Journal of Computer-Aided Molecular Design, **12:** 383–396, 1998.
KLUWER/ESCOM

Conformational energy penalties of protein-bound ligands

Jonas Boström , Per-Ola Norrby & Tommy Liljefors*
Department of Medicinal Chemistry, Royal Danish School of Pharmacy, Universitetsparken 2, DK-2100 Copenhagen, Denmark

Received 25 November 1997; Accepted 21 January 1998

Key words: AMBER*, bioactive conformation, conformational analysis, GB/SA hydration model, ligand–protein interactions, MM3*

Summary

The conformational energies required for ligands to adopt their bioactive conformations were calculated for 33 ligand–protein complexes including 28 different ligands. In order to monitor the force field dependence of the results, two force fields, MM3* and AMBER*, were employed for the calculations. Conformational analyses were performed in vacuo and in aqueous solution by using the generalized Born/solvent accessible surface (GB/SA) solvation model. The protein-bound conformations were relaxed by using flat-bottomed Cartesian constraints. For about 70% of the ligand–protein complexes studied, the conformational energies of the bioactive conformations were calculated to be $\leq$3 kcal/mol. It is demonstrated that the *aqueous conformational ensemble* for the unbound ligand must be used as a reference state in this type of calculations. The calculations for the ligand–protein complexes with conformational energy penalties of the ligand calculated to be larger than 3 kcal/mol suffer from uncertainties in the interpretation of the experimental data or limitations of the computational methods. For example, in the case of long-chain flexible ligands (e.g. fatty acids), it is demonstrated that several conformations may be found which are very similar to the conformation determined by X-ray crystallography and which display significantly lower conformational energy penalties for binding than obtained by using the experimental conformation. For strongly polar molecules, e.g. amino acids, the results indicate that further developments of the force fields and of the dielectric continuum solvation model are required for reliable calculations on the conformational properties of this type of compounds.

Introduction

Molecules which bind to proteins are in general flexible and may change their conformation when binding to the protein. Thus, the bioactive conformation of the protein-bound ligand does not necessarily have to be the global energy minimum conformation for the unbound molecule, not even a local energy minimum conformation. As the difference in conformational free energy of the ligand between its bound and unbound states subtracts from the free energy of binding due to intermolecular interactions between the protein and the ligand (hydrogen bonding, van der Waals interactions, etc.), the conformational energy penalty may significantly affect the affinity of the ligand. Each 1.4 kcal/mol of increased conformational energy of the bioactive conformation leads to a decrease in affinity by a factor of 10 [1,2]. Thus, the conformational energy required for the ligand to adopt its bioactive conformation is an important factor in structure–activity studies [3,4] and in computer-aided ligand design [1]. In 3D pharmacophore identification for ligands at receptors with unknown 3D structures, an important problem is how much above the global energy minimum we must go in a conformational search in order to be reasonably sure to have included the bioactive conformation [2]. This problem is also highly relevant in molecular alignments for 3D-QSAR studies [5]. In structure-based de novo ligand design, it is of importance to assure that the designed ligand does not have a prohibitively high conformational energy.

*To whom correspondence should be addressed.

Various estimates of the upper conformational energy limit to be used in a search for candidates for a bioactive conformation have been given in the literature. In the context of pharmacophore identification, Marshall and Motoc [6] find that reasonable estimates vary in the range of 5–20 kcal/mol. Siebel and Kollman [1] find it reasonable to expect that the bioactive conformations for highly active molecules have conformational energy penalties of less than 4–5 kcal/mol (15–20 kJ/mol). Liljefors and Pettersson [2] advocate the use of an upper conformational energy limit of 2.5–3 kcal/mol (10–12 kJ/mol) in 3D pharmacophore identification.

The rapidly increasing number of protein–ligand complexes included in the Brookhaven National Laboratory Protein Data Bank (PDB) [7] and the development of powerful conformational search methods have made it possible to calculate the conformational energy difference between the bound and unbound states for a number of ligands. Recently, Nicklaus et al. [8] reported a study on the conformational energies of ligands in ligand–protein complexes present in the PDB using the CHARMm force field. In this study, only ligands which are also present in the Cambridge Structural Database were considered. For 27 flexible ligands, the calculated conformational energy differences between protein-bound and unbound conformations of the ligands varied between 0 and 39.7 kcal/mol, with an average of 15.9 kcal/mol. Previously, Spark et al. [9], using three different experimentally determined methotrexate–dihydrofolate reductase complexes, calculated conformational energy penalties for methotrexate in the range of 27–71 kcal/mol. However, it should be noted that the results obtained in these studies are based on the *gas-phase conformational ensemble* for the unbound ligand and thus (vide infra) the results can not be related to experimentally observed affinities and binding energies, in which case the *conformational ensemble in aqueous solution* must be used. The high conformational energies obtained in the studies referred to above most probably result from an electrostatic 'collapse' of the ligand in vacuo (vide infra). The calculated energy for the protein-bound conformation is particularly high for molecules with polar atoms or highly charged groups which could give rise to strong intramolecular interactions in the gas-phase.

In order to get information on the magnitude of conformational energies to be expected for protein-bound ligands, we have in the present work investigated 33 ligand–protein complexes including 28 different ligands. For each ligand–protein complex, the conformational energy differences between the conformation of the ligand bound to the protein and the preferred conformation of the ligand in aqueous solution have been calculated.

Methods

Selection of ligand–protein complexes

The protein–ligand complexes studied in the present work were taken from the PDB. The following criteria were applied in the selection: (i) the X-ray structure of the protein–ligand complex should have high resolution, preferably better than 2.0 Å; (ii) the ligand should be reasonably small and flexible; and (iii) the ligand should be suitable for conformational analysis by using force field methods. Thirty-three protein–ligand complexes were selected containing 28 different ligands. Data for these complexes are given in Table 1. For 27 of the 33 complexes, the resolution is $\leq$2.1 Å and the overall average resolution is 1.9 Å. The structures of the ligands are shown in Figure 1. Note that the phenylimidazole ligand in the PDB file 1phe contains an error in the assignment of atom types. The ligand should be 2-phenylimidazole as in Ref. 10. The ligands were extracted from the ligand–protein complex, bond orders were assigned and hydrogens added in MacroModel v.5.0 [11]. Hydroxyl torsions were determined from an examination of possible hydrogen bonding patterns in the protein–ligand complex using INSIGHTII v. 95.0 [12].

