Comparative Analysis of Protein-Bound Ligand Conformations with Respect to Catalyst's Conformational Space Subsampling Algorithms

Johannes Kirchmair,[†] Christian Laggner,[†] Gerhard Wolber,[‡] and Thierry Langer*[,†,‡]

Department of Pharmaceutical Chemistry, Institute of Pharmacy, University of Innsbruck, Innrain 52,
A-6020 Innsbruck, Austria, and Inte:Ligand Software-Entwicklungs und Consulting GmbH,
Clemens Maria Hofbauer-Gasse 6, A-2344 Maria Enzersdorf, Austria

We examined the quality of Catalyst's conformational model generation algorithm via a large scale study based on the crystal structures of a sample of 510 pharmaceutically relevant protein−ligand complexes extracted from the Protein Data Bank (PDB). Our results show that the tested algorithms implemented within Catalyst are able to produce high quality conformers, which in most of the cases are well suited for in silico drug research. Catalyst-specific settings were analyzed, such as the method used for the conformational model generation (FAST vs BEST) and the maximum number of generated conformers. By setting these options for higher fitting quality, the average RMS values describing the similarity of experimental and simulated conformers were improved from an RMS of 1.06 with max. 50 FAST generated conformers to an RMS of 0.93 with max. 255 BEST generated conformers, which represents an improvement by 12%. Each method provides best fitting conformers with an RMS value < 1.50 in more than 80% of all cases. We analyzed the computing time/quality ratio of various conformational model generation settings and examined ligands in high energy conformations. Furthermore, properties of the same ligands in various proteins were investigated, and the fitting qualities of experimental conformations from the PDB and the Cambridge Structural Database (CSD) were compared. One of the most important conclusions of former studies, the fact that bioactive conformers often have energy high above that of global minima, was confirmed.

INTRODUCTION

In 3D space, any arrangement of the atoms of a molecule in space defines its 'geometry'. Hence, the geometry of a molecule is the exact orientation of all atoms of a molecule with defined configuration and constitution in space. Geometries that represent various local energy minima are termed 'conformers'. Mostly, the energetic differences between conformations are very low, and thus they cannot be isolated. In our study, we often refer to higher energetic geometries as conformers because of Catalyst's conformational model generation algorithms calculating high energy values that might not represent the real conformational energy.

The biological activity of drug molecules is supposed to be related to only a very limited conformational space. Since the internal energy of the ligand contributes to the total free binding energy, bioactive conformations of ligands consequentially have a relatively low energy. However, several studies[1−5] examined that most of the time the conformational energy of bioactive conformers is well above the global energy minimum. The induced fit theory[7−11] affirms this observation by postulating that the binding of a ligand to a protein is based on a forced reorganization of both ligand and protein in order to adopt complementary shapes required for binding. The reorganization energy of the ligand-protein binding process is compensated by newly established interactions between ligand and protein.

Conformational model generation attempts to representatively cover the conformational space of a molecule, within a given energy range, by a limited number of low-energy conformers. The most common methods for conformational analysis are those that are able to identify minima on the potential energy surface. The time required for conformation generation depends on the hardware and on the algorithm used for the energy calculation. Most of the conformational model generation programs use force fields and quantum mechanics to determine the preferred conformations of a ligand in gaseous phase or in the bioactive conformation.[12] A number of computational approaches for the analysis of the binding of small ligands to proteins have been developed, including pharmacophore search and ligand docking. All these methods have to be evaluated by the use of experimental data in order to get successfully established.

Within the past few years, five rather limited studies and one large scale analysis on the conformational strain of bound ligands were reported. The first study analyzed the energetic difference between bioactive conformations of 27 Protein Data Bank (PDB)[13] ligands and their lowest energy conformation in a vacuum calculated using the CHARMm force field.[1] To minimize artifacts due to electrostatic collapse of charged groups calculated in vacuo, ionizable molecules were treated as neutral. The average calculated strain energy of all calculated ligands was 15.9 ± 11.5 kcal/mol. The authors found the number of rotatable bonds and, to a lower extent, the number of hydrogen bond centers in the ligand to have a major impact on average energy and suggested to calculate a huge number of conformers in order to get the biologically

* Corresponding author phone: +43-512-507-5252; fax: +43-512-507-5269; e-mail: Thierry.Langer@uibk.ac.at.
† University of Innsbruck.
‡ Inte:Ligand Software-Entwicklungs und Consulting GmbH.

PROTEIN-BOUND LIGAND CONFORMATIONS

J. Chem. Inf. Model., Vol. 45, No. 2, 2005 **423**

relevant conformation. A further study attempted to determine the preferred conformations of 33 protein−ligand complexes selected from the PDB in aqueous solution.[2] Both the MM3 and the AMBER force fields were used to determine the difference between the experimental bound conformation and the lowest energy conformation of each ligand in water. 30% of the compounds had calculated strain energies higher than 3 kcal/mol. However, the authors suggested that a 3 kcal/mol energy threshold could be used in general and that higher energy thresholds were caused by inadequate computational methods or insufficient resolution of the corresponding crystal structure. A study by Poulsen on the evaluation of Catalyst's conformational search algorithm examined the energy thresholds of conformations of 11 NK1/NK2 antagonists.[3] The study confirms that bioactive conformers may have higher strain energy thresholds than the respective conformers in gaseous phase. In the fourth study, a systematic conformational search was performed on 10 small, polar ligands to calculate the solution free energies using a continuum solvation model.[4] Thus generated low energetic conformers were observed to have only little similarity in torsional space with the crystallographic conformers. Since the calculated energy thresholds for the crystallographic conformers reached from 0 to 9 kcal/mol, the authors suggested using a 10 kcal/mol energy cutoff for small, polar ligands. However, an investigation of the anchoring points of the lowest energetic conformers in solution and the bioactive conformation detected similar positions of the same atoms to a large extent causing the authors to infer that pharmacophore models based on low energetic conformers are still an efficient approach for molecular drug design. Boström published a study on 32 PDB ligands evaluating Catalyst, Flo99, Confort, OMEGA, and Macromodel.[5] He concluded that the Low-Mode conformational search method implemented in MacroModel performs better than other algorithms. Confort and Catalyst, with diverse conformational sampling, performed worst. The conformational models produced by OMEGA were found to be strongly depending on the input conformation, since OMEGA does not change bond lengths or standard bond angles.

The insights of these five studies are limited by the small test sets employed and the fact that the majority of ligands used in these samples are not drug-like.

