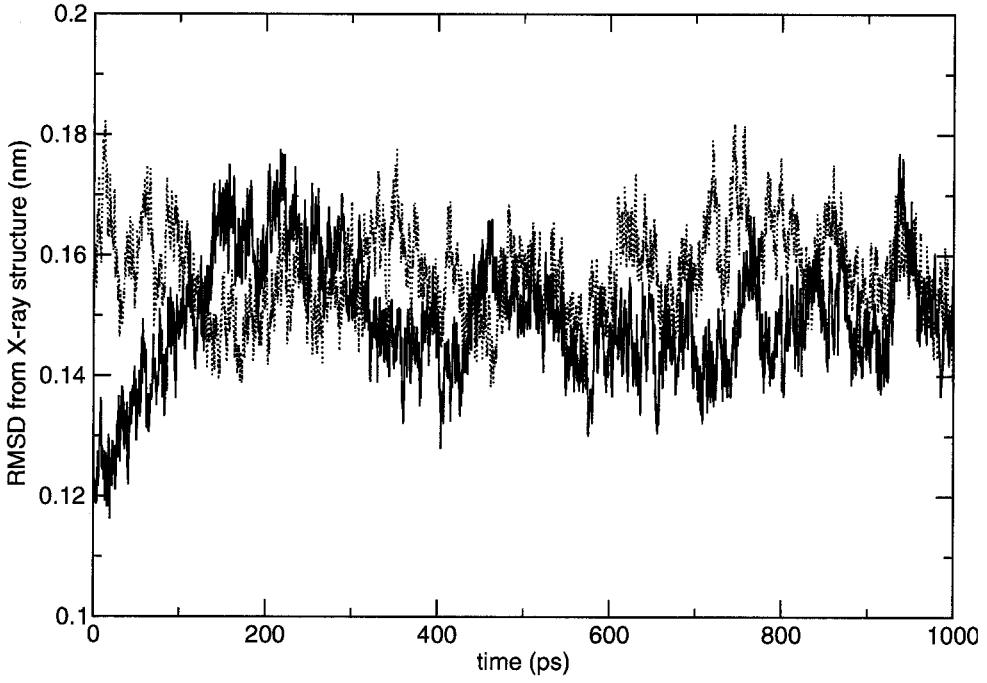

Figure 3. Atom-positional root-mean-square deviations (RMSD) for the protein backbone atoms with respect to the X-ray structure of the protein factor Xa. The solid line corresponds to the first nanosecond of simulation, the dotted line to the second nanosecond, with a particular dihedral angle in the reference ligand changed to a *trans* conformation.

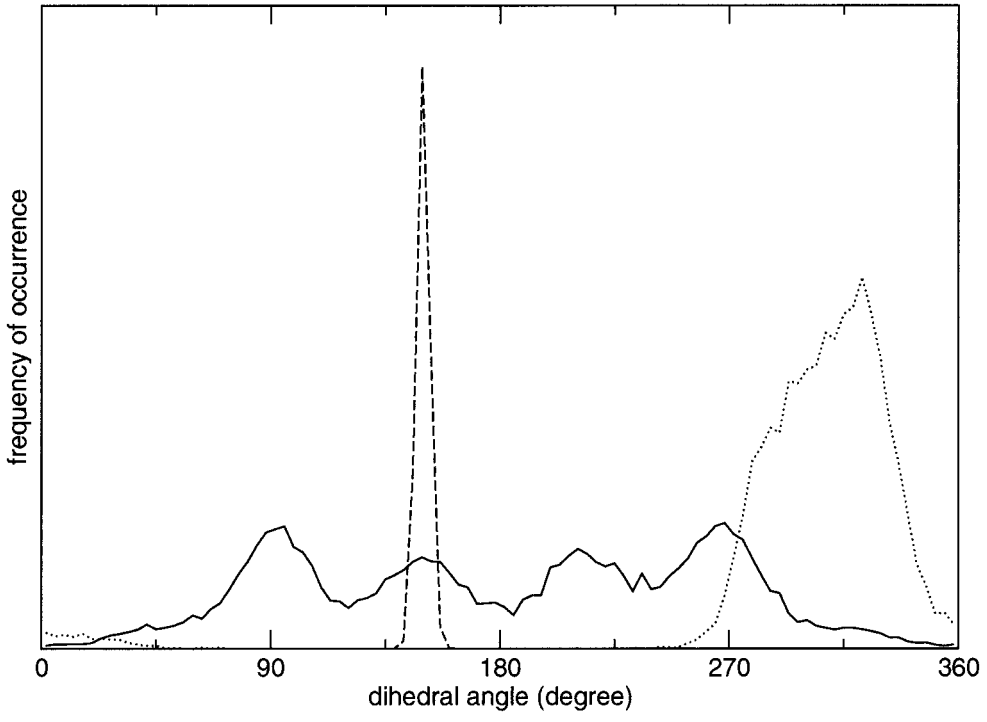

Figure 4. Distribution of the values of the dihedral angle defined by atoms 18, 38, 39, and 40 (Fig. 1) in the reference ligand from the solvent simulation (solid line) and the protein simulations (dotted line: first ns; dashed line: second ns). For clarity, the height of the dashed line has been reduced by a factor of 5.