Computational procedures

Force fields

In order to monitor the force field dependence of the calculated results, two different force fields, MM3* and AMBER* as implemented in MacroModel v. 5.0 [11], were employed in all calculations. MM3* and AMBER* have been proven to be two of the best force fields in reproducing experimental conformational energy data [13]. All calculations were run on an SGI Indigo2 workstation. The force fields were used with default settings, except for a few parameter extensions. For structures containing pyridines (compounds **5** and **7**, Figure 1), the 'benzene' substructure in MM3* was replaced with the 'benzenoid' and 'pyridinoid' substructures from MacroModel v. 5.5. For compound **9**, the parameters for the torsion around the phenyl-N^+

Table 1. The ligands studied, names of the PDB files, crystallographic parameters and the protein in the ligand–protein complex

No.	Ligand	PDB file	Resolution (Å)	R-factor	Protein
1	Ethylbenzene	1nhb	1.7	0.165	Lysozyme mutant
2	Isobutylbenzene	184l	1.8	0.162	Lysozyme mutant
3	Butylbenzene	186l	1.8	0.164	Lysozyme mutant
4a	5-Phenylimidazole	1phd	1.6	0.190	Cytochrome P450-cam
4b	5-Phenylimidazole	1phe	1.6	0.190	Cytochrome P450-cam
4c	2-Phenylimidazole	1phf	1.6	0.190	Cytochrome P450-cam
5	Metyrapone	1phg	1.6	0.190	Cytochrome P450-cam
6	Dichloroethane	2dhc	2.4	0.195	Haloalkane dehalogenase
7	Tacrine	1acj	2.8	0.195	Acetylcholine esterase
8	Decamethonium	1acl	2.8	0.199	Acetylcholine esterase
9	Edrophonium	1ack	2.8	0.199	Acetylcholine esterase
10	Aminomethylcyclohexane	1tng	1.8	0.172	Trypsin
11	Tranylcypramine	1tnl	1.9	0.164	Trypsin
12	*p*-Fluorobenzylamine	1tnh	1.8	0.170	Trypsin
13	2-Phenylethylamine	1tnj	1.8	0.170	Trypsin
14	3-Phenylpropylamine	1tnk	1.8	0.171	Trypsin
15	4-Phenylbutylamine	1tni	1.9	0.161	Trypsin
16a	Retinoic acid	1cbs	1.8	0.200	Retinoic acid binding protein
16b	Retinoic acid	1epb	2.2	0.182	Retinoic acid binding protein
17	Retinol	1crb	2.1	0.188	Retinol binding protein
18	Biotin	1stp	2.6	0.22	Streptavidin
19	Myristic acid	1icm	1.5	0.159	Fatty acid binding protein
20a	Palmitic acid	2ifb	2.0	0.178	Fatty acid binding protein
20b	Palmitic acid	1lie	1.6	0.198	Adipocyte lipid binding protein
21a	Stearic acid	1hmt	1.4	0.114	Fatty acid binding protein
21b	Stearic acid	1lif	1.6	0.180	Adipocyte lipid binding protein
22a	Oleic acid	1hms	1.4	0.121	Fatty acid binding protein
22b	Oleic acid	1lid	1.6	0.172	Adipocyte lipid binding protein
23	Elaidic acid	1hmr	1.4	0.132	Fatty acid binding protein
24	Chloramphenicol	3cla	1.75	0.157	Acetyltransferase
25	Thyroxine	1eta	1.7	0.184	Transthyretin
26	Histidine	1lag	2.06	0.162	Lys, Arg, Orn binding protein
27	Ornithine	1lah	2.06	0.158	Lys, Arg, Orn binding protein

bond were lacking in MM3* as well as in AMBER*. Semi-empirical AM1 calculations indicate that the torsional profiles are very similar for rotation about the $C(sp^2)$-$C(sp^3)$ bond in *t*-butyl benzene and the $C(sp^2)$-N^+ bond in trimethylammonium benzene. Therefore, the required torsional parameters were copied from the corresponding hydrocarbon parameters. The protonation state in neutral aqueous solution (pH = 7) was used for the calculations on acids and bases.

Constrained relaxation of the protein-bound ligand structures

Crystallographically determined atomic positions in ligand–protein structures contain significant uncertainties. As the energies of the protein-bound ligands are to be compared with energies calculated for fully relaxed structures of the unbound ligand, it is necessary to remove incompatibilities of, in particular, bond lengths between those displayed by the X-ray structure and the ones calculated by the force field method. The large force constants for bond stretching/compression

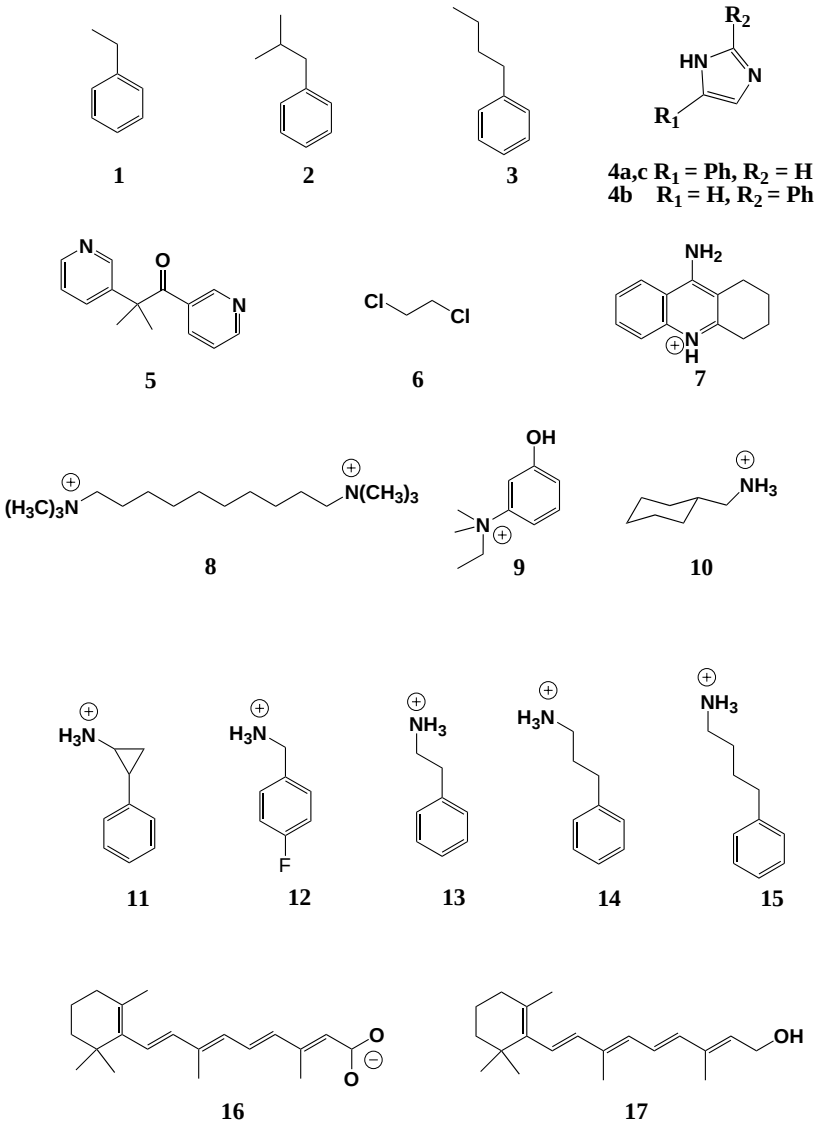

Figure 1. Structures of the ligands studied in this work.