The most extensive study on this issue that is concurrently focused on drug-like molecules is an analysis of the nature and the energetics of the conformational changes ligands experience during the binding process.[6] 150 crystal structures of pharmaceutically relevant protein−ligand complexes were analyzed, and the common knowledge that ligands rarely bind in their lowest calculated energy conformation was confirmed. Using various methods, approximately 60% of the ligands were determined to bind with strain energies lower than 5 kcal/mol and at least 10% with energies above 9 kcal/mol. Ligand flexibility was found to correlate with strain energy, while there was no correlation between binding affinity and strain energy. The study provides an overview across three different programs for conformational analysis: Macromodel 8.0, Catalyst 4.6, and ICM 3.0. These tools were compared in their ability to find the bioactive conformation and the global minimum. The diversity of the generated conformational models and the distribution of the strain energies of the generated conformers relative to the closest local minima were investigated.

Although this study gives some insights to address this topic, until now there is no study providing a detailed analysis of Catalyst's[14] CHARMm force field and its ability of determining the bioactive conformation. Moreover, there was yet no analysis examining Catalyst's different conformational model generation algorithms and the influence of the number of generated molecules on the quality of the conformational model. Our study provides an extensive guide for the Catalyst user to optimize generation settings including algorithm settings and energy threshold settings. We highlight the advantages and disadvantages of Catalyst's conformational model generation based on the evaluation of six different Catalyst settings and give advanced insight to the strain energy issue. Our results uncovered a weakness of Catalyst's CHARMm force field obtaining considerably higher RMS and energy values for sulfonamides. Last, the paper also analyzes the differences between ligand structures taken from the PDB and the Cambridge Structural Database CSD.[15]

METHODS

The study on PDB protein−ligand complexes is based on a carefully selected test set of 510 drug-like ligands. For the evaluation of Catalyst's conformational model generation, the following major steps were performed: (a) extraction of a ligand set from the PDB to be tested, (b) conformational model generation with different methods and settings, (c) computation of fit values between the bioactive conformation and all generated conformations by usage of the module Citest, and (d) data analysis of the best fitting conformation of each ligand.

Catalyst ConFirm operations for this study were calculated with Accelrys Catalyst 4.8 under Red Hat Linux 9.0. The calculations were executed on a cluster of three Pentium 4 2800 MHz PCs with 1GB RAM and two Athlon XP 1800+ PCs with 512MB RAM, connected by Open-Mosix 2.4.22.

Extraction of a Ligand Set from the PDB. The primary data source for this evaluation was the conformation of ligand molecules stored in the PDB. The software LigandScout,[16,17] a novel program for analysis and visualization of ligand protein complexes, was used to extract the small organic ligands from their respective complexes and to store the structures as MDL MOL files. LigandScout extracts ligands and their close protein environment from the PDB and assigns bond characteristics to all PDB entries in a fully automated and distributed way.[18]

Conformational Model Generation with Different Methods and Settings. To avoid any influences on the generation of conformers by presenting a starting conformation, the extracted MOL files were stripped of their 3D-coordinates via conversion into the SMILES file format. While in MOL files the type, connectivity, and position of each atom are defined, the SMILES files only contain information about the configuration in a very packed but clear way. Existing chirality specifications are retained.

Catalyst's ConFirm Module. Catalyst's ConFirm module is a tool for conformational model generation. It allows the user to calculate coverage-based conformational models using the CHARMm force field.

In command line mode, ConFirm supports automated conformational model generation of sets of compounds.[19] Within Catalyst, ConFirm is used for conformational model generation of single compounds, additionally providing a tethers function for alignment of structurally diverse compounds and hypotheses.

ConFirm is designed to provide comprehensive coverage of the low-energy conformational space by including a representation of all conformations within a user-adaptable energy threshold. It is of extraordinary importance to set a suitable energy limit because if the bioactive conformation has a higher energy than the energy limit, ConFirm will not generate suiting conformers. ConFirm's highly diverse conformational models are based on a poling algorithm that repels too similar conformers.[20,21]

There are two algorithmic methods of conformation generation integrated in ConFirm: the BEST and the FAST algorithm. While both methods emphasize the coverage of conformational space, BEST uses more extensive algorithms to explore conformational space, whereas FAST is more approximate and requires less time.

FAST is recommended for rapidly building large multi-conformer databases. The conformational search algorithm is based on a grid search of torsion values in combination with some energy minimization to resolve high energy configurations such as overlaps of van der Waals radii. FAST's performance is largely improved by the use of a library of stored ring conformations containing simplified approaches about ring systems. However, FAST conformer generation should not be used for molecules containing large, flexible ring systems that have more than eight members, nor should it be used for large, flexible molecules with more than 12 rotatable bonds. In these circumstances, the FAST method does not provide representative coverage of conformational space; the BEST method should be used instead.

The BEST conformational search mode is based on an algorithm that allows all internal coordinates to vary. The method supports user-adaptable poling to ensure maximized diversity of generated conformations. Poling is a technique that forces ConFirm to repel newly generated conformers that are too similar to already existing conformations by assigning an energy penalty.

Conformational Model Generation with Catalyst's ConFirm Module. The ConFirm module takes input files of compounds and produces conformational models for each input compound. The quality of the results mainly depends on two settings: the number of generated conformers (NOC) and the algorithm used for the calculation (BEST or FAST). In our study, we defined the settings used by the maximum number of generated conformers and the method; e.g. 255 FAST is used for a maximum of 255 generated conformers with the FAST algorithm. Starting from their respective SMILES files, six different sets of conformers were generated for each examined compound: 50 FAST, 100 FAST, 255 FAST, 50 BEST, 100 BEST, and 255 BEST. The conformational models were generated automatically by using the ConFirm syntax for sets of compounds, which is able to handle large numbers of compounds. The energy threshold was left at the default value of 20 kcal/mol for all these calculations.

Catalyst's Citest Module. Citest is another module of Catalyst which allows the user to compare two conformers
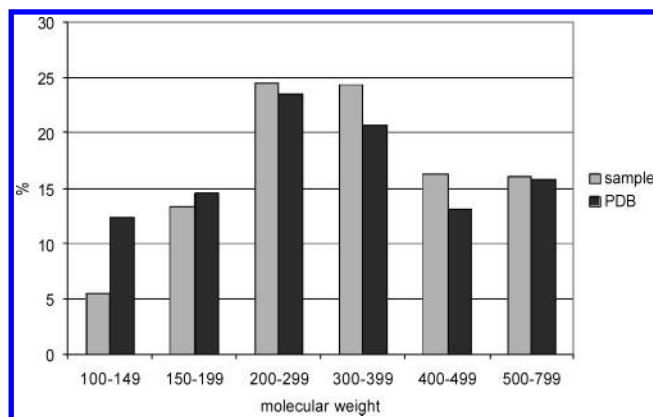

Figure 1. Molecular weight distribution of our sample and the PDB.

of both same or different molecules as well as fit-value calculation of molecules on feature-based hypotheses.[22] In contrast to ConFirm, Citest does not support batch processing, and so it was necessary to use a shell script to automate this process.