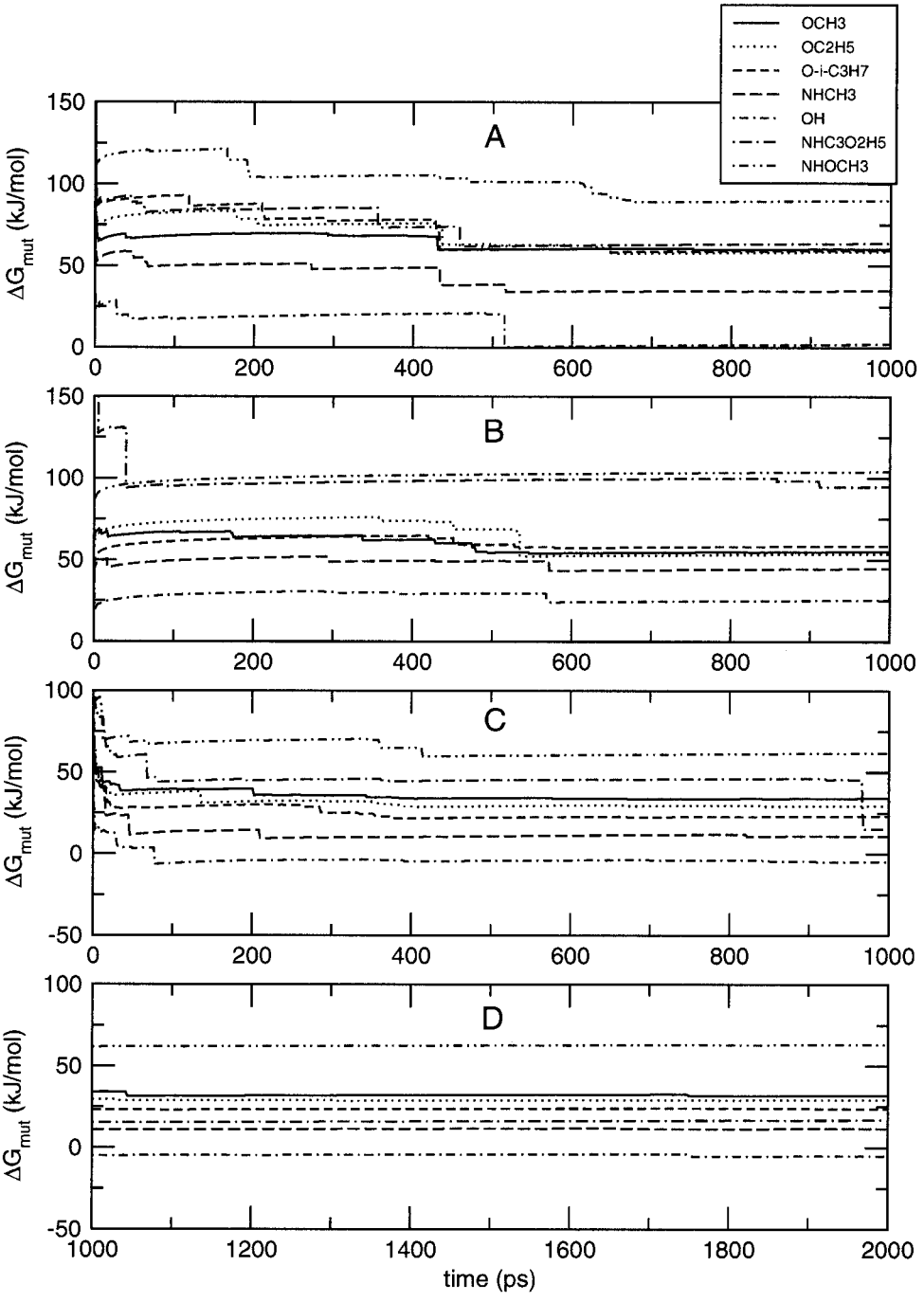

Figure 5. Development of the free energy difference between the real ligands and the reference ligand ΔG_{mut} (see Table 3) as a function of time for all real ligands, in both protein simulations and in pure solvent. (A) Initial simulation of ligand-protein complex. (B) Second simulation of ligand-protein complex in which a particular dihedral angle in the ligand is kept near the *trans* conformation. (C and D) Simulation of the ligand in pure solvent.

ester group in $NHC_3O_2H_5$, whereas the same atoms are neutral for all other compounds. In the second place, it has been shown[3] that in the case of soft charges, water molecules are likely to get caught inside the soft parts of the molecules, aligning either directly with the soft charges or with the first shell of the polarized surroundings. Without charges on the soft atoms, however, the reference

Table 3. Free Energy Estimates Calculated Using Eq. (1).