would otherwise severely distort the energy calculations. The crystallographic structures were therefore partially relaxed within each force field. Two methods of relaxation were used. In the first method, which was also used by Nicklaus et al. [8] all non-hydrogen dihedral angles in the X-ray structure were constrained by a harmonic potential with a force constant of 240 kcal/mol deg^2 (1000 kJ/mol deg^2), whereas the remaining degrees of freedom were optimized. Dihedral angles in ring systems, however, were fully relaxed. In the second relaxation method, all non-hydrogen atoms were tethered to the crystallographic positions by harmonic, flat-bottomed Cartesian constraints, whereupon the molecule was minimized. The flat-bottomed radius, i.e. the distance each atom is allowed to move from the tether position before an energy penalty is applied, was set to 0.3 Å. At larger distances, a harmonic penalty function with a force constant of 120 kcal/mol Å^2 (500 kJ/mol Å^2) was applied. In a few cases (indicated in Tables 2 and 3), the radius was set to 0.4 Å. Hydroxyl torsions (determined from plausible hydrogen bonding patterns) were in all cases restrained by dihedral constraints.

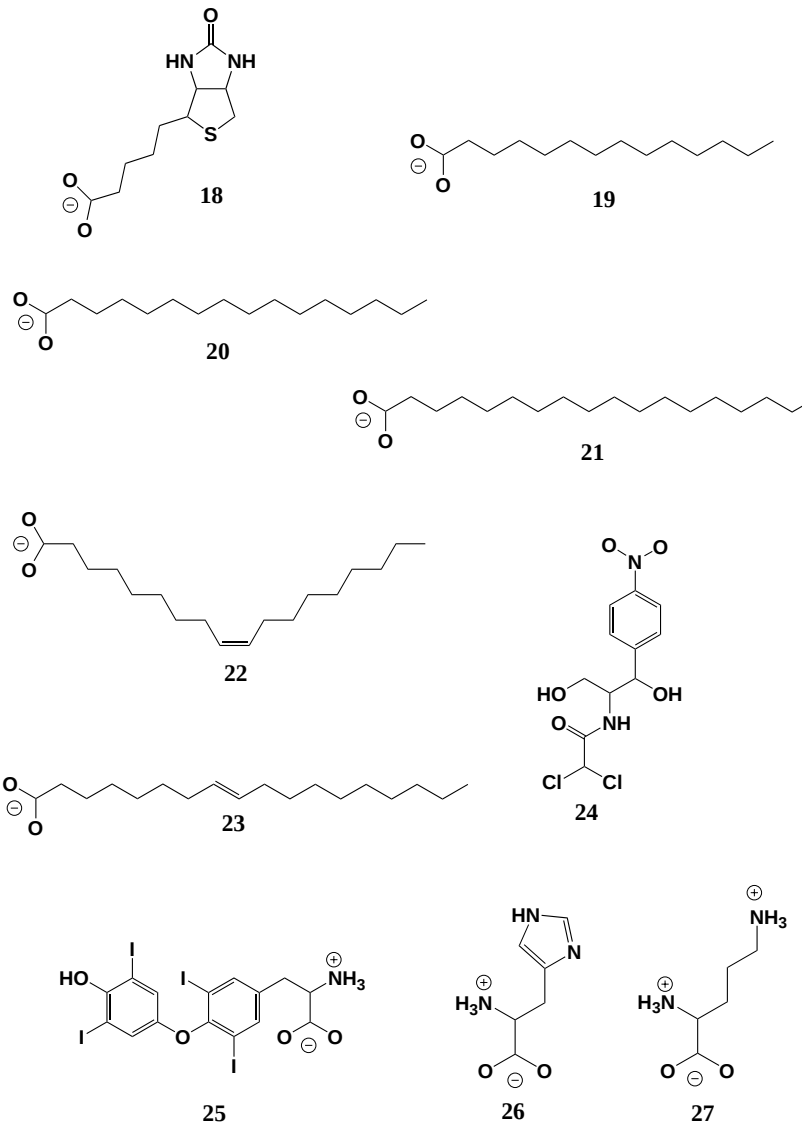

Figure 1. (Continued) Structures of the ligands studied in this work.

Conformational analysis of the ligands

The conformational space was searched by using the Monte Carlo multiple minimum (MCMM) method [14] implemented in the MacroModel program. All heavy atoms and hydroxyl protons were superimposed in the test for duplicate structures. The energy minimizations were carried out employing the truncated Newton conjugate gradient (TNCG) algorithm. The conformational searches were continued until all low energy minima had been found multiple times. In cases for which two (or more) ring conformations were possible, the conformational search was shortened by using the low energy ring conformations as starting structures. In the general case, any structure with an energy of more than 12 kcal/mol (50 kJ/mol) above the current global minimum was discarded. In order to speed up the conformational searches for the long-chain fatty acids **19**–**23**, two energy windows were set: (i) any structure which after 20 iterations had an energy of more than 7 kcal/mol higher than the current lowest energy structure was rejected; and (ii) any minimized structure with an energy of more than 2.4 kcal/mol above the current minimum energy conformation was discarded.

For each ligand, the conformational analysis was performed for aqueous solution as well as *in vacuo*.

The calculations for aqueous solution were performed using the generalized Born/solvent accessible surface area (GB/SA) continuum dielectric solvation model developed by Still et al. [15] and implemented in the MacroModel program.

Calculation of the conformational energy penalty for the protein-bound ligand

The conformational energy penalty for each protein-bound ligand was calculated by subtracting the internal (steric) energy of the preferred conformation in aqueous solution (i.e. the energy of the global energy minimum in aqueous solution excluding the hydration energy) from the calculated energy of the protein-bound ligand after constrained structure optimization. Thus, conformational entropy effects were not taken into account; the conformational ensemble was in each case represented by the lowest energy conformation obtained by the conformational analysis. In cases where several low energy conformers of the ligand are present in aqueous solution, the neglect of conformational entropy leads to an underestimation of the conformational energy penalty for the protein-bound ligand. This underestimation is expected to increase as the number of rotatable bonds in the ligand increases.

Thermodynamics of ligand binding

The binding of a ligand to a protein consists of several steps, all of which must be correctly evaluated in order to predict absolute or relative binding constants. The overall free energy of binding is a simple function of the equilibrium constant (Equation (1)):

$$\Delta G = -RT \ln K \tag{1}$$

A thermodynamic cycle describing the complete binding event is depicted in Figure 2. It is important to realize that ligand binding takes place in an aqueous environment; the solvated ligand–protein complex will be in equilibrium with unbound solvated ligand and solvated protein. The actual binding will consist of (partial) desolvation of both ligand and protein, a conformational change, and formation of binding intermolecular interactions. The transformations depicted in Figure 2 do not correspond to actual physical events, but to computational events whose sum is equivalent to the actual binding event.