Data Analysis. For statistical analysis, the huge amount of data that was stored in log files by citest was processed in an automated way with a Perl script.

RESULTS AND DISCUSSION

The Sample. All used data of bioactive conformers was taken from the PDB database. We filtered the compounds for duplicates and excluded ligands with MW < 100 and > 800 as well as molecules without any carbon atom since they are not likely to be relevant for pharmaceutical protein targets. The sample of 510 ligands was taken from these remaining 2478 compounds. Since compounds of the PDB are known to be moderately error-prone, valid geometry of the ligands as well as the quality of the complexes were affirmed by visual inspection, while concurrently ensuring drug-likeness of the sample.

All results of conformational models are also depending e.g. on the molecular weight and the number of rotatable bonds. Thus, the representative quality of the sample had to be investigated. Besides the constraints described above, the compounds of our sample were selected by randomized search of the edited PDB with respect to the molecular weight distribution and the flexibility distribution to ensure statistical relevance. Both property distributions of the PDB (edited as described above) and our sample are matching very well. Figure 1 shows that there are only minor differences in the molecular weight distribution between the edited PDB and our sample except in the fraction of molecular weight 100–149. The distribution of the rotatable bonds (Figure 2), which represents the molecule's flexibility, also demonstrates large accordance between the edited PDB and the sample. There are less compounds of the lightest and the most rigid fraction in our sample because these are not able to utilize ConFirm's capability of generating huge numbers of conformers. Drug-likeness also determines both properties. Thus, there are less ligands represented in the fraction of highest flexibility (>19 rotatable bonds) shifting the focus to medium flexible compounds containing 11–14 rotatable bonds. These criteria reflect our efforts to take a statistical and pharmaceutical relevant sample of the PDB as the basis of our study.
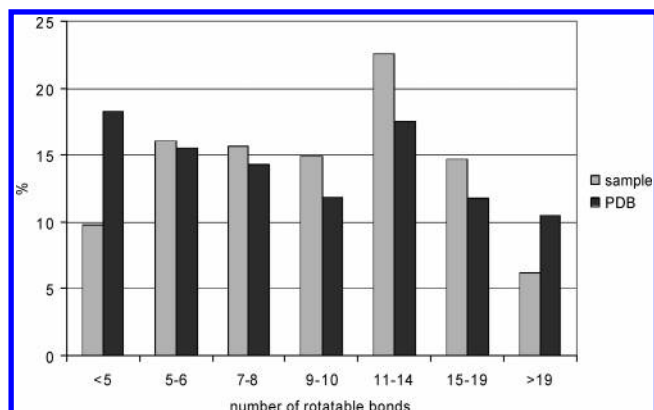
PROTEIN-BOUND LIGAND CONFORMATIONS

J. Chem. Inf. Model., Vol. 45, No. 2, 2005 **425**

Figure 2. Rotatable bonds distribution of our sample and the PDB.

Number of Generated Conformers. The evaluation was performed with maxima of 50, 100, and 255 generated conformers. This setting for the number of generated conformers defines a maximum, while the actual number of generated conformers is often smaller, depending on the molecular weight, the number of rotatable bonds, and the rigidity of the ligands. There is no significant elevation in the number of generated conformers between FAST and BEST (Table 1). The highest average number of generated conformers is 131 (with max. 255 BEST). Models generated with 50 FAST produce only 29% of the conformers that are generated with 255 BEST.

Evaluation of the Fitting (RMS). The quality of fitting is defined by the RMS value. The models are aligned using superposition of the matching heavy atoms between each pair of models. The function to be minimized is defined as the sum of squares of the distances between all pairs of heavy atoms to be superimposed.[23]

Catalyst generates high quality conformers (Table 1) and is able to lower the average RMS from 50 FAST to 255 BEST. Every method provides best fitting conformers with RMS < 1.50 in more than 80% of all cases. Figure 3 shows the spreading of RMS values depending on ConFirm settings (cumulative).

RMS < 0.1. RMS below 0.1 was found in less than 2% of the sample ligands. The fraction of RMS < 0.1 does not rise when generating more conformers using the BEST method, but in FAST mode results are better when creating more conformers. RMS < 0.1 means almost perfect fit.

0.1 ≤ RMS < 0.5. RMS below 0.5 means excellent fit and is achieved for 16−18% of the ligands.

0.5 ≤ RMS < 1.0. 0.5 ≤ RMS < 1.0 means good fit. Catalyst generates the best fitting conformers within this fraction in 32−40% of the time.

1.0 ≤ RMS < 1.5. Acceptable fit is achieved with 1.0 ≤ RMS < 1.5 (29−32%). This is the first fraction in which the number of fittings within this fraction decreases with settings for higher quality (except with 100 BEST), because more ligands can already be fitted with lower RMS.

1.5 ≤ RMS < 2.0. Fittings within this fraction are more or less acceptable (8−12%). Some features may already drift apart.

2.0 ≤ RMS < 3.0. In this RMS range the trend seen in the two previous fractions is continued. Results with 2.0 ≤ RMS < 3.0 are low quality fits (4−6%); the chemical features drift far apart. Thus, these conformations must not be considered as appropriate representations of the biologi-

cally active one. The number of ligands within this fraction descends with more extensive settings.

RMS ≥ 3.0. RMS **above** 3.0 was achieved in less than 0.6% and is not acceptable for molecular drug design. Most fittings (three ligands) were found with 100 FAST. Ligand 638−1QBR achieved the worst fitting of all 510 ligands (RMS = 3.871). This ligand is both very flexible (16 rotatable bonds) and heavy (molecular weight 759), which seem to be the reasons for Catalyst failing to exhaustively explore this large conformational space. The high RMS value is mainly a result of the false position of both benzyl moieties and is not a problem affected by the molecule's symmetry.

Catalyst's average RMS values between the best fitting conformer and the bioactive conformation are improved from RMS = 1.06 with 50 FAST generated conformers to RMS = 0.93 with 255 BEST generated conformers, which represents a reduction of 12%. Average RMS values decrease from 50 FAST generated conformers (RMS = 1.06) to 255 FAST generated conformers (RMS = 0.97) by 8.5% and from 50 BEST (RMS = 1.01) to 255 BEST (RMS = 0.93) by 7.9%. Altogether, the FAST procedure performed well; at least 93% of the ligands are fitting perfectly, well, and more or less acceptable (RMS ≤ 2).

