Conformational Analysis of Drug−Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding

ARTICLE *in* JOURNAL OF MEDICINAL CHEMISTRY · JUNE 2004

Impact Factor: 5.45 · DOI: 10.1021/jm030563w · Source: PubMed

2 AUTHORS, INCLUDING:

Paul S Charifson
Vertex Pharmaceuticals
58 PUBLICATIONS **3,104** CITATIONS

SEE PROFILE

Conformational Analysis of Drug-Like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization upon Binding

Emanuele Perola* and Paul S. Charifson

Vertex Pharmaceuticals, 130 Waverly Street, Cambridge, Massachusetts 02139

Received November 6, 2003

This paper describes a large-scale study on the nature and the energetics of the conformational changes drug-like molecules experience upon binding. Ligand strain energies and conformational reorganization were analyzed with different computational methods on 150 crystal structures of pharmaceutically relevant protein−ligand complexes. The common knowledge that ligands rarely bind in their lowest calculated energy conformation was confirmed. Additionally, we found that over 60% of the ligands do not bind in a local minimum conformation. While approximately 60% of the ligands were calculated to bind with strain energies lower than 5 kcal/mol, strain energies over 9 kcal/mol were calculated in at least 10% of the cases regardless of the method used. A clear correlation was found between acceptable strain energy and ligand flexibility, while there was no correlation between strain energy and binding affinity, thus indicating that expensive conformational rearrangements can be tolerated in some cases without overly penalizing the tightness of binding. On the basis of the trends observed, thresholds for the acceptable strain energies of bioactive conformations were defined with consideration of the impact of ligand flexibility. An analysis of the degree of folding of the bound ligands confirmed the general tendency of small molecules to bind in an extended conformation. The results suggest that the unfolding of hydrophobic ligands during binding, which exposes hydrophobic surfaces to contact with protein residues, could be one of the factors accounting for high reorganization energies. Finally, different methods for conformational analysis were evaluated, and guidelines were defined to maximize the prevalence of bioactive conformations in computationally generated ensembles.

Introduction

The binding of small organic molecules to proteins can be investigated with a number of computational methods, including docking, pharmacophore search, and 3D-QSAR. In each of these methods multiple conformations of each prospective ligand are evaluated, and their complementarity to the binding site of a given protein is estimated using specific metrics. To successfully apply these methods, it is important to have reliable criteria to evaluate the biological relevance of a given conformation. To establish such criteria is the primary objective of this study.

It is intuitive that a good ligand must bind in a relatively low energy conformation, since the internal energy of the ligand contributes to the total free energy of binding. However, it has been observed that in the majority of cases the bioactive conformation of a flexible ligand does not correspond to the global energy minimum of the same ligand in its free state, and in many cases it does not even correspond to a local minimum.[1,2] This observation is consistent with the induced fit theory,[3−7] according to which both ligand and protein can reorganize themselves in order to adopt the complementary shapes required for binding. When this happens, the interactions between the two partners compensate for the cost in reorganization energy that ligand and protein have to pay in the process.

In order to establish what is the amount of ligand strain energy that can be tolerated in this kind of process, a rigorous statistical study of the reorganization energies associated to ligand−protein binding would be required. However, the contribution of the internal strain of the ligand to the free energy of binding cannot be directly measured, and the amount of structural data available for small molecules in solution (the initial state of the process) is very limited, thereby precluding a rigorous determination of the appropriate energy thresholds. When the structure of the bound ligand is available, the strain energy can be calculated with computational methods, which rely on force fields or quantum mechanics to determine the preferred conformation of the unbound ligand and to calculate the energies of bound and unbound conformations. From the practical standpoint, a systematic computational study on the strain energies of ligands in known crystal structures would help define the appropriate energy thresholds to be used to filter conformational ensembles and docking poses. Such thresholds would be applicable as long as the conformational energies are calculated with the same approach used to derive the rules.

Another important aspect to be considered in this context is the degree of folding/unfolding a ligand undergoes during binding. Visual analysis of a large number of protein−ligand complexes leads to the qualitative conclusion that ligands tend to bind in an extended conformation, even when a folded conformation is more stable in solution. A systematic analysis of the conformations of bound ligands would allow to quantitate such tendency, and it could potentially add

* To whom correspondence should be addressed. Phone: (617) 444-6646, fax: (617) 444-6566, e-mail: emanuele_perola@vrtx.com.

10.1021/jm030563w CCC: $27.50
Published on Web 04/07/2004

one additional criterion for the assessment of the biological relevance of a given conformation.

While the issue of ligand unfolding has never been extensively treated, three studies on the conformational strain of bound ligands were reported a few years ago. In the first,[1] a set of 27 protein−ligand complexes was selected from the PDB, and the strain energies of the bound ligands were calculated relative to their lowest energy conformations in vacuum using the CHARMM force field. Ionizable molecules were treated as neutral to minimize the artifacts due to electrostatic collapse of charged groups in vacuum. The calculated strain energies ranged from 0 to 39.7 kcal/mol, with an average of 15.9 kcal/mol. The strain energy appeared to be proportional to the number of rotatable bonds and, to a lower extent, to the number of hydrogen bond centers in the ligand. The authors concluded that a wide ensemble of conformations should be used in order to be confident that the bioactive conformation is included, but no energy thresholds were suggested. In the second study,[2] 33 protein−ligand complexes were selected from the PDB, and conformational analyses were performed for each ligand in continuum solvent using both the MM3 and the AMBER force fields, thus attempting to determine the preferred conformations in aqueous solution. The conformational strain energies were calculated as the energy difference between the experimental bound conformation and the lowest energy conformation of each ligand in water. Although 30% of the compounds had calculated strain energies higher than 3 kcal/mol, the authors attributed such cases to inadequacies in the computational methods or insufficient resolution of the corresponding crystal structures, thereby suggesting that 3 kcal/mol could be used as a general strain energy threshold. In the third study,[8] performed on a set of 10 complexes, the same issue was addressed with a slightly different approach. Systematic conformational search was performed on the 10 ligands, the conformations were clustered, and the solution free energies of representative conformers were calculated using a continuum solvent model. The low energy conformers thus generated were compared to the crystallographic conformers with respect to torsion angles and positions of anchoring points, the latter defined as the key atoms responsible for tight binding. The authors found little similarity between the low energy solution conformers and the crystallographic conformers in torsional space, and the calculated strain energies for the latter ranged from 0 to 9 kcal/mol. However, the positions of the anchoring points in the lowest energy solution structures were found to be very similar to the positions of the same atoms in the active site conformations in 9 out of 10 cases. On this basis, the authors concluded that building pharmacophore models based on low energy solution conformations was still a valid approach.

Although these studies provide useful insights and suggest reasonable approaches to address this topic, their usefulness is limited by the size and composition of the test sets employed. The number of complexes included in each study is relatively small, and the vast majority of the ligands in the complexes are not drug-like. Moreover, only in the second study is a quantitative threshold for conformational strain proposed, but admittedly only 70% of an already small data set supports such a conclusion.

This analysis highlights the need for new studies addressing the reorganization of ligands upon binding on larger test sets enriched in drug-like ligands. To address the conformational strain and folding issues in a rigorous manner, we performed a systematic study on a set of 150 crystal structures of pharmaceutically relevant complexes with known binding affinities. We calculated the strain energies of the crystallographic conformer of each ligand relative to the global minimum and to the closest local minimum for the same ligand in its unbound state. Calculations were performed with two different and widely used force fields, MMFF[9−11] and OPLS-AA,[12] both in vacuum and in continuum solvent, thus addressing the dependence of the results on the method used. We also analyzed the degree of folding of the bound conformations of the ligands using two different metrics. Finally, we evaluated three reputable programs for fast conformational analysis on the same test set of ligands.

The objectives of this work were to (a) define reliable strain energy thresholds to be used when filtering ensembles of conformers or docking poses, (b) estimate the tendency of ligands to unfold upon binding, (c) analyze the main causes of conformational strain in bound ligands, and (d) define a protocol for efficient and reliable conformational analysis.

We have also addressed methodological aspects of conformational analysis, with particular attention to the advantages and disadvantages of the application of continuum solvent models.