Following the thermodynamic cycle in Figure 2, the total free energy of binding can be expressed as a sum of discrete components, Equation (2):

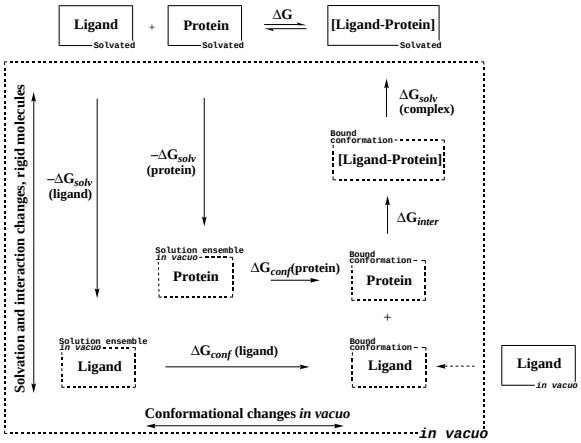

Figure 2. Thermodynamic cycle for ligand binding (solid arrows). The free energy of binding is divided into contributions from intermolecular changes (vertical arrows) and intramolecular (conformational) changes (horizontal arrows). The intermolecular energies include solvation and intermolecular interactions between protein and ligand, evaluated for rigid conformations. Internal energies correspond exactly to conformational free energy changes in vacuo.

$$\begin{aligned}\Delta G \;=\; & \Delta G_{solv} + \Delta G_{conf}(\text{ligand}) \\ & + \Delta G_{conf}(\text{protein}) + \Delta G_{inter}\end{aligned} \tag{2}$$

In Equation (2), ΔG_{solv} represents the overall change in solvation energy in the reaction, i.e. $\Delta G_{solv}(\text{complex}) - \Delta G_{solv}(\text{ligand}) - \Delta G_{solv}(\text{protein})$. The interaction energy ΔG_{inter} is the in vacuo energy required to separate ligand and protein. The conformational free energy ΔG_{conf} pertains to the in vacuo change of the bound conformational ensemble to the solvated conformational ensemble. Note that the gas-phase conformational ensemble (bottom right, Figure 2) is not a part of the thermodynamic cycle.

It must be stressed that all terms in Equation (2) correspond to isolated components of the total binding energy. For example, as opposed to the more common usuage of the same term, the solvation energies in Figure 2 (ΔG_{solv}) *only* include the solvation component of the total energy, not the conformational rearrangement on going between gas-phase and the solvated state. Similarly, the interaction energy between ligand and protein in the complex (ΔG_{inter}) refers purely to the intermolecular interactions, the energy necessary to separate the rigid ligand from the likewise rigid protein. In the context of the current work, we define that the vertical transformations in Figure 2 refer to purely intermolecular changes, i.e. ΔG_{solv} and ΔG_{inter} correspond to solvation and interactions of *rigid* molecules,

whereas horizontal transformations in Figure 2 denote conformational reorganization (changes in internal energy). The top equilibrium includes changes in both intermolecular and internal interactions.

Only the solid boxes in Figure 2 represent real physical states. Thus, the component free energies cannot be directly compared to experimental measurements. It should be noted that the thermodynamic cycle in Figure 2 is not the only way to represent the binding process. It is entirely possible to define a cycle consisting of physical events with both internal and external energy changes. This may be an advantage if single steps in the cycle are to be correlated with experimental measurements (e.g. experimental free energies of solvation). However, the cycle depicted in Figure 2 is more suited to an analysis of the relative importance and accuracies of different computational aspects of ligand binding. In the present work, we will investigate the steric (conformational) energy penalty on going from the solvated to the bound state. It must be stressed that for the thermodynamic cycle to be valid, we evaluate the steric energy in vacuo, but the geometry optimization of the free ligand is performed in the presence of a solvation model.

Results and discussion

The calculated results for ligands **1**–**27** employing the MM3* force field are given in Table 2 and the corresponding results obtained by using the AMBER* force field are shown in Table 3. In addition to the relative energies of the global energy minima in vacuo and in aqueous solution and the conformational energy differences between the protein-bound ligand and its global energy minimum in aqueous solution (ΔE_{conf}), Tables 2 and 3 also include the energy difference between the ligand in its protein-bound conformation and the nearest local energy minimum (ΔE_{local}) obtained by unconstrained energy minimization of the X-ray structure of the ligand in vacuo. A non-zero value of ΔE_{local} indicates that the protein-bound ligand is not in a local energy minimum.

Constrained structure optimizations of the protein-bound ligand

Two types of constraints for constrained structure optimizations of the protein-bound ligands were used in the present work: (i) dihedral constraints; and (ii) flat-bottomed Cartesian constraints (see the Methods section). Comparing the two methods, it was noted that the former often resulted in large Cartesian deviations for extended structures. Since the dihedral angles in such cases frequently were locked in high energy conformations, the energy of the constrained optimized structure was generally higher with the former method. The selection of the restraining method is particularly important for the long-chain fatty acids (**19**–**23**). For example, stearic acid (**21**) is a C_{18} fatty acid with 16 torsional degrees of freedom. If each dihedral angle incorporates a relatively small torsional energy error of 0.5 kcal/mol, the total energy error will be as large as 8 kcal/mol. This problem may be avoided by the use of flat-bottomed Cartesian constraints. This method consistently yielded an optimized structure which should be able to fit into the electron density envelope from X-ray crystallography. The results from flat-bottomed Cartesian constrained relaxations were therefore used in all comparisons of energies in the present study and only the results employing this constraining technique are reported. The structures obtained after constrained relaxation of the crystallographic ligand structures were least-squares superimposed on the unmodified X-ray structures (including all heavy atoms). The average rms deviations obtained by these superimpositions were for both force fields 0.22 Å. An example of a superimposition with an rms deviation of 0.21 Å is given in Figure 3.

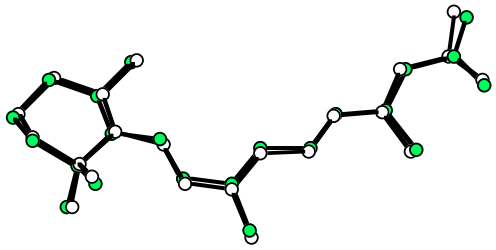

Figure 3. A least-squares superimposition of the unmodified X-ray structure of the protein-bound ligand **16b** (unfilled atoms) and the corresponding constrained optimized structure using flat-bottomed Cartesian constraints with a half-width of 0.3 Å. The rms value is 0.21 Å.