Catalyst's Citest works fine fitting most of the ligands. However, Citest may not manage to fit highly rigid ligands with only one or two rotatable bonds. Most of the time these ligands only have one to five generated conformers registered in the conformational model, as a result of the poling function's restrictions. Easing the user-adaptable poling restrictions of the BEST algorithm may enable Catalyst to get fittings also in these special cases. If the automated fitting procedure of Citest does not work, it is recommended to use the Catalyst GUI for fitting the ligands manually. The GUI provides a tether function which lets the user define atoms of both ligands that mate. Defining and connecting identical atoms in both molecules enables Catalyst to find the best fitting conformer.

In less than 1.5%, Catalyst is able to achieve a fitting only in BEST mode and not in FAST mode as well. Usually, fittings that failed with 50 BEST cannot be accomplished by using settings for the generation of more conformers, either. Thus, only in very special cases, an additional try with 255 BEST might be successful if the ligand does not already fit with lower quality settings. A fitting achieved under these circumstances is mostly due to the generation of a larger number of conformers and not to improving the quality of the generation algorithm.

Energy. In our introduction, we already presented the energy of the best fitting conformer as an interesting aspect in conformational model generation. All former studies and also this large scale study have a high average energy value of the best fitting conformer in common. It should be pointed out that these high internal energy values are artifacts of the programs, possibly arising from minor deviations in bond lengths and bond angles. Both geometric parameters usually do not change significantly from one low-energy conformation to another one. A further source of error is the consideration of the conformations in vacuo. Differences in internal energies between the ensemble of conformations in vacuo and the bioactive conformation might be a further reason for the lack of precision of our current scoring function. However, our experience with Catalyst caused us

Table 1. Average Number of Generated Conformers, Average RMS, and Average Energy of Generated Conformational Models

	50 FAST	100 FAST	255 FAST	50 BEST	100 BEST	255 BEST
average number of generated conformers	38	66	124	43	77	131
relative % of average number of generated conformers	29	50	95	33	59	100
average RMS	1.060	1.009	0.967	1.013	0.970	0.932
% of ligands in RMS fractions (noncumulative)						
RMS < 0.1	1.37	1.76	1.76	1.37	1.37	1.37
0.1 ≤ RMS < 0.5	16.08	15.88	17.65	16.27	17.06	18.24
0.5 ≤ RMS < 1.0	32.16	35.88	38.82	35.49	37.84	39.61
1.0 ≤ RMS < 1.5	31.96	31.37	28.82	29.22	29.41	28.82
1.5 ≤ RMS < 2.0	11.76	10.00	9.02	11.57	8.82	8.43
2.0 ≤ RMS < 3.0	6.47	4.51	3.53	5.88	5.29	3.53
3.0 ≤ RMS < 4.0	0.20	0.59	0.39	0.20	0.20	0.00
average energy [kcal/mol]	5.24	5.64	6.35	4.57	5.71	6.52
ratio of ligands with energy > 10 kcal/mol and RMS < 1 [%]	9.92	12.13	9.46	7.75	12.89	16.56
average RMS of sulfonamides	1.48	1.35	1.29	1.45	1.40	1.30
deviation of sulfonamides from average RMS [%]	40.07	33.56	33.82	43.30	44.69	39.30
average energy of sulfonamides [kcal/mol]	5.47	6.98	8.08	4.91	6.50	9.27
deviation of sulfonamides from average energy [%]	4.42	23.89	27.29	7.35	13.82	42.12
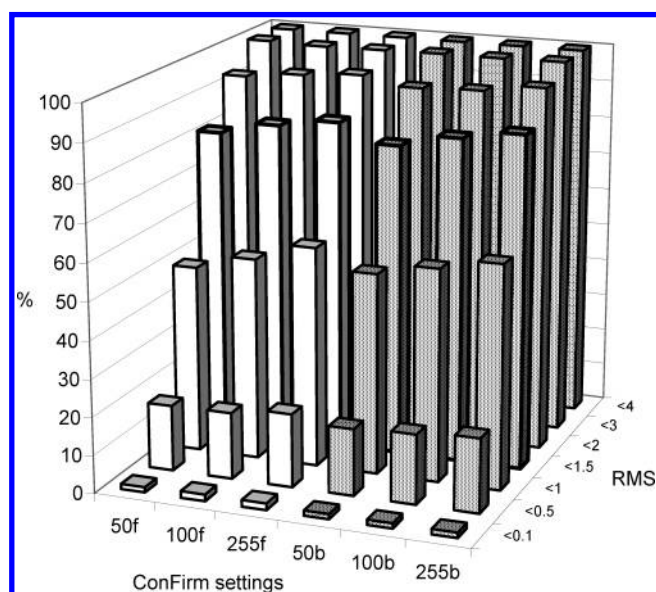
Figure 3. Spreading of the RMS values.

to use ConFirm's default energy range of 20 kcal/mol for all the basic calculations. During our study, we could confirm that this high energy threshold setting, even if its value appears unreasonably high, is necessary to find the bioactive conformation among all geometries calculated.

Table 1 shows an increase of the average energy when generating a higher number of conformers. Thus, with an increasing number of conformers, better fittings, yet energetically higher conformers are achieved. This means that the average energy values calculated for bioactive conformers are significantly higher than those estimated for the minimum energy conformers. The energetic difference to the gas-phase minimum has to be compensated by interactions with the protein at the active site. The average energy of all 510 ligands is between 4.6 and 6.5 kcal/mol, depending on the applied settings. It is noteworthy that the energy values of the best fitting conformers are spread over the full range of 0−20 kcal/mol. The results show that generated conformers with higher average energy are more similar to bioactive conformations than the minimum energy conformers. The ratio of ligands with energy > 10 kcal/mol and RMS < 1 shows that the fraction of higher energetic ligands rises with the lowering of the RMS (Table 1). Compared to FAST,

BEST nearly reduplicates the fraction of the ligands with high energy and low RMS. Figure 5 shows the spreading of the energy of all best fitting conformers with RMS < 2, generated with 255 FAST and 255 BEST. BEST generates higher energetic best fitting conformers; in the lower energetic fraction, 255 FAST achieves more fittings.

The Influence of the Molecular Weight and the Number of Rotatable Bonds. During our investigation of conformer models, we observed a decisive impact of the molecular weight and the number of rotatable bonds on the resulting RMS and energy values of the best fitting conformers. Table 2 describes these dependencies with conformational models generated with the 255 BEST method. In contrast to the heaviest fraction (molecular weight > 500), which reaches an average energy of 9 kcal/mol, light molecules (molecular weight 100−149) have a three times lower average energy level than heavy ligands. However, ligands of high molecular weight are much more likely to develop multiple inter- and intramolecular interactions such as H−bonds to stabilize high energetic conformers. Average RMS rises by 400% from 0.4 to 1.6 between the lightest and the heaviest fraction.