Methods

Selection of Protein−Ligand Complexes. Test sets of complexes have been described in the literature in the context of various computational studies.[13−15] A common denominator of all these sets is the low occurrence of drug-like ligands, which in our opinion limits the usefulness of such sets for applications or studies that are pertinent to drug discovery. To address this issue, we carefully constructed a test set of pharmaceutically relevant complexes suitable for a variety of tasks: evaluation of docking programs and existing scoring functions, development and calibration of new scoring functions, and analysis of various aspects of protein−ligand binding and conformational analysis.

One hundred complexes were selected from the Protein Data Bank[16] and 50 from the Vertex structure collection according to the following criteria:

General: (a) binding constant (K_i or K_d) available; (b) noncovalent binding between ligand and protein; (c) crystallographic resolution < 3.0 Å.

Ligands: (a) molecular weight between 200 and 600; (b) 1 to 12 rotatable bonds; (c) drug-/lead-like; (d) structurally diverse.

Proteins: (a) multiple classes; (b) diverse within classes; (c) relevant to drug discovery.

Complexes containing ligand classes that are of limited interest to drug discovery programs, such as sugars, nucleotides, and macrocycles, were excluded, as well as ligands containing atoms other than C, O, N, S, F, Cl, Br, and H. A small number of peptide-containing structures were included. No ligand was selected twice from different complexes. These criteria reflect our intention to include the maximum amount of structural information on systems that are of high interest in a structure-based drug design context and exclude those that are only rarely considered. Additional criteria were used to exclude structures that would not be suitable for docking or scoring studies. Details on such criteria will be given in a separate paper.

The test set includes 63 different proteins from a variety of classes, including proteases, kinases, nuclear receptors, phosphatases, oxido-reductases, isomerases, and lyases. Kinases (43 complexes) and proteases (42 complexes) are the most widely represented. There are 24 metalloprotein complexes, all of them with a zinc ion in the active site. Several examples of approved drugs in complex with their targets are also included (e.g., Agenerase/HIV protease, Aricept/acetylcholinesterase, Lisinopril/ACE). The PDB codes of the complexes selected from the Protein Data Bank are reported in Table 1, along with crystallographic resolutions, protein names and sources and binding affinities, expressed as pK_i ($-\log_{10} K_i$).

Search for the Global Minima. To identify the lowest energy conformation (global minimum) of each ligand in the test set, conformational analysis was performed on the 150 ligands with three different programs. In this part of the study we attempted to achieve a high degree of thoroughness, and we were less concerned with the speed of the methods used. The ligand structures were extracted from the pdb files of the complexes, and correct bond orders and protonation states were assigned upon visual inspection. The structures thus obtained were used as input for the conformational analyses, performed with each program using the settings summarized below.

Macromodel v8.0 (Schrodinger Inc.). The low-mode conformational search method[17] was used, and the calculations were performed with the OPLS-AA force field using the GB/SA continuum solvent model for water.[18] Structures with energy over 50 kJ/mol higher than the current global minimum were discarded, and each search was carried out until 5000 conformations were found. Each conformation was minimized for up to 500 steps using the conjugate gradient method. Redundant conformations were removed after heavy atom superimposition (RMSD cutoff = 0.5 Å).

Catalyst v4.6 (Accelrys). The conformational analyses were performed in the BEST mode, which uses the poling search algorithm[19] to maximize diversity and a modified version of the CHARMM force field[20] with no electrostatic terms for the energy calculations. Each conformation is minimized in both Cartesian and torsional space. The default energy threshold of 20 kcal/mol above the global minimum was used in these calculations, and the maximum number of conformations was set to 1000.

ICM v3.0 (Molsoft LLC). The calculations were performed using the Monte Carlo search algorithm implemented in the program.[21] Each conformation is minimized in torsional space with the ICM force field, which combines parameters from ECEPP/3[22] and MMFF. Default settings were used for the length of the Monte Carlo runs. An energy threshold of 20 kcal/mol above the global minimum was used, and the maximum number of conformations was set to 500. Duplicate conformations were discarded based on a torsion angle cutoff of 30°.

All the conformations generated with the three methods were combined and energy-minimized with Macromodel both in vacuum and in continuum solvent. The minimizations were performed with two different force fields, OPLS-AA and MMFF, thus generating four sets of minimized conformers (OPLS-AA/vacuum, OPLS-AA/water, MMFF/vacuum and MMFF/water). In the calculations in vacuum a distance-dependent dielectric constant ($4r_{ij}$) was used. In the calculations in continuum solvent the GB/SA continuum solvent model for water was used, and the dielectric constant was set to 1. All the calculations were carried out with the truncated Newton conjugate gradient method to the final convergence (0.01 kJ/Å/mol). Four sets of global minima (one for each force field/medium combination) were identified upon analysis of the final calculated conformational energies.

Calculation of Strain Energies. The strain energies of the crystallographic conformations of the ligands were calculated with MMFF and OPLS-AA both in vacuum and in continuum solvent, using the same force field settings listed above for each case. The global strain energy of a protein-bound ligand can be defined as the energy difference between its bound (bioactive) conformation and the lowest energy conformation of the unbound ligand. The local strain energy of a protein-bound ligand can be defined as the energy difference between its bound (bioactive) conformation and the conformation corresponding to the closest local minimum on the conformational energy surface of the unbound ligand. While the crystallographic conformation of a bound ligand can be considered a valid approximation of its bioactive conformation, the crystallographically determined positions of ligand atoms in a protein−ligand complex contain significant uncertainties, and the bond lengths and angles in the crystal structure are ultimately assigned by the force field used for refinement. Different force fields have different optimal values for bond lengths in particular, and small adjustments of such lengths when a structure is relaxed in a different force field may result in an artificially large energy change, not reflective of the actual strain of the original conformation. In order to normalize bond lengths and angles with respect to the force fields used in this study, we performed restrained minimization of the crystallographic structures of the ligands in both force fields, using harmonic flat-bottomed Cartesian constraints to tether the non-hydrogen atoms to their original position (half-width of flat bottom restraint = 0.5 Å, force constant = 500 kcal/mol/ Å).[2] The minimizations were carried out to convergence (0.01 kJ/Å/mol). The normalized crystallographic conformations were then fully minimized until convergence (0.01 kJ/Å/mol) to the closest local minimum. The local strain energy of each crystallographic conformer was calculated as the energy difference between the partially minimized (normalized) and the fully minimized crystallographic conformation.

The global strain energy of each crystallographic conformer was calculated as the energy difference between the partially minimized crystallographic conformation and the lowest energy conformation of that ligand (the global minimum) determined as described above.

Calculation of the Degree of Folding. The degree of extension of the bioactive conformations of the ligands was calculated with two different methods. In the first method the geometric extent of a given conformation was defined as the largest pairwise interatomic distance in that conformation. This distance was determined for all the crystallographic conformers as well as for all the minimized conformers generated for each ligand. The degree of extension for a given conformer was then calculated as the ratio between the extent of that conformer and the maximum extent determined from all the minimized conformers of the same molecule.

In the second method the solvent-accessible surface areas (SASA) of crystallographic and minimized conformers were calculated, and the degree of extension of a given conformer was calculated as the ratio between the SASA of that conformer and the largest SASA determined from all the minimized conformers of the same molecule.

Evaluation of Tools for Fast Conformational Analysis. The test set described above was used in the evaluation of three programs: ICM, Catalyst, and Omega. The 2D structures of the crystallographic ligands were converted to 3D using Corina,[23] and the resulting conformations were used as input for the calculations. The settings of the calculations were chosen as described below to generate an average between 150 and 200 conformers/molecule with each of the three programs.

ICM v3.0 (Molsoft LLC). The parameter that controls the length of the Monte Carlo runs was set to 12. The maximum number of conformers was set to 1000, and conformers with energy over 20 kcal/mol above the corresponding global minimum were discarded. Duplicate conformations were discarded based on a torsion angle cutoff of 30 degrees of arc. A total of 24 425 conformations were generated for the 150 compounds (163 confs/mol average).