Hydration effects on the conformational properties of the ligand

For non-polar ligands as compounds **1**–**3** or molecules with polar functional groups but without the possibility of forming strong polar intramolecular interactions (hydrogen bonds, cation–π interactions, etc.) such as **4**–**12**, **16**, **17** and **19**–**23**, the energy difference between the global minima in vacuo and in aqueous solution is

Table 2. Calculated MM3* results for compounds **1**–**27** (energies are in kcal/mol)

No.	Global energy minimum for unbound ligand		Protein-bound ligand		Local energy minimum	
	In vacuo[a]	Aqueous	ΔE_{conf}[b]	Rms[c] (Å)	ΔE_{local}[d]	Rms[e] (Å)
1	0.0	0.0	0.0	0.04	0.0	0.04
2	0.0	0.0	0.0	0.14	0.0	0.14
3	0.0	0.0	1.4	0.14	0.0	0.14
4a	0.0	0.0	0.0	0.24	0.0	0.27
4b	−0.1	0.0	0.0	0.07	0.1	0.07
4c	0.0	0.0	0.1	0.23	0.1	0.30
5	0.0	0.0	0.3	0.17	0.3	0.35
6	0.0	0.0	0.0	0.17	0.0	0.17
7	0.1	0.0	0.7	0.06	0.7	0.06
8	0.0	0.0	4.5[f]	0.33	0.9	0.50
9	−0.1	0.0	1.9	0.10	1.3	0.09
10	−0.1	0.0	−0.1	0.08	0.0	0.09
11	−0.1	0.0	1.6	0.22	1.7	0.64
12	−0.1	0.0	−0.1	0.22	0.0	0.24
13	−2.0	0.0	0.9	0.24	2.9	0.60
14	−0.9	0.0	3.4[f]	0.32	1.5	0.49
15	−0.2	0.0	2.7[f]	0.28	2.9	0.72
16a	−1.9	0.0	−1.6	0.22	0.3	0.33
16b	−1.9	0.0	0.2	0.21	1.0	0.83
17	−0.7	0.0	0.5	0.21	0.3	0.34
18	−13.4	0.0	−0.6	0.27	1.9	0.47
19	−1.4	0.0	3.4[f]	0.30	3.4	0.62
20a	−1.4	0.0	6.1[f]	0.31	4.4	1.31
20b	−1.4	0.0	2.5[f]	0.28	0.6	0.51
21a	−1.4	0.0	7.4[f]	0.28	4.8	0.82
21b	−1.4	0.0	0.4[f]	0.28	0.6	0.51
22a	−1.7	0.0	5.3[f]	0.28	3.1	0.76
22b	−1.7	0.0	4.3[f]	0.33	2.6	2.8
23	−1.8	0.0	5.5[f]	0.32	3.6	2.12
24	−0.8	0.0	7.2[f]	0.23	4.9	1.48
25	−4.6	0.0	5.2	0.21	6.5	0.84
26	−4.0	0.0	6.5[f]	0.29	5.0	0.89
27	−32.3	0.0	−2.0	0.23	22.0	1.08

[a] Energies relative to the global energy minimum in aqueous solution.
[b] Calculated by using flat-bottomed Cartesian constraints. The applied half-width was 0.3 Å, unless otherwise indicated. Energies are relative to the global energy minimum in aqueous solution.
[c] The rms deviation between the structure obtained by constrained optimization and the X-ray structure.
[d] The energy difference between the protein-bound conformation and the closest local energy minimum.
[e] The rms deviation between the X-ray structure and its closest local energy minimum structure.
[f] A flat-bottomed half-width of 0.4 Å was applied.

Table 3. Calculated AMBER* results for compounds **1**–**27** (energies are in kcal/mol)

No.	Global energy minimum for unbound ligand		Protein-bound ligand		Local energy minimum	
	In vacuo[a]	Aqueous	ΔE_{conf}[b]	Rms[c] (Å)	ΔE_{local}[d]	Rms[e] (Å)
1	0.0	0.0	0.0	0.06	0.0	0.06
2	0.0	0.0	0.0	0.17	0.0	0.14
3	0.0	0.0	0.3	0.23	0.0	0.22
4a	−0.1	0.0	0.0	0.24	0.1	0.32
4b	0.0	0.0	0.0	0.01	0.0	0.07
4c	−0.1	0.0	0.2	0.22	0.3	0.35
5	−0.5	0.0	0.6	0.20	0.5	0.38
6	−0.7	0.0	−0.7	0.17	0.0	0.17
7	−0.1	0.0	−0.1	0.09	0.0	0.09
8	−0.1	0.0	3.6[f]	0.32	0.9	0.53
9	0.0	0.0	2.7	0.10	1.2	0.10
10	−0.4	0.0	−0.2	0.09	0.0	0.09
11	−0.2	0.0	−0.2	0.10	0.0	0.10
12	−0.2	0.0	−0.2	0.23	0.0	0.23
13	−0.9	0.0	1.6	0.23	2.5	0.62
14	−2.4	0.0	0.4[f]	0.30	0.6	0.45
15	−2.7	0.0	2.0[f]	0.26	2.4	0.68
16a	−0.6	0.0	0.2	0.22	0.3	0.30
16b	−0.6	0.0	3.0	0.24	0.7	0.48
17	−0.5	0.0	1.1	0.23	1.1	0.31
18	−4.9	0.0	1.5	0.23	0.7	0.44
19	−0.1	0.0	3.8[f]	0.30	2.9	0.60
20a	−0.1	0.0	6.1[f]	0.31	4.1	1.29
20b	−0.1	0.0	2.3[f]	0.30	0.4	0.52
21a	−0.1	0.0	6.2[f]	0.25	4.7	0.88
21b	−0.1	0.0	1.8[f]	0.30	0.4	0.52
22a	−0.9	0.0	4.3[f]	0.26	0.2	0.88
22b	−0.9	0.0	3.0[f]	0.32	0.3	0.89
23	−1.2	0.0	5.4[f]	0.32	3.2	2.28
24	−1.1	0.0	9.1[f]	0.28	5.6	1.38
25	−2.3	0.0	5.9	0.20	6.8	0.91
26	−2.0	0.0	7.1[f]	0.29	9.1	0.96
27	−7.1	0.0	19.1	0.20	0.0	0.20

[a–f] As in Table 2.

in general calculated to be less than 1 kcal/mol (Tables 2 and 3). The somewhat larger difference in energy between the preferred conformations in vacuo and in aqueous solution for the monocarboxylate substituted compounds **16** and **19**–**23** displayed by the MM3* calculations (Table 2) is mainly due to differences in the O-C-O bond angle in the two phases.

As mentioned above, molecules with polar atoms or charged groups which can form intramolecular hydrogen bonds or cation–π interactions will in general 'collapse' electrostatically in vacuo. In such cases it is highly probable that the most stable conformation in vacuo will be a conformation in which intramolecular electrostatic interactions and hydrogen bonding are optimized. In aqueous solution, hydrogen bonding between the ligand and the surrounding water molecules competes with the intramolecular electrostatic interactions and the preferred conformation may then adopt a more extended shape. Examples of this conformational behavior are displayed by compounds **13**–**15**, **18** and the amino acids **26** and **27**. The AMBER* as well as the MM3* results display coiled lowest

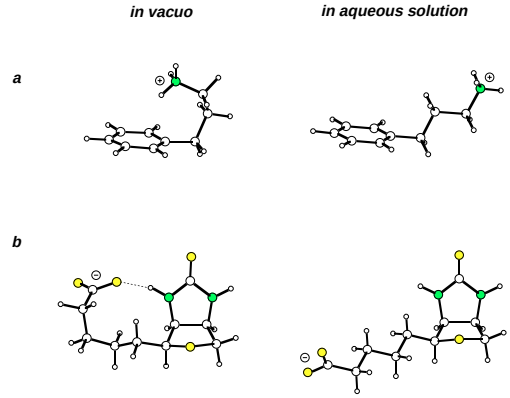

Figure 4. Calculated global energy minimum conformations (MM3*) for (a) compound **14** and (b) compound **18** in vacuo and in aqueous solution.

energy conformations in vacuo for these compounds (with the exception of compound **15** in MM3*). Figure 4a displays the preferred MM3* conformations of **14** in vacuo and in aqueous solution. The positively charged ammonium group is in vacuo attracted by the negatively charged carbon atoms of the phenyl ring. However, in water the hydration of the molecule, in particular the strong hydration of the ammonium group, favors an extended alkylammonium chain.