Both energy and RMS values increase with the number of rotatable bonds. The average energy rises from 1.7 kcal/mol with <5 rotatable bonds to 9.6 kcal/mol with >19 rotatable bonds and average RMS from 0.5 with <5 rotatable bonds to 1.8 with >19 rotatable bonds. These effects are caused by the fact that more rotatable bonds result in a larger conformational space.

Combination of the FAST and BEST Methods. With the *keep* option, ConFirm saves previously generated conformers and adds new ones to the set if the calculation procedure is repeated. This experiment was performed to analyze if Catalyst can improve the fitting quality by iterated conformational model generation.

Conformational space of 48 randomly selected ligands of the 510 ligands sample was first analyzed with 255 FAST, and then a second step with 255 BEST was applied on this set, to generate further conformers. The average number of conformers ascended from 96 to 123 when using this combination. However, while the average energy value rose (7.03 with 255 BEST; 7.57 with 255 COMB), the RMS value was only slightly reduced (1.051 with 255 BEST; 1.046 with 255 COMB). Therefore, it is obvious that the combination
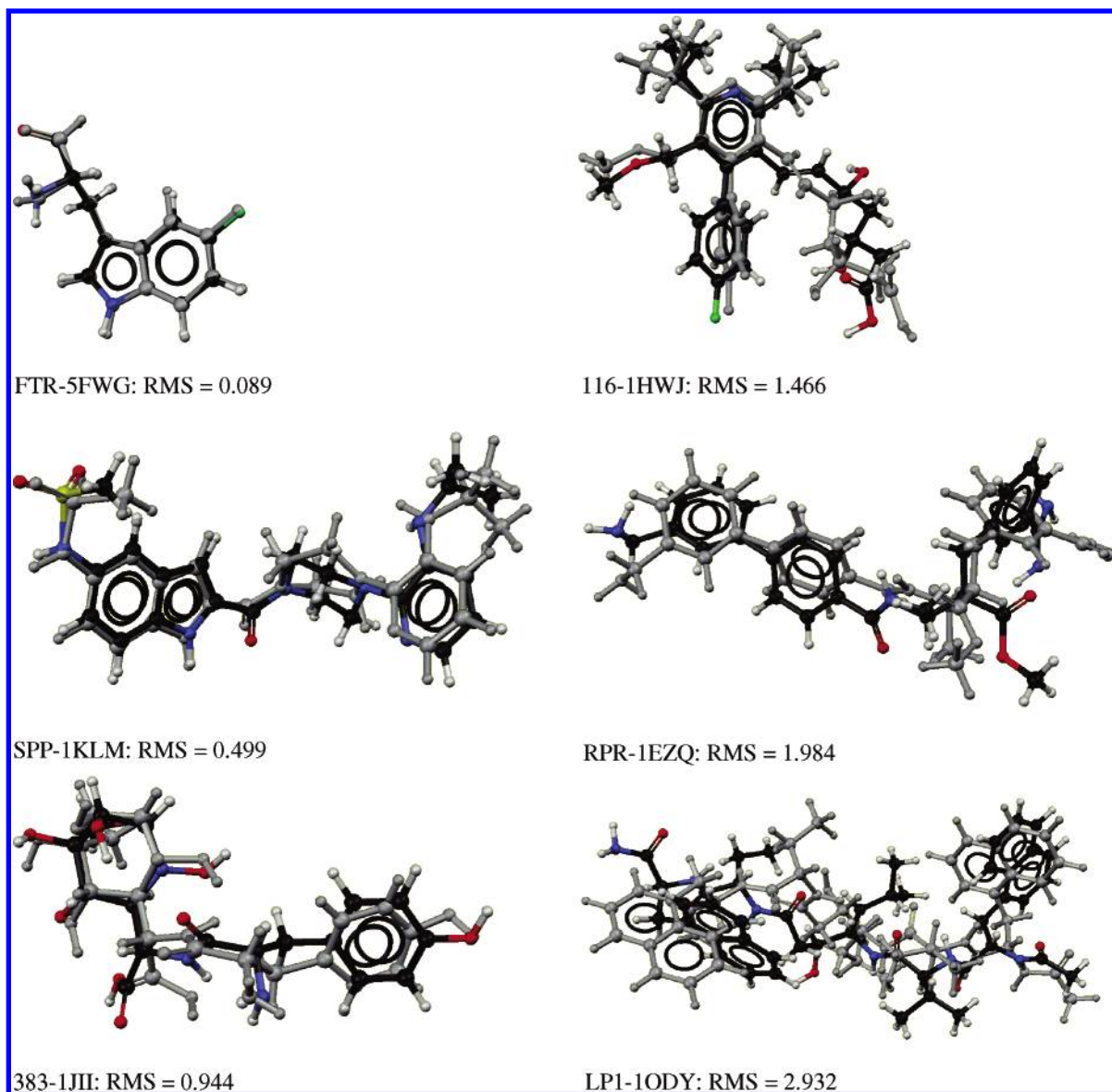
PROTEIN-BOUND LIGAND CONFORMATIONS

J. Chem. Inf. Model., Vol. 45, No. 2, 2005 **427**

Figure 4. Fittings of bioactive conformations of PDB ligands with the best fitting generated conformers.
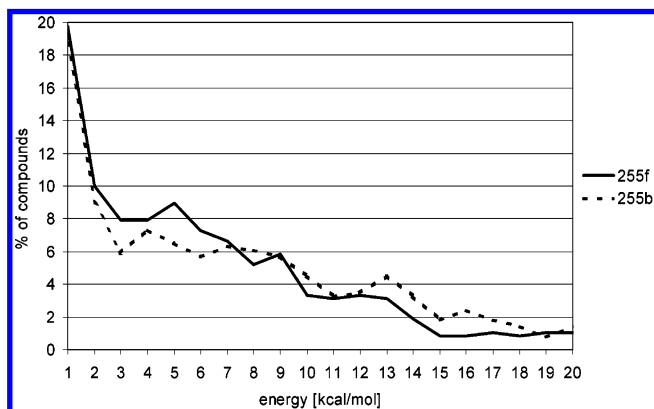

Figure 5. Spreading of the energy of ligands with RMS < 2.

of FAST and BEST is not a suitable way to obtain better fitting results because generation expenditure rises without significantly improving the results.