Catalyst v4.6 (Accelrys). The calculation was performed in the FAST mode, in which the conformers are not energy-minimized. The maximum number of conformers was set to 1000, and the energy window was set to 20 kcal/mol. Default

Table 1. Composition of the PDB Portion of the Test Set with Binding Data Expressed as pK_i

PDB code	resolution, Å	protein	organism	pK_i
13gs	1.90	glutathione S-transferase	human	4.62
1a42	2.25	carbonic anhydrase II	human	9.89
1a4k	2.40	antibody Fab	mouse	8.00
1a8t	2.55	metallo β-lactamase	*Bacteroides fragilis*	5.80
1afq	1.80	γ-chymotrypsin	bovine	6.21
1aoe	1.60	dihydrofolate reductase	*Candida albicans*	9.66
1atl	1.80	atrolysin C	*Crotalus atrox*	6.28
1azm	2.00	carbonic anhydrase I	human	6.14
1bnw	2.25	carbonic anhydrase II	human	9.08
1bqo	2.30	stromelysin I	human	7.74
1br6	2.30	ricin	*Ricinus communis*	3.22
1cet	2.05	lactate dehydrogenase	*Plasmodium falciparum*	2.89
1cim	2.10	carbonic anhydrase II	human	9.55
1d3p	2.10	thrombin	human	5.11
1d4p	2.07	thrombin	human	6.30
1d6v	2.00	oxy-Cope catalytic antibody germline precursor	human/mouse hybrid	6.17
1dib	2.70	methylenetetrahydrofolate dehydrogenase	human	7.74
1dlr	2.30	dihydrofolate reductase	human	9.18
1efy	2.20	poly(ADP-ribose) polymerase	chicken	8.22
1ela	1.80	elastase	pig pancreas	6.35
1etr	2.20	thrombin	bovine	7.41
1ett	2.50	thrombin	bovine	6.19
1eve	2.50	acetylcholinesterase	*Torpedo californica*	8.48
1exa	1.59	retinoic acid receptor γ-2	human	6.30
1ezq	2.20	coagulation factor Xa	human	9.05
1f0r	2.10	coagulation factor Xa	human	7.66
1f0t	1.80	trypsin	bovine	6.00
1f4e	1.90	thymidylate synthase	*Escherichia coli*	2.96
1f4f	2.00	thymidylate synthase	*Escherichia coli*	4.62
1f4 g	1.75	thymidylate synthase	*Escherichia coli*	6.48
1fcx	1.47	retinoic acid receptor γ-1	human	7.12
1fcz	1.38	retinoic acid receptor γ-1	human	9.22
1fjs	1.92	coagulation factor Xa	human	9.70
1fkg	2.00	FKBP-12	human	8.00
1fm6	2.10	PPAR-γ	human	7.33
1fm9	2.10	PPAR-γ	human	8.82
1frb	1.70	FR-1 aldo-keto reductase	mouse	7.77
1 g4o	1.96	carbonic anhydrase II	human	8.68
1gwx	2.50	PPAR-γ	human	7.30
1h1p	2.10	cyclin-dependent kinase II	human	4.92
1h1s	2.00	cyclin-dependent kinase II	human	8.22
1h9u	2.70	retinoid X receptor β	human	8.52
1hdq	2.30	carboxypeptidase A	bovine pancreas	5.82
1hfc	1.56	fibroblast collagenase	human	8.15
1hpv	1.90	HIV-1 protease	HIV-1	9.22
1htf	2.20	HIV-1 protease	HIV-1	8.09
1i7z	2.30	antibody Gnc92H2	human	6.40
1i8z	1.93	carbonic anhydrase II	human	9.82
1if7	1.98	carbonic anhydrase II	human	10.52
1iy7	2.00	carboxypeptidase A	bovine	6.19
1jsv	1.96	cyclin-dependent kinase II	human	5.70
1*k*1j	2.20	trypsin	bovine	7.68
1k22	1.93	thrombin	human	8.40
1k7e	2.30	tryptophan synthase	*Salmonella typhimurium*	2.92
1k7f	1.90	tryptophan synthase	*Salmonella typhimurium*	3.32
1kv1	2.50	p38 map kinase	human	5.94
1kv2	2.80	p38 map kinase	human	10.00
1l2s	1.94	β-lactamase	*Escherichia coli*	4.59
1l8g	2.50	protein tyrosine phosphatase 1B	human	6.22
1lqd	2.70	coagulation factor xa	human	8.05
1m48	1.95	interleukin-2	human	5.09
1mmb	2.10	MMP-8	human	9.22
1mnc	2.10	neutrophil collagenase	human	9.00
1mq5	2.10	coagulation factor Xa	human	9.00
1mq6	2.10	coagulation factor Xa	human	11.15
1nhu	2.00	HCV RNA polymerase	hepatitis C virus	5.66
1nhv	2.90	HCV RNA polymerase	hepatitis C virus	5.66
1o86	2.00	angiotensin converting enzyme	human	9.57
1ohr	2.10	HIV-1 protease	HIV-1	8.70
1ppc	1.80	trypsin	bovine	6.16
1pph	1.90	trypsin	bovine	6.22
1qbu	1.80	HIV-1 protease	HIV-1 virus	10.24
1qhi	1.90	thymidine kinase	herpes simplex virus	7.30
1ql9	2.30	trypsin	rat	5.36
1qpe	2.00	lymphocyte-specific kinase	human	8.40
1r09	2.90	human rhinovirus 14	human rhinovirus 14	4.90

Table 1. (Continued)

PDB code	resolution, Å	protein	organism	pK_i
1syn	2.00	thymidylate synthase	*Escherichia coli*	9.05
1thl	1.70	thermolysin	*Bacillus thermoprot.*	6.42
1uvs	2.80	thrombin	human	5.40
1uvt	2.50	thrombin	bovine	7.64
1ydr	2.20	c-AMP dependent kinase	bovine	5.52
1yds	2.20	c-AMP dependent kinase	bovine	5.92
1ydt	2.30	c-AMP dependent kinase	bovine	7.32
2cgr	2.20	immunoglobulin-2 β Fab fragment	mouse	7.27
2csn	2.50	casein kinase-1	*Schizosaccharomyces* p.	4.41
2pcp	2.20	antibody Fab	mouse	8.70
2qwi	2.00	influenza A neuraminidase	influenza A virus	8.40
3cpa	2.00	carboxypeptidase A	bovine	4.00
3erk	2.10	erk2 kinase	rat	5.12
3ert	1.90	estrogen receptor α	human	9.60
3std	1.65	scytalone dehydratase	*Magnaporthe grisea*	11.11
3tmn	1.70	thermolysin	*Bacillus thermoprot.*	5.90
4dfr	1.70	dihydrofolate reductase	*Escherichia coli*	8.62
4std	2.15	scytalone dehydratase	*Magnaporthe grisea*	10.33
5std	1.95	scytalone dehydratase	*Magnaporthe grisea*	10.49
5tln	2.30	thermolysin	*Bacillus thermoprot*	6.37
7dfr	2.50	dihydrofolate reductase	*Escherichia coli*	4.96
7est	1.80	elastase	pig pancreas	7.60
830c	1.60	collagenase-3	human	9.28
966c	1.90	fibroblast collagenase I	human	7.64

settings were used for the remaining parameters. A total of 25 962 conformations were generated (173 confs/mol average).

Omega v1.2 (OpenEye). The maximum number of conformers was set to 300, and the energy window was set to 20 kcal/mol. The threshold for duplicate removal was set to 0.75 Å RMSD. Default settings were used for the remaining parameters. A total of 25 500 conformations were generated (170 confs/mol average).

In each of the generated ensembles, the RMSD was calculated for each conformer both relative to the corresponding bioactive conformation (from the crystal structure) and to the corresponding global minimum. The ability of each program to find these two relevant conformations was thus determined. The global minima previously calculated in vacuum with MMFF were used as a reference in this case. The strain energy of each conformer relative to the closest local minimum was then calculated with MMFF, and the tendency of each program to generate strained conformations was thus evaluated. Finally, the diversity of the conformational ensembles generated by the three programs was evaluated. In each of the three ensembles, the RMSD's between each conformer and all the other conformers generated for the same molecule were calculated, and the percentages of conformers differing from each other more than a series of defined thresholds (0.5, 1.0, 2.0, 3.0, 4.0, 5.0 Å) were determined as a measure of diversity.