For biotin (**18**), the calculated preferred in vacuo conformation displays a strong hydrogen bond between the carboxylate anion and an NH of the ureido group (Figure 4b). However, in water an extended structure is strongly preferred. This solution structure is not significantly populated in vacuo and the gas-phase structure is very little populated in aqueous solution. The preferred gas-phase structures calculated for **18** by using AMBER* and MM3* are virtually identical. However, the preferred bicyclic conformation in aqueous solution differs between the two force fields and the energies of the preferred structures in relation to those in vacuo are accordingly substantially different (see Tables 2 and 3). The preferred conformation in aqueous solution obtained by AMBER* calculations has a bicyclic trans conformation, whereas the MM3* results display a cis conformation. Ab initio calculations (HF/6-31G*) predict essentially identical energies for the two conformations of the bicyclic ring ($\Delta E = 0.2$ kcal/mol) in agreement with the results obtained with the MM3* force field ($\Delta E = 0.4$ kcal/mol). However, AMBER* gives an energy difference of 2.5 kcal/mol favoring the trans conformation.

The largest force field dependence in the calculated results of the compounds studied in the present work is displayed by the amino acid ornithine (**27**). The results of the conformational analysis in vacuo and in aqueous solution are shown in Figure 5. As expected, both force fields predict a coiled ('electrostatically collapsed') conformation of **27** in vacuo. In aqueous solution, the MM3* force field calculates an extended structure similar to the protein-bound conformation, whereas AMBER* predicts a coiled structure also for the hydrated molecule. It is not clear which of the predictions are correct, but the results indicate that reliable force field calculations on this type of highly charged molecules may be beyond the capability of standard force fields and dielectric continuum hydration models. Most probably, polarization effects need to be included in the force field [16]. Furthermore, the dielectric continuum model only produces an 'average' solvation field. Particularly strongly bound water molecules, as may be expected in the hydration of highly charged molecules, may be of importance in the calculation of conformational properties of polyionic compounds such as **27**. The inclusion of explicit water molecules may be necessary in such cases.

Conformational energy penalties of the protein-bound ligands

The conformational energy differences (ΔE_{conf}) between the protein-bound ligand and its global energy minimum in aqueous solution are given in Tables 2 and 3. These results are also shown in graphical form in Figure 6.

The two force fields applied in the present study display similar results for the calculated conformational energy penalties (ΔE_{conf}) for the protein-bound ligands. Excluding the strongly deviating case of compound **27** discussed above, the average difference in the calculated ΔE_{conf} between the two force fields is 0.9 kcal/mol.

For 22 (MM3*) or 23 (AMBER*) of the ligand–protein complexes studied in the present work, the conformational energy penalty of the bound ligand is calculated to be less than or equal to 3.0 kcal/mol. In the following, we will discuss the results of the calculations in terms of two energy intervals. First we will discuss the protein-bound ligands with calculated conformational energy penalties of $\leq$3.0 kcal/mol and then ligands with higher conformational energies for the bioactive conformation. For 30 of the 33 protein–

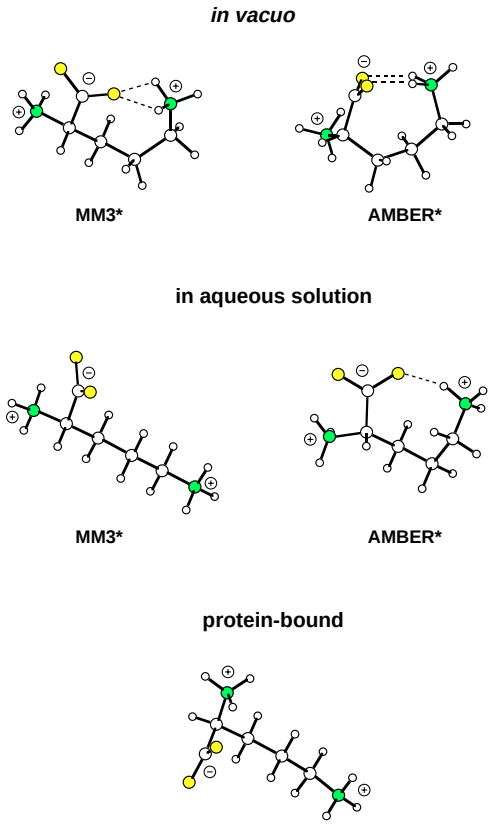

Figure 5. Calculated preferred conformations of ornithine (**27**) in vacuo and in aqueous solution calculated by MM3* and AMBER* and the conformation of **27** bound to the Lys, Arg, Orn-binding protein (LAO).

ligand complexes the calculated ΔE_{conf} is in the same energy interval for both force fields employed.

Ligands with a calculated conformational energy penalty ≤ 3.0 kcal/mol

In this group of ligands, the bioactive conformation is often identical to, or differs only slightly from, the global minimum conformation in aqueous solution. For example, for **1** and **2** the preferred conformations are, for both force fields employed, identical to the bioactive conformation. In these cases, the preferred conformations *in vacuo* and in solution are also identical. The ligands **1**–**3** are all low affinity ligands bound to lysozyme mutants. The dissociation constants range from 14 to 68 μm [17].

When employing the lowest energy conformation in aqueous solution for the calculation of the conformational energy penalty for the protein-bound ligand, the energy penalty may very well be negative. This happens when the lowest energy conformation in aqueous solution is higher in energy than the preferred conformation in vacuo and the bioactive conformation is closer to the in vacuo structure than to the solution structure. An example of this is 1,2- dichloroethane (**6**) in the AMBER* calculations (Table 3 and Figure 6). The calculated preferred conformation in vacuo is anti, whereas it is gauche in aqueous solution. The conformation of **6** when bound to the haloalkane dehalogenase enzyme is anti as calculated for the gas-phase structure.

The preferred conformations in vacuo and in aqueous solution for compound **9** are identical, whereas the protein-bound conformation is somewhat different. The conformational energy penalty for binding is calculated to be less than 3 kcal/mol with both force fields. The energy penalty is mainly due to the adoption of a higher energy hydroxyl torsion in the bioactive conformation of **9** to make it possible for the hydroxyl group to act as a hydrogen bond donor to a serine residue and as an acceptor to a histidine residue in acetylcholine esterase. The bioactive conformation is in this case not in a local energy minimum ($\Delta E_{local} > 0$, Tables 2 and 3).