	Ligand in pure water	Ligand-protein complex in water						
		Protein simulation 1			Protein simulation 2			Exp.
Compound	ΔG_{mut}	ΔG_{mut}	$\Delta\Delta G$	$\Delta\Delta G'$	ΔG_{mut}	$\Delta\Delta G$	$\Delta\Delta G'$	$\Delta\Delta G'$
OCH3	31.9	61.6	29.7	0.0	56.2	24.3	0.0	0.0
OC2H5	29.3	57.9	28.6	−1.1	55.4	26.1	1.8	0.53
O-i-C3H7	23.8	60.2	36.4	6.7	55.0	31.2	6.9	4.3
NHCH3	11.7	22.7	11.0	−18.7	46.2	34.5	10.1	8.3
OH	−5.1	3.8	8.8	−20.9	23.0	28.1	3.7	9.3
NHC3O2H5	16.9	63.6	46.6	16.9	96.3	79.3	55.0	10.6
NHOCH3	63.1	90.4	27.3	−2.5	105.0	41.9	17.6	10.6

Columns 2, 3, and 6 contain the free energies ΔG_{mut} of mutating the reference ligand into the real ligands, in pure solvent and when bound to the protein in solution (two simulations), respectively. Columns 4 and 7 are relative free energies of binding $\Delta\Delta G = \Delta G_{mut}$ (complex) $- \Delta G_{mut}$ (pure solution), using the two different simulations of the ligand-protein complexes. Columns 5 and 8 contain these quantities relative to the one for OCH_3 ($\Delta\Delta G'$). Column 9 lists the experimental values taken from ref. 13. All values in kJ/mol.

state lacks the electrostatic characteristics of the real ligands and thus the ensemble that is generated for the reference state will resemble the ensembles for the real ligands less closely. For example, no significant hydrogen bonding was observed between the protein and reference ligand involving atom 42 of the reference ligand. For a neutral part of the unphysical reference ligand, this is not surprising, but it is unlikely that none of the compounds with a real hydrogen atom would form a hydrogen bond.

In summary, three important lessons can be learned from these simulations regarding the design of an unphysical reference state for the calculation of free energies. First, the flexibility of the reference ligand should be as similar as possible to that of the real ligands one is interested in. Because the dynamical behavior of soft atoms can be quite different from that of normal atoms, a successful one-step free energy calculation is more likely to occur for more rigid compounds. Yet, for flexible compounds one could either choose the reference state such that a particular degree of freedom, for example, the dihedral (18 − 38 − 39 − 40) in the present case, is more uniformly sampled, or add a second reference state to this end. Second, one should be aware of the fact that the size of the unphysical part of the reference ligand should be limited if one intends to sample conformations that give space to all soft atoms simultaneously. This was also observed in a previous study where inaccuracies arose when trying to mutate six out of nine soft atoms into normal atoms.[24] This limitation could be alleviated either by reducing the soft-core interactions thereby enhancing the sampling, or by adding a second reference state for large compounds in which the size of the soft part is reduced. Third, one should try to reproduce the character of the real ligands in the reference ligand. In the ideal case the reference ligand does not only resemble the real ligands in their topology, but also in their electrostatic behavior. This further limits the allowed differences between the real ligands. Again, one could solve this problem by adding a slightly polar second reference state to the original neutral, nonpolar one in order to broaden the sampling.

Conclusion

Previous studies have shown that accurate free energy differences can be calculated from a simulation of an unphysical reference state by applying a single perturbation.[1–5] In the present study we have applied the one-step perturbation method to a set of ligands that differ in one relatively large group of atoms, with a variety of electrostatic and dynamical behavior. Three limitations to the design of the ideal reference ligand have been demonstrated and discussed, being its size, its flexibility, and its electrostatic character. Because the one-step perturbation technique has as major assets its computational efficiency and the possibility to identify ligand conformations that significantly contribute to the binding free energy, it seems worthwhile to further attempt to push back the accuracy-limiting components of the technique by increasing the range of its sampling either through the use of torsional angle potential energy terms that enhance sampling, or through the use of one or two additional reference states that, when combined, encompass the various polarities of the compounds.