Results and Discussion

The first issue we addressed in this study was the choice of the method to calculate conformational energies. The importance of considering solvent effects in conformational energy calculations was pointed out in one of the previous studies.[2] When the calculations are performed in vacuum, internal electrostatic interactions are magnified relative to the solution state. As a consequence of this, folded conformations with internal electrostatic collapse can be generated. Such conformations, characterized by intramolecular interactions between topologically distant atoms involving at least one charged group, are much less populated in solution, where polar or charged groups must be desolvated in order to interact with each other.

Different approaches can be used to account for solvent effects. Methods that include explicit solvent are the most accurate, but they can be very CPU-intensive and they are not practical in most cases. When more than a few molecules must be processed, solvent effects can be accounted for in two different ways: the use of a distance-dependent dielectric constant in the calculation of electrostatic interactions and the use of continuum solvent models. We applied both of these methods in this study and compared the results. In the following we will refer to the calculations performed with distance-dependent dielectric constant as calculations in vacuum. Our calculations show that the strain energies calculated with the two methods are highly correlated and similar in magnitude. As shown in Table 2, the correlation coefficients between vacuum and continuum solvent ranged from 0.79 to 0.87 in our test set for both local and global strain energies.

Table 2. Correlation Coefficients R^2 between the Strain Energies Calculated in Vacuum and Those Calculated Using the Continuum Solvent Model

	local	global
MMFF	0.79	0.80
OPLS-AA	0.82	0.87

The structures of the global minima calculated in vacuum and in continuum solvent differed significantly (RMSD > 1.0 Å) in 66% of the cases when MMFF was used and in 58% of the cases when OPLS-AA was used. In order to determine whether the occurrence of electrostatically collapsed conformations was more frequent in the calculations in vacuum, we analyzed the structures of each global minimum. Structures possessing internal hydrogen bonds involving at least one charged group, with donor and acceptor atoms separated by at least three noncyclic atoms were classified as electrostatically collapsed. As shown in Table 3, global minima affected by electrostatic collapse were only slightly more frequent in the calculations in vacuum than in the calculations in continuum solvent, and in both cases the occurrence of such conformations was relatively low. Moreover, when the ligands with electrostatically collapsed global minima were excluded, the average strain energies of the remaining ligands were no different than

Table 3. Number of Global Minima with Electrostatically Collapsed Structures[a]

	no. of structures	% of structures
MMFF/vacuum	16	10.7
MMFF/GBSA	13	8.7
OPLS-AA/vacuum	16	10.7
OPLS-AA/GBSA	11	7.3

[a] A structure is classified here as electrostatically collapsed if there is an internal hydrogen bond involving at least one charged group and the donor/acceptor pair is separated by at least three non-cyclic atoms. There are a total of 79 ligands in the test set bearing at least one charged group.

those calculated on the entire set, thus showing that even when electrostatic interactions between topologically distant groups were prioritized, the resulting energies did not offset the other contributions. On the basis of these results, it appears that the use of a distance-dependent dielectric constant ($\epsilon = 4r$) and the use of the GBSA continuum solvent model are both viable methods to minimize the artifacts resulting from the overestimation of electrostatic interactions. Since the statistical analysis of the results obtained with the two methods yielded similar conclusions, we will discuss the results of the calculations in vacuum, and we will mention the results obtained in continuum solvent only when they support different conclusions.

The strain energies calculated for the crystallographically determined bound ligands with respect to the closest local minimum ranged from 0 to 16.5 kcal/mol when the calculations were performed with MMFF and from 0 to 13.3 kcal/mol when the calculations were performed with OPLS-AA. The average local strain energies were 1.9 and 2.1 kcal/mol, respectively.

The strain energies calculated relative to the global minima ranged from 0 to 26.2 kcal/mol for MMFF and from 0 to 20.5 kcal/mol for OPLS-AA, with average values of 4.0 and 5.1 kcal/mol, respectively.

The local and global strain energies calculated in vacuum with the two force fields and sorted from lowest to highest are displayed in Figure 1.

The observation that flexible ligands rarely bind in their lowest energy conformation was confirmed by this study. On the basis of the calculations performed with the MMFF force field, only in 14.7% of the cases was the energy of the crystallographic conformation within 0.5 kcal/mol of the global minimum, while the percentage went down to 11.3% when the calculations were performed with OPLS-AA. Surprisingly, the majority of the ligands in our set do not appear to bind in a local minimum conformation either. The percentages of bioactive conformations within 0.5 kcal/mol of the closest local minimum were 34.0% and 35.3% for MMFF and OPLS-AA, respectively.

The correlation between calculated strain energies and different structural descriptors was analyzed. The dependence of the deformation energy on ligand flexibility had been pointed out in one of the previous studies.[1] In the present study the ligands in the test set were divided in four groups based on their flexibility (1−3, 4−6, 7−9, and 10−12 rotatable bonds), and a statistical analysis of the distribution of the strain energies in each group was performed. The results of this analysis are illustrated in Figure 2, which reports for each group the percentage of bioactive conformers

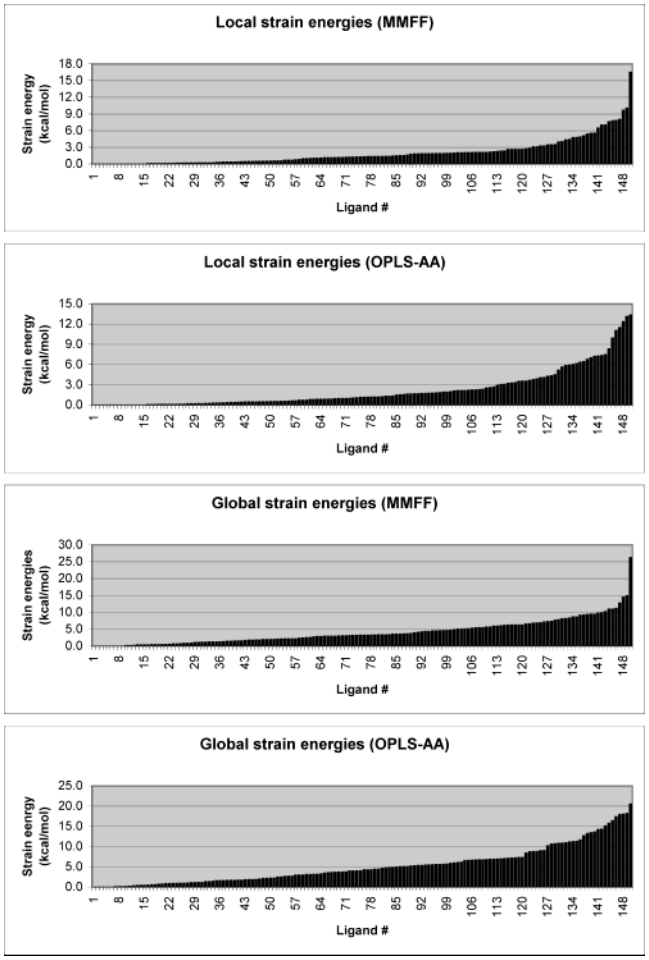

Figure 1. Local and global strain energies calculated in vacuum with MMFF and OPLS-AA for the bioactive conformations of the ligands in the test set, sorted in ascending order.

with calculated strain energies lower than a given cutoff. The two top panels illustrate the distributions of global strain energies according to MMFF and OPLS-AA, respectively, while the bottom panels refer to the local strain energies calculated with the same force fields. This analysis shows a clear correlation between acceptable strain energy and degree of flexibility of the ligand and suggests that different energy thresholds should be applied to ligands with different flexibility when filtering computationally generated conformers. While Bostrom et al. proposed a general threshold of 3 kcal/mol for the global strain of bioactive conformations,[2] these plots indicate that higher strain energies are not uncommon, especially for ligands with more than six rotatable bonds. According to the calculation performed with MMFF, over 60% of the ligands with seven to nine rotatable bonds have global strain energies higher than 3 kcal/mol, and over 25% have local strain energies exceeding the same cutoff. From a practical standpoint, these plots help define thresholds to filter conformations obtained by conformational analysis, docking, or de novo design. For example, if the global strain energies of a given conformational ensemble are calculated with MMFF, the top left panel (Figure 2) suggests that in order to retain 90% of the bioactive conformers present in the ensemble a threshold of 5 kcal/mol should be used for ligands with one to three rotatable bonds, but with the same threshold the expected rate of recovery is only