   The Sulfonamide Issue. During our study, we detected a difference in the fitting quality of ligands with and without *N*-mono- or *N,N*-disubstituted sulfonamide features when we

Table 2. Average RMS and Energy Values Depending on Molecular Weight and on the Number of Rotatable Bonds, Calculated with the 255 BEST Method

molecular weight	no. of molecules	RMS	energy [kcal/mol]
100–149	28	0.42	2.88
150–199	68	0.58	4.61
200–299	125	0.72	5.21
300–399	124	0.88	6.92
400–499	83	1.14	8.31
500–799	82	1.59	8.96

no. of rotatable bonds	no. of molecules	RMS	energy [kcal/mol]
<5	50	0.54	1.71
5–6	82	0.55	3.77
7–8	80	0.69	6.77
9–10	76	0.80	6.74
11–14	115	1.10	8.21
15–19	75	1.38	8.35
>19	32	1.78	9.64

compared 46 sulfonamides of the 510 ligands sample with the sample's average RMS and energy values.
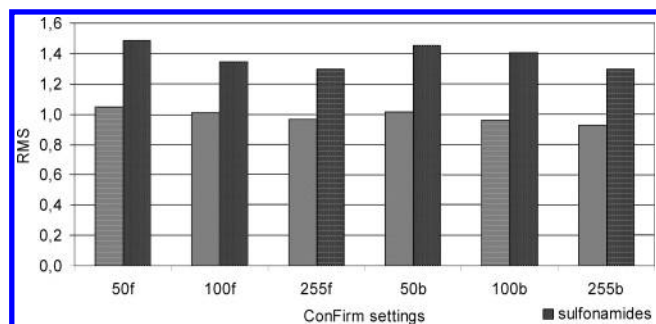
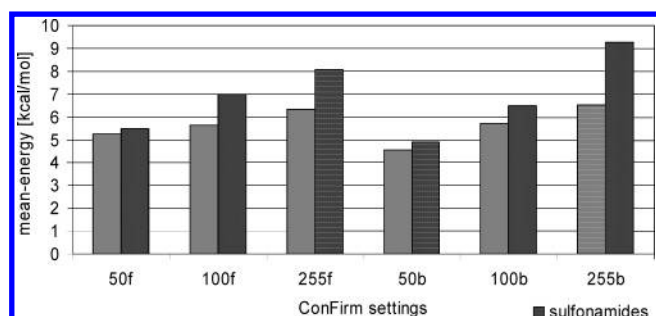**Figure 6.** Average RMS values of sulfonamides and the sample.

Figure 7. Average energy of sulfonamides and the sample.

The RMS values of the best fitting conformers are 34−45% higher than the average values of the whole set (Figure 6, Table 1), and also the average energy rises by 4−42% (especially with 255 BEST) when generating conformers of sulfonamides (Figure 7, Table 1).

The geometry of arylsulfonamides had been studied by ^{15}N NMR[24,25] and X-ray crystallography. It was shown that the nitrogen atom adopts a tetrahedral-like geometry. The most favored conformation is demonstrated in Figure 8.

Catalyst's force field, however, assumes a wrong torsion angle R−S−N−R1/R2 to the lowest energetic conformer of sulfonamides, most of which are S-arylsulfonamides, in about half of all cases. The calculated lowest energy conformation of *N*-monosubstituted sulfonamides often adopts a 'trans'-like conformation, where the two substituents at nitrogen and sulfur share a torsional angle τ of 180°, while experimental data from the PDB and CSD databases confirm a very high preference for a torsional angle of about 60°. In a PDB sample of 53 sulfonamides we detected only two compounds with a torsion angle τ of about 180°, and in a 186 compound sample of the CSD, there were no conformers of this kind at all. Furthermore, there is a strong tendency of Catalyst to produce lowest energetic *N,N*-disubstituted sulfonamide conformers which are arranged as depicted in Figure 8. Within the energy range of 20 kcal/mol, Catalyst is able to generate also conformations that are fitting to the bioactive conformer, but its force field penalizes these conformations with a too high energy. While this may not be a big issue for database screening, special care should be taken when sulfonamides are included in the process of hypothesis generation.

Energy Threshold Settings. These options are used to specify an energy range for the generated conformers. Unless otherwise instructed, the ConFirm module uses a default value of 20 kcal/mol. ConFirm generates conformers in such a way that the energy of each conformer differs from the minimum by no more than the range that was set previously.

Those conformers, whose energy differs from the lowest energy conformer by more than the defined units, are deleted.

We investigated the impact of energy threshold settings on conformational models with three additional maximum energy settings: 10, 15, and 30 kcal/mol. The number of generated conformers rises with the energy range (Table 3). An energy range below 20 kcal/mol is not recommended because of a higher resulting RMS value. The largest improvement of fitting quality can be achieved between 10 and 15 kcal/mol; between these settings, the RMS value decreases by 8.6%. A value higher than 20 kcal/mol does not improve the fitting results, while the average energy rises. Upon fitting to their bioactive conformers, only 5% of the ligands with molecular weight below 200 have an energy higher than 15 kcal/mol and 11% have an energy higher than 10 kcal/mol. Thus, for molecules with low molecular weight an energy threshold of 15 kcal/mol would suffice most of the time, while in general, because of the approximations of the CHARMm force field, one is advised to use the default energy range of 20 kcal/mol.

Computational Resources. The CPU time required for the conformational model generation with various settings was compared by analyzing the calculation performance of 10 ligands on an Octane 1200 with 300 MHz and 1GB RAM. When comparing 50 FAST to 255 BEST, about 230 times more computing power is needed to lower the average RMS by 12% (Figure 9). In economic considerations, this high demand of computing time is a important aspect. In contrast to the FAST method, there is a rapid increase in calculation time with BEST when generating more conformers. Because of these results, 255 FAST appears to have reasonable cost-value ratio.

Examination of Energetically Extreme Value Ligands. For the study of ligands exhibiting extreme energy values, many characteristics of the ligands and the proteins had to be analyzed. Again, LigandScout was used to extract this information. 40 high energetic (energy 10−20 kcal/mol, average energy 13.57 kcal/mol) and 38 low energetic (energy 0−4 kcal/mol, average energy 1.36 kcal/mol) ligands of the sample set were analyzed regarding their molecular interactions with the surrounding protein. The number of chemical features such as H-bond donor, H-bond acceptor, positive ionizable, negative ionizable and hydrophobic, the protein classification by PDB entry 'classification', and the localization inside the protein or on its surface were investigated.

First, the correlation between the RMS and the energy attracts attention: most of the time extremely high energy means also high RMS. However, no connection between the protein classification and the high- or low-energy levels could be found. Analysis of the protein−ligand complex shows that there is a trend to higher energetic ligands buried inside the proteins; lower energetic ligands tend to be on the surface of the protein. Frequently, high energy ligands exhibit several hydrophobic interactions with binding sites (e.g. tight pockets in the center of the protein interacting with alkyl chains).