Biotin (**18)** is one of two compounds in the present study which Nicklaus et al. [8] also included in their study. The conformational energy penalty of binding for **18**, calculated by Nicklaus et al., is 8.2 kcal/mol (calculated for the un-ionized state). As mentioned above, this value is the difference between the energy of the protein-bound structure and the preferred conformation in vacuo. The high conformational energy penalty is due to the 'collapse' of **18** in vacuo as shown in Figure 4b. The corresponding energy calculated by using MM3* and AMBER* for the ionized state is 12.8 and 6.4 kcal/mol, respectively. These high conformational energies of the bioactive conformation are clearly not compatible with the exceptionally high affinity of **18** to streptavidin ($K_a = 2.5 \times 10^{13}$, $\Delta G° = -18.3$ kcal/mol, $\Delta H° = -32.0$ kcal/mol [18]). When applying the GB/SA dielectric continuum hydration model, the global energy minimum conformations of both force fields display an extended structure (Figure 4b). The energy difference between the protein-bound conformation and this extended conformation of **18** is small, −0.6 and 1.5 kcal/mol for MM3* and AMBER*, respectively (Tables 2 and 3). This example clearly demonstrates the importance of using the aqueous conformational ensemble (here represented by the lowest energy conformer) if conformational energies

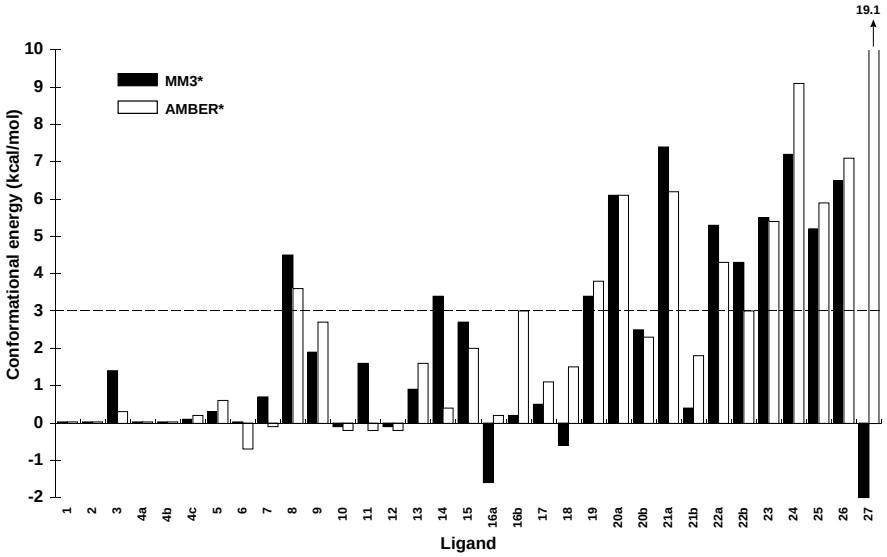

Figure 6. Calculated conformational energy penalties of the protein-bound ligands **1**–**27**. The dashed horizontal line indicates a conformational energy penalty of 3 kcal/mol. In the cases for which the conformational energy differences are calculated to be 0.0 kcal/mol, for clarity they are represented by a bar corresponding to 0.1 kcal/mol.

of polar protein-bound ligands are to be discussed in the context of experimentally determined affinities and for the prediction of the conformational contribution to the binding energy.

Ligands with a calculated conformational energy penalty >3.0 kcal/mol

The results of the present study display nine ligand–protein complexes for which the conformational energy penalties for ligand binding are calculated to be higher than 3 kcal/mol with MM3* as well as with AMBER* (Tables 2 and 3, Figure 6). In eight of these cases, the ligand is a long-chain very flexible molecule as **8** or the fatty acids **19**–**23** or an amino acid (**25**, **26**). In addition, both force fields calculate the conformational energy required for compound **24** to adopt its bioactive conformation to be significantly larger than 3 kcal/mol. In this context it is also of interest to discuss the very high conformational energy penalty (19.1 kcal/mol) calculated for **27** by AMBER*.

Palmitic acid (**20**) binds to the rat intestinal fatty acid binding protein (**20a**) and to the mouse adipocyte lipid binding protein (**20b**) with high and similar affinities, $K_i = 30$ and 77 nM, respectively [19]. However, whereas the calculated conformational energy penalty is low for the ligand in complex **20b** (MM3* 2.5 kcal/mol, AMBER* 2.3 kcal/mol), it is significantly higher (MM3* and AMBER* 6.1 kcal/mol) in complex **20a**. The isotropic temperature factors for **20a** increase from about 20 close to the carboxylate group to more than 50 at the terminal carbon atoms. This suggests that there may be other conformations than the reported X-ray structure of **20** in complex **20a** which may fit the electron density envelope and which may display lower conformational energy penalties. To investigate this, we searched the calculated aqueous conformational ensemble for **20**. The results are shown in Figure 7, which displays a least-squares superimposition of the X-ray structure of **20** in complex **20a** and five conformations with calculated conformational energy penalties of less than 2 kcal/mol. The five structures are very similar to the X-ray structure and they most probably fit well into the experimental electron density envelope. These structures are not obtainable by constrained structure optimization of the experimental structure. Similarly, we found three conformations of decamethonium (**8**) with rms deviations with respect to the X-ray structure of 0.6–0.8 Å and with calculated conformational energy penalties of less than 3 kcal/mol.

These results indicate that in these cases the resolution of the experimental data may not be sufficient to unambiguously determine the conformation of the long-chain flexible ligands, and that the high conformational energy penalties calculated for the majority of such compounds in the present study do not represent the real situation.

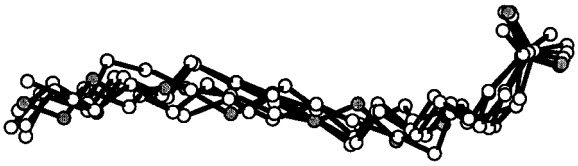

Figure 7. A least-squares superimposition of the X-ray structure of **20a** as bound to the rat intestinal fatty acid binding protein (filled atoms) and five conformations with calculated energy penalties of $\leq$2 kcal/mol (unfilled carbon atoms). The rms deviations are in the range 0.5–0.8 Å.

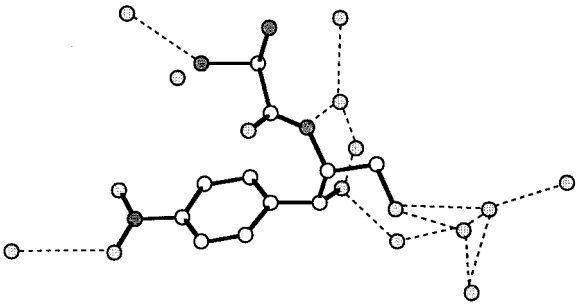

Figure 8. The protein-bound structure of **24** and 12 surrounding water molecules in the active site. The dashed lines indicate plausible hydrogen bonds.