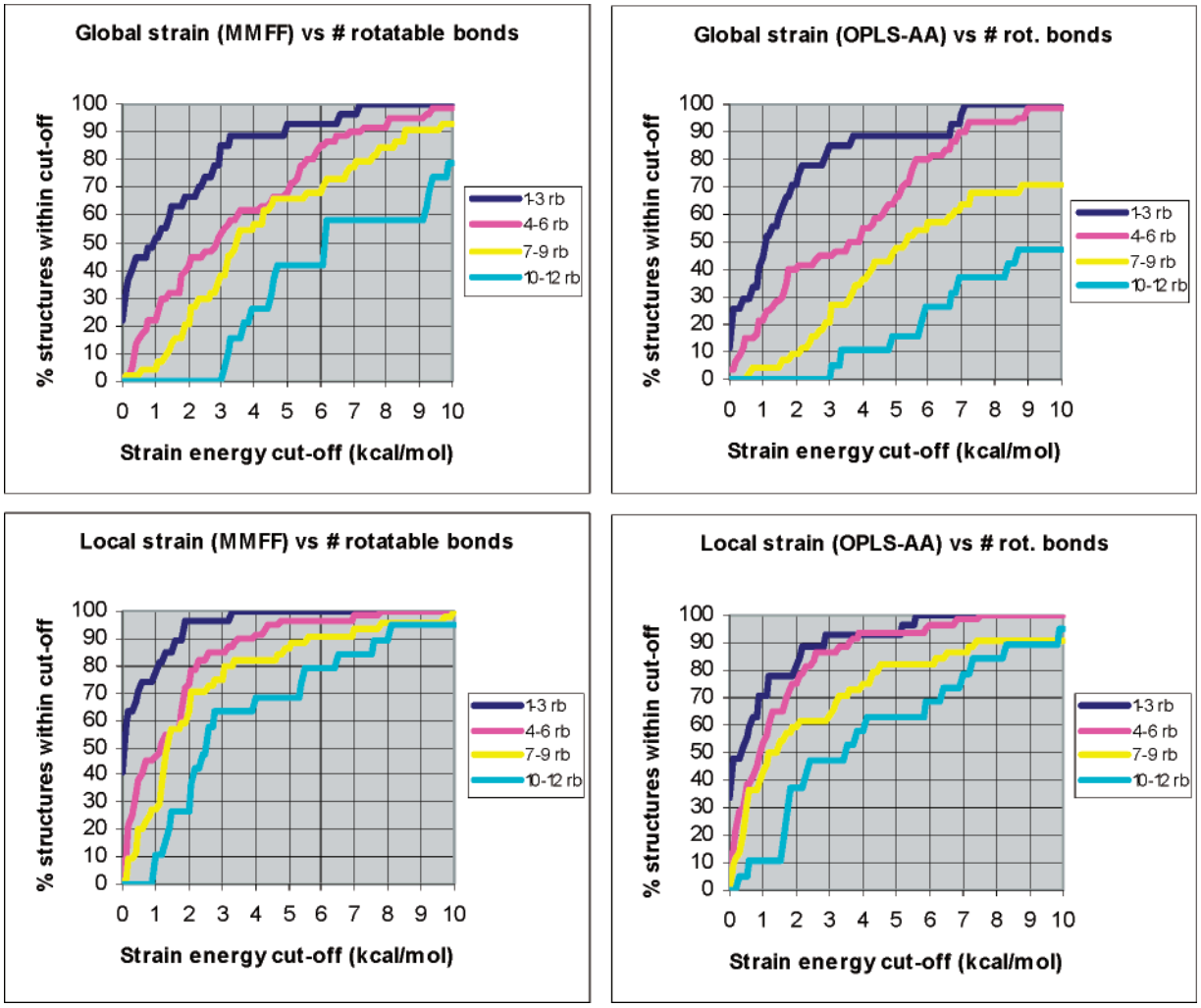

Figure 2. Distribution of the local and global strain energies calculated in vacuum with MMFF and OPLS-AA for the bioactive conformations of the ligands in the test set with respect to their flexibility. The ligands have been divided in four groups, with 1−3, 4−6, 7−9, and 10−12 rotatable bonds. Each curve represents the percentage of bioactive conformers (*y*-axis) with calculated strain energies lower than a given energy cutoff (*x*-axis) for one of the four groups.

Table 4. Local and Global Strain Energy Cutoffs That Retain 90% of the Bioactive Conformations for Molecules with Various Degrees of Flexibility (rb = number of rotatable bonds; energy values in kcal/mol)

	1−3 rb	4−6 rb	7−9 rb	10−12 rb
MMFF/local	1.6	3.5	5.6	7.6
OPLS-AA/local	2.2	3.6	7.4	8.3
MMFF/global	3.3	6.9	8.6	11.1
OPLS-AA/global	3.7	6.9	14.2	17.4

∼40% for ligands with 10 to 12 rotatable bonds. In general, the use of lower thresholds minimizes the occurrence of false positives at the cost of missing potential bioactive conformations, while higher thresholds allow a more quantitative recovery of the relevant conformations but increase the occurrence of false positives. The trends reported in Figure 2 provide the necessary information to make an educated choice in this respect. A possible set of thresholds for local and global strain energies based on the trends described in Figure 2 is reported in Table 4. Such thresholds correspond to the energy cutoffs necessary to retain 90% of the bioactive conformers in our test set. Values are given for both MMFF and OPLS-AA force fields.

In principle the global strain energy provides the best assessment of the price a ligand pays to adopt a bioactive conformation. While an estimate of the global strain is possible whenever an exhaustive conformational analysis is performed and the global minimum can be reliably identified, the local strain energy can be more readily calculated on each individual conformation when docking poses are being analyzed. Hit lists from docking can therefore be pruned based on the local strain energy of the calculated poses using the thresholds determined in this study. Although the local strain energy may only partially account for the cost paid in the deformation associated with binding, using it as a criterion will impact structures that may also be removed if global strain was used. In the test set used in this study local and global strain energies correlate reasonably well ($R^2 = 0.65$ for both MMFF and OPLS-AA), which further supports the use of local strain as a criterion for filtering. The limitation of this approach is that conformations close to a local minimum are going to be preserved whether or not their global strain exceeds the acceptable thresholds.

An important descriptor in a protein−ligand complex is the number of hydrogen bonds between the two partners. Since hydrogen bonds are important contributors to the overall binding energy, it is not unreasonable to expect that hydrogen bonds may be among the main

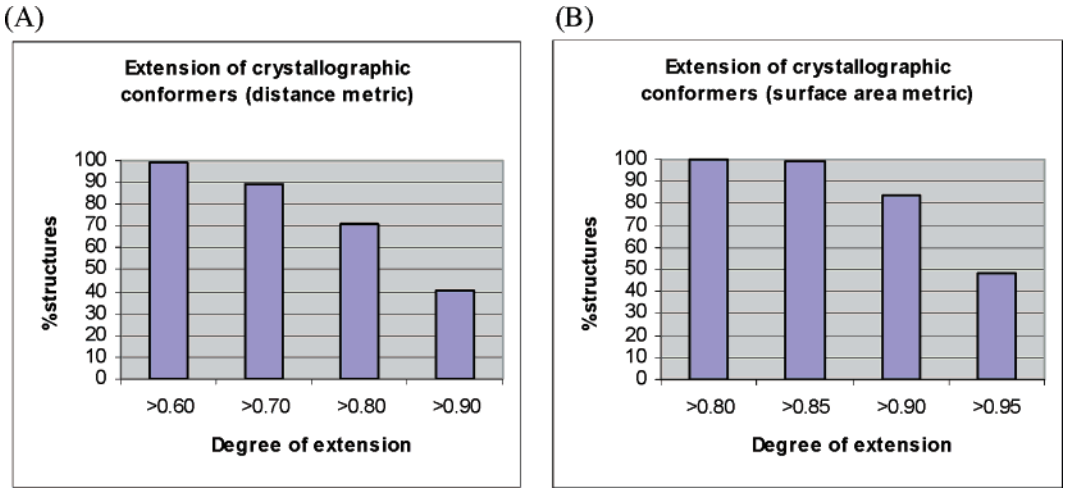

Figure 3. Statistical distribution of the degrees of extension of the bioactive conformations of the ligands in the test set based on two different metrics. In both metrics the degree of extension of a given conformation is calculated as the ratio between the extension of such conformation and the extension of the most extended minimized conformer. A: distribution based on the distance metric; *x*-axis: degree of extension; *y*-axis: percentage of crystallographic conformers with a degree of extension higher than a given cutoff. B: distribution based on the surface area metric; *x*-axis: degree of extension; *y*-axis: percentage of crystallographic conformers with a degree of extension higher than a given cutoff. Details on the two metrics are given in the Methods section.