Naturally, the number of interactions with the protein is depending on many characteristics of the ligand, most of all on the surface area of the molecule, which is in part reflected by the molecular weight. The ratio of the molecular weight of a ligand and the number of interactions confirms suggestions that higher energetic conformers are stabilized by more interactions with the binding site of the protein. For exact
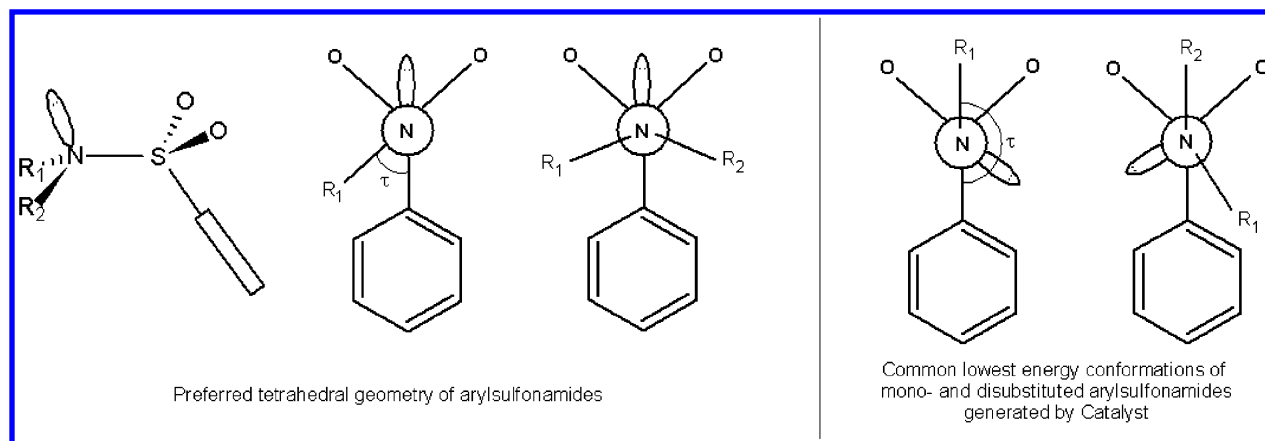
PROTEIN-BOUND LIGAND CONFORMATIONS

J. Chem. Inf. Model., Vol. 45, No. 2, 2005 **429**

Figure 8. Conformation of arylsulfonamides.

Table 3. RMS and Energy Values Depending on Energy Threshold Settings

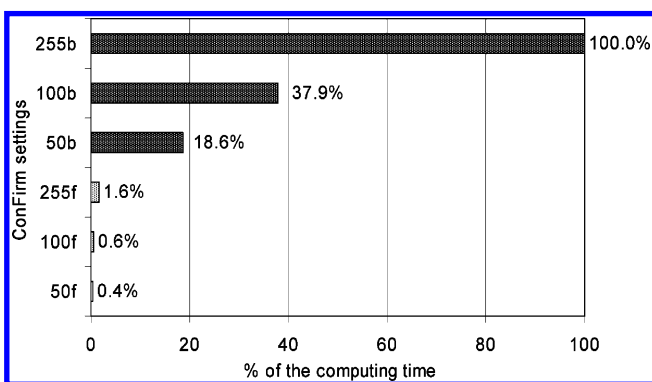| energy threshold setting [kcal/mol]	average no. of conformers	average RMS	average energy [kcal/mol]
10	52	1.203	4.21
15	84	1.099	6.02
20	112	1.057	7.27
30	138	1.057	8.87

Figure 9. Computing time ratio for the six different methods.

evidence of a correlation between the energy and the number of interactions the relative intensity of the interactions would have to be quantified.

Furthermore, the number of rotatable bonds represents a very important parameter. Though having comparable molecular weight in both groups, ligands with higher energy possess more rotatable bonds than the 'low energetic' ligands. Therefore, low energetic ligands tend to be more rigid.

Conformational Model Generation with Ligands Found in Both the PDB and the CSD. The aim of this analysis was to test if there are any differences in Catalyst's prediction between the bioactive conformation of a compound within a protein binding site and the conformation of the pure compound (sometimes cocrystallized with small molecules such as water) in the crystal state.

For this purpose, the PDB and the CSD were checked for identical entries via export and conversion of the compounds from both databases into canonical SMILES strings, which were then compared for identifying identical entries. The search algorithm developed by Inte:Ligand[16] was able to retrieve 450 molecules that are present in both databases. The detected compounds were visually inspected for drug-likeness and chemical diversity, and thus 15 of these ligands

Table 4. Comparison of Generated Conformers Fitting to the PDB and CSD Conformations (255 BEST)

	PDB	CSD
average RMS	0.83	0.81
average energy [kcal/mol]	3.74	3.56

were selected for the investigation. The models calculated with the 255 BEST method were compared to the conformers stored in both databases, and the resulting RMS values were used for the evaluation (Table 4).

The results show that there are only minor differences of Catalyst's fitting quality between the CSD and the PDB (Table 4). Conformers fitting to the CSD conformers have a 4.8% lower average energy and a 2.4% lower average RMS than calculated conformers fitting to the PDB conformers. In some cases, single values of both energy and RMS are even higher for CSD conformations.

Thus, these results suggest that Catalyst's CHARMm force field is able to predict bioactive conformations and conformations predominating in the crystal phase at virtually the same level of quality.

CONCLUSIONS

The aim of this large scale study was to evaluate Catalyst's conformational model generation with a large sample of 510 ligands and to develop a reliable protocol for the modeler on how to use Catalyst's ConFirm settings. The results show that the CHARMm force field implemented within Catalyst is a powerful tool which is able to produce high quality conformers that most of the times are well suited for in silico drug research. The RMS and energy values of the best fitting conformer are depending on the features of the calculated ligand and on Catalyst's ConFirm settings. Mainly the number of rotatable bonds, the molecular weight, and the chemical features are ligand-specific attributes affecting the quality of the fitting results. The average energy is depending on the molecular weight of the ligand. In contrast to the heaviest fraction (molecular weight > 500), which reaches an average energy of 9 kcal/mol, light molecules (molecular weight < 149) have a three times lower average energy level than heavy ligands. The RMS value rises by 400% from 0.4 to 1.6 between the lightest and the heaviest fraction of molecules in the sample. Both energy and RMS values increase by generating conformational models of ligands with more rotatable bonds.

430 *J. Chem. Inf. Model., Vol. 45, No. 2, 2005*

KIRCHMAIR ET AL.