References

1. Liu, H. Y.; Mark, A. E.; van Gunsteren, W. F. J Phys Chem 1996, 100, 9485.
2. Schäfer, H.; van Gunsteren, W. F.; Mark, A. E. J Comput Chem 1999, 20, 1604.
3. Pitera, J. W.; van Gunsteren, W. F. J Phys Chem B 2001, 105, 11264.
4. Mark, A. E.; Xu, Y. W.; Liu, H. Y.; van Gunsteren, W. F. Acta Biochim Pol 1995, 42, 525.
5. Oostenbrink, B. C.; Pitera, J. W.; Van Lipzig, M. M. H.; Meerman, J. H. N.; van Gunsteren, W. F. J Med Chem 2000, 43, 4594.
6. Beutler, T. C.; Mark, A. E.; Van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Chem Phys Lett 1994, 222, 529.
7. A literature search on "factor Xa inhibitors" yielded 167 articles, published in the last 5 years. See, for example, refs. 8–12.

8. Matter, H.; Defossa, E.; Heinelt, U.; Blohm, P. M.; Schneider, D.; Müller, A.; Herok, S.; Schreuder, H.; Liesum, A.; Brachvogel, V.; Lönze, P.; Walser, A.; Al-Obeidi, F.; Wildgoose, P. J Med Chem 2002, 45, 2749.
9. Ewing, W. R.; Becker, M. R.; Manetta, V. E.; Davis, R. S.; Pauls, H. W.; Mason, H.; Maignan, S.; Guilloteau, J. P.; Brown, K.; Colussi, D.; Bentley, R.; Bostwick, J.; Kasiewski, C. J.; Morgan, S. R.; Leadley, R. J.; Dunwiddie, C. T.; Perrone, M. H.; Chu, V. J Med Chem 1999, 42, 3557.
10. Herron, D. K.; Goodson, T. J.; Wiley, M. R.; Weir, L. C.; Kyle, J. A.; Yee, Y. K.; Tebbe, A. L.; Tinsley, J. M.; Mendel, D.; Masters, J. J.; Franciskovitch, J. B.; Sawyer, J. S.; Beight, D. W.; Ratz, A. M.; Milot, G.; Hall, S. E.; Klimkowski, V. J.; Wikel, J. H.; Eastwood, B. J.; Towner, R. D.; Gifford-Moore, D. S.; Craft, T. J.; Smith, G. F. J Med Chem 2000, 43, 859.
11. Klein, S. I.; Czekaj, M.; Gardner, C. J.; Guertin, K. R.; Cheney, D. L.; Spada, A. P.; Bolton, S. A.; Brown, K.; Colussi, D.; Heran, C. L.; Morgan, S. R.; Leadley, R. J.; Dunwiddie, C. T.; Perrone, M. H.; Chu, V. J Med Chem 1998, 41, 437.
12. Han, Q.; Dominguez, C.; Stouten, P. F. W.; Park, J. M.; Duffy, D. E.; Galemmo, R. A. J.; Rossi, K. A.; Alexander, R. S.; Smallwood, A. M.; Wong, P. C.; Wright, M. M.; Luettgen, J. M.; Knabb, R. M.; Wexler, R. R. J Med Chem 2000, 43, 4398.
13. Maduskuie, T. P. J.; McNamara, K. J.; Ru, Y.; Knabb, R. M.; Stouten, P. F. W. J Med Chem 1998, 41, 53.
14. Beveridge, D. L.; DiCapua, F. M. Ann Rev Biophys Biophys Chem 1989, 18, 431.
15. Kollman, P. Chem Rev 1993, 93, 2395.
16. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krüger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Biomolecular Simulation: The GROMOS96 Manual and User Guide; Vdf Hochschulverlag AG an der ETH Zürich: Zürich, 1996.
17. Schuler, L. D.; Daura, X.; van Gunsteren, W. F. J Comput Chem 2001, 22, 1205.
18. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Intermolecular Forces; Reidel: Dordrecht, The Netherlands, 1981; p 331.
19. Maignan, S.; Guilloteau, J. P.; Pouzieux, S.; Choi-Sledeski, Y. M.; Becker, M. R.; Klein, S. I.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. J Med Chem 2000, 43, 3226.
20. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. J Chem Phys 1984, 81, 3684.
21. Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. J Chem Phys 1995, 102, 5451.
22. Heinz, T. N.; van Gunsteren, W. F.; Hünenberger, P. H. J Chem Phys 2001, 115, 1125.
23. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327.
24. Oostenbrink, C.; van Gunsteren, W. F., accepted for publication in Proteins 2003.