compensating factors when ligands bind in a high-energy conformation. In other words, the optimal spatial disposition of the hydrogen bond centers of the ligand relative to the protein partner may only be possible in some cases if the ligand incurs some degree of strain. If this was common, some degree of correlation may be observed between the global strain of the bound ligands and the number of hydrogen bonds between ligand and protein. In the test set used in this study no significant correlation was observed. The correlation coefficient R^2 was lower than 0.04 regardless of the method used to calculate the strain energy, and it did not improve when the number of hydrogen bonds was normalized with respect to the number of heavy atoms in the ligand.

Another important descriptor of a protein−ligand complex is the degree of burial of the ligand, which can be quantified as the fraction of the solvent-accessible surface area of the ligand that is buried upon binding. Intuitively, a buried binding site is more likely to force a ligand into a higher energy conformation in order to complement its shape, while the requirements of a more open binding site should be less stringent. Based on this assumption, a certain correlation might be expected between degree of burial and global strain energy of the ligands. Again, no significant correlation was observed in this study, even when the strain energies were normalized relative to the number of rotatable bonds ($R^2 < 0.01$ regardless of the method).

Finally, since the internal strain of the ligand contributes to the overall free energy of binding, a negative correlation should be expected between global strain energy and binding affinity, with tighter ligands generally binding with lower strain than weaker ligands. However, no correlation between global strain and pK_i was found in our test set. The correlation coefficient R^2 was lower than 0.01 regardless of the method used in the energy calculations, and it did not improve when the strain energy was normalized relative to the number of rotatable bonds. This observation seems to indicate that in the majority of the cases higher-energy conformational rearrangements in the small molecule are triggered by overwhelmingly favorable interactions with the protein, which offset the penalty paid in reorganization energy.

An aspect that is closely related to the energetics of binding is the degree of unfolding observed in bound conformations of ligands relative to their free state. As described in the Methods section, we used two metrics to quantitate the extension of a structure: the distance between the two most widely separated atoms and the solvent-accessible surface area. When a molecule is completely extended, these two parameters are maximized. The statistical analysis of the degrees of folding of the ligands in our test set is illustrated in Figure 3. Panel A shows the distribution of the degrees of extension based on the distance metric, panel B shows the distribution based on the surface area metric (see Methods section for details on the two metrics). The analysis seems to indicate that the majority of the ligands bind in an extended conformation. However, the average extensions of the crystallographic conformations are only slightly higher than those of the corresponding global minima. Based on the distance metric, the average degree of extension of the bound ligands is 0.85, while the average degree of extension of the global minima ranges from 0.82 to 0.85 depending on the method used to generate them. Based on the surface area metric, the average degree of extension of the bound ligands is 0.94, while for the global minima it ranges from 0.91 to 0.94.

The difference is more significant when the ligands with the highest global strain energies in their bound conformation are considered. The 20 most strained ligands according to MMFF have global strain energies ranging from 8.0 to 26.2 kcal/mol. Based on the distance metric, the average degree of extension of the bound conformations of these ligands is 0.85, while the average degree of extension of the corresponding global minima is 0.78. Based on the surface area metric, the average values are 0.94 and 0.89 for the bound conformations and the global minima, respectively. These differences are significant considering that over 95% of the crystal-

lographic conformers have degrees of extension over 0.65 according to the distance metric and over 0.85 according to the surface area metric. Analysis of the 20 global minima shows that only in two cases hydrogen bonds between topologically distant atoms, involving one or more charged groups (electrostatic collapse), are present, while at least 12 of the global minima are characterized by hydrophobic interactions between topologically distant groups, resulting in a folded conformation (hydrophobic collapse). Analysis of the corresponding crystal structures shows that in at least 6 of the 12 cases most of such hydrophobic interactions are removed as a result of unfolding. The five ligands with the largest degree of unfolding between global minimum and crystal structure (the largest positive change in the degree of extension based on the two metrics used in this study) are among those six.

The same analysis on the 20 most strained ligands according to OPLS-AA, with global strain energies ranging from 10.8 to 20.5 kcal/mol, yields similar results. Based on the distance metric, the average degree of extension of the bound conformations of these ligands is 0.82, while the average degree of extension of the corresponding global minima is 0.69. Based on the surface area metric, the average values are 0.93 and 0.84 for the bound conformations and the global minima, respectively. Analysis of the structures of the 20 global minima shows 4 cases of electrostatic collapse and 15 cases of hydrophobic collapse. Analysis of the corresponding crystal structures shows that in 9 of the 15 cases most of the hydrophobic interactions between topologically distant groups are removed, again as a result of unfolding. Eight of the nine ligands with the largest degree of unfolding between global minimum and crystal structure are among those nine.

This analysis suggests that one of the factors determining high strain energies in bound ligands may be the process of unfolding, in which the intramolecular interactions between hydrophobic groups of the ligand in aqueous solution are released to increase the hydrophobic surface available for similar interactions with the binding site of the protein. When the difference between calculated global minimum and experimental bioactive conformation is indicative of this process, the calculated reorganization energy reflects the cost involved in disrupting such intramolecular interactions.

The tendency of hydrophobic molecules to adopt folded conformations in aqueous solutions has been observed experimentally in several cases.[24−26] The same phenomenon is at the basis of the globular shape that most proteins adopt, confining the hydrophobic residues to the core of the structure and predominantly exposing polar residues to the interaction with solvent. Although the hypothesis emerging from this study is consistent with this knowledge, only experiments comparing the conformation of the same ligand in solution and in the protein binding site could provide the necessary evidence to confirm our observation.

In the final part of this study we evaluated three tools for fast conformational analysis, encompassing three different types of search algorithm: Omega, which employs a rule-based algorithm,[27,28] Catalyst (FAST mode), which uses the poling algorithm to maximize diversity, and ICM, which uses a stochastic search algorithm combined with force field minimization in internal coordinates. The ligands from our test set of 150 complexes were used in the evaluation, and the programs were ranked based on four different metrics: the ability to find the bioactive conformation, the ability to find the global minimum, the diversity of the generated conformational ensembles, and the distribution of the strain energies of the generated conformers relative to the closest local minima. The results of this comparison are illustrated in Figure 4. As shown in panel A, the ability of the three programs to generate conformers close to the bioactive conformation was comparable. Omega found at least one conformer within 0.5 Å of the bioactive conformation 27% of the times, versus 20% for both ICM and Catalyst, while for all three programs the closest conformer was within 1.0 Å of the crystallographic conformation about 65% of the times. As shown in panel B, ICM outperformed Omega and Catalyst with respect to finding conformers close to the global minimum. ICM, which is the only program of the three that energy-minimizes every conformer in the search, found at least one conformer within 0.5 Å of the global minimum 35% of the times, versus 28% for Catalyst and 26% for Omega. The relative percentages were 77% for ICM, 67% for Catalyst, and 66% for Omega at 1.0 Å cutoff. Panel C clearly shows that the conformational ensembles generated by Catalyst were significantly more diverse than those generated by ICM, which in turn outperforms Omega at RMSD cutoffs of 1.0 Å or higher. This observation seems in contrast with the fact that Catalyst and Omega performed similarly in finding the bioactive conformation, and it points to the fact that Catalyst may extensively explore regions of the conformational space that are not relevant for binding. ICM had the highest tendency to generate duplicate conformers, as indicated by the percentages of the generated conformers differing by less than 0.5 Å from each other. Finally, as shown in panel D, the conformers generated by ICM had a generally lower local strain energy relative to those generated by Omega and Catalyst. This is also consistent with the fact that the ICM-generated conformers are energy-minimized during the search, unlike those generated by Omega and Catalyst.