The calculated conformational energy penalty for chloramphenicol (**24**) bound to its acetyltransferase is 7.2 kcal/mol with MM3* and 9.1 kcal/mol with AMBER*. In spite of the low affinity of **24** ($K_d = 13\ \mu M$ [20]), these high energies are surprising. A large part of these energies comes from repulsive electrostatic interactions between the hydroxyl groups in the protein-bound structure. An inspection of the ligand–protein complex reveals that the hydroxyl groups of the ligand do not interact directly with an amino acid residue or backbone atom of the protein. The ligand is surrounded by a large number of water molecules as shown in Figure 8. This makes it difficult to determine the directions of the OH bonds of the ligand. Furthermore, the water molecules which are hydrogen bonded to the ligand, in particular those hydrogen bonded to the hydroxyl groups, may be considered to be part of the ligand. The electrostatic interactions calculated between the hydroxyl groups may thus be highly exaggerated in our calculations.

The three amino acids included in the present study (**25**–**27**) all display high conformational energies for the protein-bound structure (Tables 2 and 3; Figure 6). For **25**, the diphenyl ether part of the ligand is virtually identical in aqueous solution and in the protein-bound structure. Thus, this part of the ligand does not give any contribution to the high conformational energy penalty (MM3* 5.2 kcal/mol, AMBER* 5.9 kcal/mol); the entire conformational energy penalty comes from interactions within the amino acid part and interactions between the amino acid part and the nearby aromatic system. However, the amino acid part of the ligand in the ligand–protein complex is fully exposed to bulk water and should be completely hydrated. The electrostatic interactions calculated for the fully desolvated molecule are thus most probably highly exaggerated. If this is taken into account, it is highly probable that ligand **25** has a low conformational energy penalty for binding.

The calculated results for histidine (**26**) and ornithine (**27**) demonstrate serious problems in calculations of conformational properties of highly polar molecules within the framework of force field calculations and a dielectric continuum solvation model. This is most obvious, as discussed above, for ligand **27** which displays very different calculated results by using MM3* and AMBER* (Figure 5). The preferred 'electrostatically collapsed' conformation in aqueous solution calculated by AMBER* results in a very high calculated conformational energy penalty for **27**, 19.1 kcal/mol (Table 3). This high energy is clearly unrealistic. The situation for **26** is similar; a hydrogen bond between the non-protonated imidazole nitrogen and the ammonium group is predicted for the molecule in vacuo as well as in aqueous solution by both force fields, whereas the protein-bound conformation does not display any intramolecular hydrogen bonding.

Conclusions

In the great majority of ligand–protein complexes studied in this work (22 and 23 out of 33 complexes for MM3* and AMBER*, respectively), the conformational energies required for the ligands to adopt their bioactive conformations are calculated to be less than or equal to 3 kcal/mol. It should be noted that the aqueous conformational ensemble for the unbound ligand must be used for this type of calculations. The results of the calculations on ligand–protein complexes for which the ligand is calculated to have a high conformational energy penalty in all cases suffer from uncertainties in the interpretation of the experimental data or inadequacies of the computational methods. For long-chain flexible ligands, e.g. fatty acids, with calculated conformational energy penalties significantly higher than 3 kcal/mol, the resolution of the experimental

data is most probably not sufficient to unambiguously determine the protein-bound conformation. It is demonstrated that several conformations may be found which are similar to the experimental one and which display low conformational energy penalties for binding. For strongly polar molecules, e.g. amino acids, the results obtained in this study indicate that further developments of the force field methodology and of dielectric continuum solvation models are required for reliable calculations on the conformational properties of this type of compounds.

Acknowledgements

This work was financially supported by the Danish Medical Research Council and the Lundbeck Foundation, Copenhagen, which is gratefully acknowledged.

References

1. Siebel, G.L. and Kollman, P.A., In Hansch, C., Sammes, P.G., Taylor, B. and Ramsden, C.A. (Eds.) Comprehensive Medicinal Chemistry, Vol. 4, Pergamon, Oxford, 1990, pp. 125–138.
2. Liljefors, T. and Pettersson, I., In Krogsgaard-Larsen, P., Liljefors, T. and Madsen, U. (Eds.) A Textbook of Drug Design and Development, 2nd ed., Overseas Publishers Association, Amsterdam, 1996, pp. 60–93.
3. Pettersson, I. and Liljefors, T., J. Comput.-Aided Mol. Design, 1 (1987) 143.
4. Bengtsson, M., Liljefors, T. and Hansson, B.S., Bioorg. Chem., 15 (1987) 409.
5. Cramer III, R.D., DePriest, S.A., Patterson, D.E. and Hecht, P., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 443–485.
6. Marshall, G.R. and Motoc, I., In Burgen, A.S.V., Roberts, G.C.K. and Tute, M.S. (Eds.) Molecular Graphics and Drug Design, Elsevier, New York, NY, 1986, pp. 115–156.
7. Protein Data Bank, Brookhaven National Laboratory, http://www.pdb.bnl.gov.
8. Nicklaus, M.C., Wang, S., Driscoll, J.S. and Milne, G.W.A., Bioorg. Med. Chem., 3 (1995) 411.
9. Spark, M.J., Winkler, D.A. and Andrews, P.R., Int. J. Quantum Chem., 9 (1982) 321.
10. Poulos, T. and Howard, A.J., Biochemistry, 26 (1987) 8165.
11. MacroModel v. 5.0, Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T. and Still, W.C., J. Comput. Chem., 11 (1990) 440.
12. INSIGHTII (v. 95.0), Molecular Simulations, Cambridge, U.K.
13. Gundertofte, K., Liljefors, T., Norrby, P.-O. and Pettersson, I., J. Comput. Chem., 17 (1996) 429.
14. Chang, G., Guida, W.C. and Still, W.C., J. Am. Chem. Soc., 111 (1989) 4379.
15. Still, W.C., Tempczyk, A., Hawley, R.C. and Hendrickson, T., J. Am. Chem. Soc., 112 (1990) 6127.
16. a. Howard, A.E., Singh, U.C., Billeter, M. and Kollman, P.A., J. Am. Chem. Soc., 110 (1988) 6984.
    b. Meng, E.C., Cieplak, P., Caldwell, J.W. and Kollman, P.A., J. Am. Chem. Soc., 116 (1994) 12061.
17. Morton, A., Baase, W.A. and Matthews, B.W., Biochemistry, 34 (1995) 8564.
18. a. Green, N.M., Methods Enzymol., 184 (1990) 51.
    b. Weber, P.C., Wendoloski, J.J., Pantoliano, M.W. and Salemme, F.R., J. Am. Chem. Soc., 114 (1992) 3197.
19. Richieri, G.V., Ogata, R.T. and Kleinfeld, A.M., J. Biol. Chem., 39 (1994) 23918.
20. Leslie, A.G.W., J. Mol. Biol., 213 (1990) 167.