Catalyst-specific settings were used for the conformational model generation (FAST vs BEST), the maximum number of generated conformers, and the energy threshold within which the conformers should be placed. By setting the first two options for higher fitting quality, the computing time rises enormously (factor 230), while the RMS is improved from RMS = 1.06 with 50 FAST generated conformers to RMS = 0.93 with 255 BEST generated conformers, which represents a reduction of 12%. Every method provides best fitting conformers with RMS < 1.50 in more than 80% of all cases and with RMS < 2 in more than 93%. FAST conformer model generation with maximal 100 conformers should be appropriate most of the time. Since the CPU time does not rise significantly and the fitting quality can be further improved when generating 255 conformers, 255 FAST appears to have reasonable cost-value ratio. RMS values beyond 3.0, which are not acceptable for scientific research, were found in less than 0.6% of the investigated cases. However, FAST is not recommended for conformational model generation of large, flexible ligands. Ring systems that have more than eight members or flexible molecules with more than 12 rotatable bonds will produce low quality conformational models that do not provide representative coverage of conformational space; here the BEST method is recommended. The energy of the best fitting conformer is higher with the BEST method than with FAST because ConFirm reduces the RMS at the cost of raising energy. The average energy with both FAST and BEST lies between 4.6 and 6.5 kcal/mol. Because of higher generation expenditure, while producing no or little effect, the combination of the FAST and the BEST method is not appropriate to lower the RMS value.

It is more difficult for ConFirm to fit sulfonamides; they have significantly higher RMS (34−45%) and energy (4−42%) values. Within an energy range of 20 kcal/mol, Catalyst is able to generate conformations that are fitting to the bioactive conformation, but its force field penalizes this conformation with too high energy by assuming a wrong torsional angle.

20 kcal/mol is the best setting for the ConFirm force field energy range. It is not recommended to lower the adjustment because this would raise the RMS value, giving conformations which are less similar to the bioactive conformation. On the other hand, a higher setting for the energy range only raises the energy value of the best fitting conformers but does not improve the fitting quality.

Since there is evidence that conformers of the same molecule bound to proteins and in crystal phase may vary from each other, we investigated a small sample of ligands showing that there are only minor or no significant differences of Catalyst's fitting quality between CSD and PDB compounds.[26−28]

ACKNOWLEDGMENT

Supporting Information Available: List of PDB reference codes of all 510 ligands. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(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) Poulsen, A. An Evaluation of Catalyst's Conformational Search Algorithm with Regard to Conformational Diversity and Conformational Energy Penalties. *6th European Catalyst User Group Meeting* 2002, Klampenborg-Copenhagen, Denmark.

(4) Vieth, M.; Hirst, J. D.; Brooks, C. L. Do Active Site Conformations of Small Ligands Correspond to Low Free-Energy Solution Structures? *J. Comput.-Aided Mol. Des.* **1998**, *12*, 563−572.

(5) Boström, J. Reproducing the Conformations of Protein-Bound Ligands: A Critical Evaluation of Several Popular Conformational Searching Tools. *J. Comput-Aided. Mol. Des.* **2001**, *15*, 1137−1152.

(6) Perola, E.; Charifson, P. S. Conformational Analysis of Drug-like Molecules Bound to Proteins: An Extensive Study of Ligand Reorganization Upon Binding. *J. Med. Chem.* **2004**, *47*, 2499−2510.

(7) Koshland, D. E. Application of a Theory of Enzyme Specifity to Protein Synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98−105.

(8) 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.

(9) Koshland, D. E. Correlation of Structures and Function in Enzyme Action. *Science* **1963**, *142*, 1533−1541.

(10) Koshland, D. E., Jr. Conformation changes at the active site during enzyme action. *Fed. Proc.* **1964**, *23*, 719−726.

(11) Yankeelov, J. A., Jr.; Koshland, D. E., Jr. Evidence for Conformation Changes Induced by Substrates of Phosphoglucomutase. *J. Biol. Chem.* **1965**, *240*, 1593−1602.

(12) Höltje, H. D.; Folkers, G. *Molecular Modeling: Basic Principles and Applications*; VCH Verlagsgesellschaft mbH: Weinheim, 1996.

(13) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.

(14) Accelrys, San Diego, CA, U.S.A. http://www.accelrys.com.

(15) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B* **2002**, *58*, 380−388.

(16) Inte:Ligand Software-Entwicklungs und Consulting GmbH, A-2344 Maria Enzersdorf, Austria http://www.inteligand.com/technology.

(17) Wolber, G. LigandScout − Pharmacophores Derived From Protein Bound Ligands in the Protein Data Bank, Ph.D. Thesis, University of Innsbruck, 2003.

(18) Langer, T.; Wolber, G.; Laggner, C.; Kirchmair, J. The Problem of Bio-active Conformations: What Can We Learn From the PDB Database? *7th European Catalyst User Group Meeting* 2003, Astra-Zeneca, Charnwood, U.K. and *US Catalyst User Group Meeting* 2004, Emisphere Technologies, Tarrytown, NY.

(19) Accelrys, San Diego, CA, Catalyst ConFirm Help 4.8, 2003.

(20) Smellie, A.; Kahn, S.; Teig, S. Analysis of Conformational Coverage. 1. Validation and Estimation of Coverage. *J. Chem. Inf. Comput. Sci.* **1995**, *35* (2), 285−294.

(21) Smellie, A.; Kahn, S.; Teig, S. Analysis of Conformational Coverage. 2. Applications of Conformational Models. *J. Chem. Inf. Comput. Sci.* **1995**, *35* (2), 295−304.

(22) Accelrys, San Diego, CA, Catalyst Citest Help 4.8, 2003.

(23) Accelrys, San Diego, CA, Cerius^2 Help 4.5, 2000.

(24) Dorie, J.; Gouesnard, J. P. Nitrogen-15 NMR Study of Nitrogen−sulfur Bond in Sulfenamides, Sulfinamides and Sulfonamides. *J. Chim. Phys. Phys.-Chim. Biol.* **1984**, *81* (1), 15−19.

(25) Beddoes, R. L.; Kettle, J. G.; Joule, J. A. A 3-Phenylsulfinyl-1-phenylsulfonylindole. *Acta Crystallogr., Sect. C* **1994**, *C50*, 1989−92.

(26) Rybarczyk-Pirek, J.; Zgierski, Z. Intermolecular Forces and Conformational Change Upon Crystallization − The Case of Phosphorobenzopyrane Derivatives. *J. Chem. Phys.* **2001**, *115* (20), 9346−9351.

(27) Allen, F. H.; Harris, S. E.; Taylor, R. Comparison of Conformer Distributions in the Crystalline State with Conformational Energies Calculated by Ab Initio Techniques. *J. Comput.-Aided Mol. Des.* **1996**, *10* (3), 247−254.

(28) Janssen, L. H. M. Conformational Flexibility and Receptor Interaction. *Bioorg. Med. Chem.* **1998**, *6* (6), 785−788.

CI049753L