In terms of speed, the average computing time for conformer generation was 2 s/mol with Omega and 3 min/mol with ICM on a Pentium IV 2.2 GHz, while Catalyst averaged 1.2 min/mol on an SGI R10K 195 MHz.

On the basis of this analysis, Omega achieves the best balance between speed and performance among the programs examined, and it appears to be the most suitable tool for conformational analysis on large chemical databases.

In order to apply the findings of this study to refine the conformational ensembles generated with Omega, the conformational energies of each conformer should be calculated with the same force fields used here. A hypothetical protocol could include conformational search with Omega, energy minimization with MMFF or OPLS-AA, and final filtration based on the global strain energies. To apply the guidelines defined earlier in a rigorous manner, the global minima should be calculated separately with the extensive and time-consuming

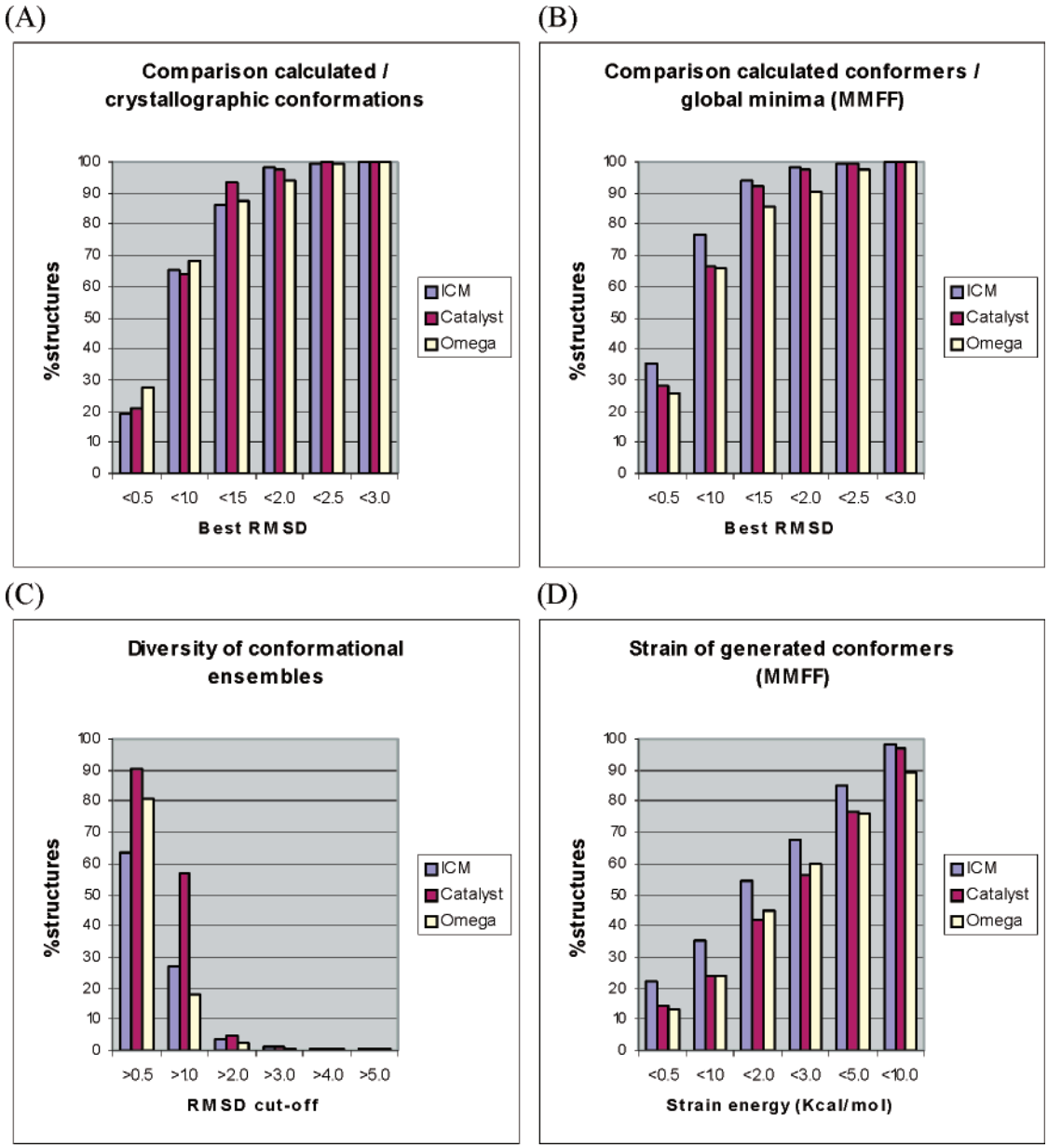

Figure 4. Graphical representation of the performance of the three conformational search engines examined in this work based on four different metrics. A: for each method, the crystallographic conformation of each ligand is compared to the closest conformer generated in the search for the same molecule; *x*-axis: cutoffs for the RMSD's of the closest conformers relative to the corresponding crystallographic conformations; *y*-axis: percentage of closest conformers with an RMSD within a given cutoff of the corresponding crystallographic conformers. B: for each method, the lowest energy conformation of each ligand determined in the first part of the study is compared to the closest conformer generated in the search for the same molecule; *x*-axis: cutoffs for the RMSD's of the closest conformers relative to the corresponding global minima; *y*-axis: percentage of closest conformers with an RMSD within a given cutoff of the corresponding global minima. C: for each method, the RMSD's between each conformer and all the other conformers generated for the same molecule were calculated; *x*-axis: RMSD cutoffs between pairs of conformers; *y*-axis: percentage of conformers that differ more than a given RMSD cutoff from each other. D: for each method, the strain energies of each generated conformer were calculated relative to the closest local minimum with MMFF in vacuum; *x*-axis: local strain energy cutoffs; *y*-axis: percentage of generated conformers with local strain energies within a given cutoff.

approaches described above. Such resolution would defeat the purpose of using a fast tool for conformational analysis. A practical and convenient solution would be to use the global minima from the minimized conformational ensemble as references for the strain energy calculations. To assess the validity of this approximation, we performed energy minimization of the Omega-generated conformers with both MMFF and OPLS-AA, and we compared the energies of the global minima thus obtained with the energies of the global minima previously determined with more expensive procedures. The result was encouraging: the global minima identified with the sequence Omega/energy minimization, with MMFF as a force field, were within 0.5 kcal/mol of the previously determined ones in 73% of the cases, and within 2.0 kcal/mol in 90% of the cases. When OPLS-AA was used in the minimization step, the outcome was almost identical (73% within 0.5 kcal/mol and 88% within 2.0 kcal/mol). This showed that the energy minima calculated with this method are a good approximation of the actual global minima in the vast majority of the cases. Analysis of the ensemble mini-

mized with MMFF showed that at least one conformation within 1.0 Å of the bioactive conformation was present for 70% of the compounds. In particular, at least one conformation meeting this criterion was found for 89% of the ligands with one to three rotatable bonds, 88% of the ligands with four to six rotatable bonds, 59% of the ligands in the seven to nine range, and in only 11% of the ligands in the 10 to 12 range. This suggests that the protocol can be effectively applied on compounds with up to eight to nine rotatable bonds, while for more flexible ligands, a more thorough and CPU-intensive search procedure is probably necessary. The energies of the conformers within 1.0 Å of the corresponding bioactive conformations were compared to those of the corresponding global minima calculated with the Omega/MMFF sequence. All the conformers in this subset had energies within 7 kcal/mol of the corresponding global minima, 92.3% had energies within 5 kcal/mol of the corresponding global minima and 79% had energies within 3 kcal/mol of the corresponding global minima. In the ensemble minimized with OPLS-AA a conformation within 1.0 Å of the corresponding bioactive conformation was present for 66% of the compounds. In this case, at least one conformation meeting this criterion was found for 81% of the ligands with one to three rotatable bonds, 80% of the ligands with four to six rotatable bonds, 61% of the ligands in the seven to nine range, and in only 16% of the ligands in the 10 to 12 range. This is consistent with the observation based on the ensemble minimized with MMFF. All the OPLS-AA-minimized conformers that best approximate the corresponding bioactive conformer within 1.0 Å had energies within 7 kcal/mol of the corresponding global minima with only one exception, while 92% were within 5 kcal/mol and 73% were within 3 kcal/mol.

These results suggest that, when a set of molecules with up to eight to nine rotatable bonds is subject to conformational search with Omega followed by energy minimization with either MMFF or OPLS-AA, an energy threshold of 5 kcal/mol above the global minima is appropriate for the final filtration step. This suggestion may seem in conflict with the results illustrated in Figure 2, which support the use of higher thresholds. However, all the conformers obtained at the end of the protocol described here are energy-minimized. The 5 kcal/mol energy window is expected to include the vast majority of those minimized conformers that best approximate the corresponding bioactive conformers. As pointed out earlier, most of the actual bioactive conformations are not in a local minimum, and their global strain is generally higher. If the conformations to be evaluated are not energy minimized, a broader spread of global strains is expected for the biologically relevant conformers, and the higher thresholds defined in Figure 2 or Table 4 should be applied at the filtration stage. A simplified approach would entail the choice of a single energy threshold based on the highest degree of flexibility represented in the compound set, while a more rigorous approach would take the flexibility of each molecule into account using different thresholds, according to the correlations we observed. A single threshold appears to be a reasonable choice on a set of minimized conformers, in which a relatively narrow spread of global strain energies is expected. When unminimized conformational ensembles or sets of docking poses are evaluated, the use of different thresholds for ligands with different degrees of flexibility is advisable.

Conclusions

Although the concept of induced fit was introduced many years ago, few reports have been published that address the nature and the energetics of the conformational changes small molecules undergo upon binding. The study described in this paper is the most thorough performed to date on this topic, and the first to be focused on drug-like molecules. The results of this study support some of the previous findings, provide some new insights on the conformational behavior of protein-bound ligands, and offer useful guidelines on how to assess the biological relevance of conformations generated with computational methods.

The common knowledge that ligands rarely bind in their lowest energy conformation was confirmed by calculations performed with two different and highly regarded force fields, MMFF and OPLS-AA. We also found that over 60% of the ligands do not bind in a local minimum conformation. The magnitudes of the strain energies calculated for some ligands were consistently high according to both force fields, with at least 10% of the ligands binding in conformations over 9 kcal/mol above the global minimum. This seems to indicate that very large and expensive conformational rearrangements can be tolerated in some cases without penalizing the tightness of binding, especially for very flexible ligands. A clear correlation was in fact observed between acceptable strain energy and number of rotatable bonds, while there was no correlation between strain energy and binding affinity. Comparison of the most strained bound conformations with the corresponding global minima suggested that one of the factors accounting for high strain energies could be the process of unfolding that exposes hydrophobic surfaces previously interacting with each other to the contact with protein residues. The analysis of the degree of folding of the bound ligands confirmed the general tendency of flexible molecules to bind in an extended conformation.

We also evaluated three reputable conformational analysis tools and identified Omega as the program that achieves the best compromise between speed and accuracy. Our results suggest that, for sets of compound with up to nine rotatable bonds, a sequence including conformational search with Omega, minimization with MMFF or OPLS-AA, and filtration of the final ensemble with an energy cutoff of 5 kcal/mol relative to global minimum identified from the search would be a suitable protocol to generate conformational ensembles enriched in biologically relevant conformations. When ensembles of unminimized conformations or sets of docking poses are evaluated, a broader energy window should be allowed, and the use of flexibility-dependent thresholds is expected to increase the prevalence of bioactive conformers. The distribution of the calculated strain energies reported in this paper for a large set of pharmaceutically relevant ligands provides guidelines for the choice of such thresholds. The results also indicate that the strain energies calculated with respect

to the closest local minimum can be efficiently used as an alternative criterion to assess the biological relevance of conformations generated with docking methods.

From a scientific point of view, more experimental data on the conformations of drug-like molecules in solution would be desirable to further support the findings of this study. In the meantime this work provides useful guidelines on how to use the available methods to maximize the prevalence of biologically relevant conformations in computationally generated conformational ensembles.

Acknowledgment. We thank Dr. John H. Van Drie for careful reading of the manuscript and valuable suggestions.

References

(1) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. Conformational changes of small molecules binding to proteins. *Bioorg Med Chem* **1995**, *3*, 411−428.
(2) Bostrom, J.; Norrby, P. O.; Liljefors, T. Conformational energy penalties of protein-bound ligands. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 383−396.
(3) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98−104.
(4) Thoma, J. A.; Koshland, D. E. Competitive Inhibition by Substrate During Enzyme Action Evidence for the Induced-Fit Theory. *J. Am. Chem. Soc.* **1960**, *82*, 3329−3333.
(5) Koshland, D. E. Correlation of Structure and Function in Enzyme Action. *Science* **1963**, *142*, 1533−1541.
(6) Koshland, D. E., Jr. Conformation Changes at the Active Site During Enzyme Action. *Fed. Proc.* **1964**, *23*, 719−726.
(7) Yankeelov, J. A., Jr.; Koshland, D. E., Jr. Evidence for Conformation Changes Induced by Substrates of Phosphoglucomutase. *J. Biol. Chem.* **1965**, *240*, 1593−1602.
(8) Vieth, M.; Hirst, J. D.; Brooks, C. L., 3rd Do active site conformations of small ligands correspond to low free-energy solution structures? *J. Comput.-Aided Mol. Des.* **1998**, *12*, 563−572.
(9) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(10) Halgren, T. A. Merck molecular force field. II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(11) Halgren, T. A. Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553−586.
(12) Jorgensen, W. L.; Maxwell, D.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(13) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(14) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(15) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−2303.
(16) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N. et al. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28*, 235−242.
(17) Kolossváry, I.; Guida, W. C. Low Mode Search. An Efficient, Automated Computational Method for Conformational Analysis: Application to Cyclic and Acyclic Alkanes and Cyclic Peptides. *J. Am. Chem. Soc.* **1996**, *118*, 5011−5019.
(18) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(19) Smellie, A.; Teig, S. L.; Towbin, P. Poling: Promoting Conformational Variation. *J. Comput. Chem.* **1995**, *16*, 171−187.
(20) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S. et al. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(21) Abagyan, R.; Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983−1002.
(22) Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G. et al. Energy parameters in polypeptides. 10. Improved geometrical parameters and nonbonded interactions for use in the ECEPP/3 algorithm, with application to proline-containing peptides. *J. Phys. Chem.* **1992**, *96*, 6472−6484.
(23) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic Generation of 3D-Atomic Coordinates for Organic Molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537−547.
(24) Jankowiak, R.; Lin, C. H.; Zamzow, D.; Roberts, K. P.; Li, K. M. et al. Spectral and conformational analysis of deoxyadenosine adducts derived from syn- and anti-Dibenzo[a, l]pyrene diol epoxides: fluorescence studies. *Chem. Res. Toxicol.* **1999**, *12*, 768−777.
(25) Smith, P. E.; Tanner, J. J. Conformations of nicotinamide adenine dinucleotide (NAD(+)) in various environments. *J. Mol. Recognit.* **2000**, *13*, 27−34.
(26) Valero, M.; Costa, S. M. B.; Santos, M. A. Conformations of a non-steroidal anti-inflammatory drug nabumetone in aqueous environments. *J. Photochem. Photobiol.* **2000**, *132*, 67−74.
(27) Bostrom, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graph. Model.* **2003**, *21*, 449−462.
(28) Klebe, G.; Mietzner, T. A fast and efficient method to generate biologically relevant conformations. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583−606.

JM030563W