Journal of Computer-Aided Molecular Design, **16:** 105–112, 2002.
KLUWER/ESCOM

Can we separate active from inactive conformations?

David J. Diller* & Kenneth M. Merz, Jr.**
Department of Molecular Modeling, Pharmacopeia, Inc., CN5350, Princeton, NJ 08543-5350, USA

Received 18 October 2001; accepted 28 February 2002

Key words: bound conformation, conformational analysis, surface area, three-dimensional descriptors

Summary

Molecular modeling methodologies such as molecular docking, pharmacophore modeling, and 3D-QSAR, rely on conformational searches of small molecules as a starting point. All of these methodologies seek conformations of the small molecules as they bind to target proteins, i.e., their active conformations. Thus the question as to whether active conformations can be separated from inactive conformations is extremely relevant. In this paper, 3D-descriptors that separate random conformations from active conformations of small molecules are sought. To select appropriate descriptors, 65 protein-ligand complexes were taken from the protein data bank. For each ligand the active conformation was compared to randomly generated low energy conformations. Descriptors such as solvent accessible surface area, number of internal interactions and radius of gyration appear to be useful for separating the active conformations from the random conformations. The results with all these descriptors indicate that active conformations are less compact that random conformations, i.e., they have more solvent accessible surface area, fewer internal interactions and a larger radius of gyration than random conformations. Thus these descriptors could be useful as weights to bias conformational search procedures to conformations more likely to bind to proteins or as filters to eliminate conformations unlikely to bind to any protein.

Introduction

Many molecular modeling methodologies rely on conformational search as a starting point. Many of these methods, such as molecular docking [1–4], pharmacophore searches [5, 6], and 3D-QSAR [7], need to find bioactive conformations, i.e., the conformations of the small molecules as they bind to a given protein. Thus these methods would benefit greatly from models that could bias conformational searches and conformational databases to conformations that are more likely to be bioactive.

There is some evidence that molecules bind to proteins in low strain conformations. Several groups [8–11] have investigated the strain of conformations of small molecules extracted from protein-ligand complexes deposited in the Protein Data Bank [12, 13] (PDB). Initial studies found that the bound conformations were in fact highly strained[9] with estimates of their strain ranging from 5–40 kcal/mol. These studies, however, calculated the strain using CHARMm in vacuum and did not consider the potential for some coordinate error in the structures. Bostrom et al. [10, 11] showed that by using a solvation correction the strain estimates dropped significantly. Furthermore they found a case were the actual conformation of the ligand in the deposited structure was highly strained, but showed that there were several conformations within the error of the structure that were of low strain. Finally, they showed that the force field used in the calculation could have a dramatic effect on the calculated strain of the conformation. There are several potential conclusions from these series of studies. First, as methods improve the estimates of the strain of the bound conformations have decreased significantly: the majority of the bound conformations appear to have a strain below 3–4 kcal/mol [11]. A second conclusion is that given the coordinate errors, force fields are

*To whom correspondence should be addressed. E-mail: ddiller@pharmacop.com
**Permanent address: 152 Davey Laboratory, Department of Chemistry, Penn State University, University Park, PA 16802, USA.

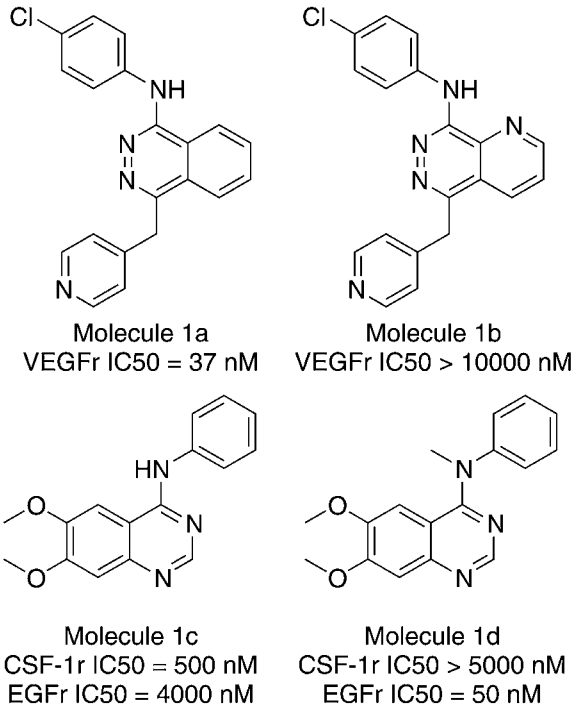


Figure 1. Two examples demonstrating the importance of conformation is structure activity relationships.

still too sensitive to be used to effectively estimate the strain of the conformations of small molecules as they are deposited in the PDB.

A second piece of evidence for the belief that small molecules bind to proteins in low energy conformations is the development of the empirically derived scoring functions used for estimating the binding constant of protein ligand complexes [14–18]. None of these models has a term to account for the strain of the ligand but all of these models achieve a fit to the experimental binding constants to within 1–1.5 kcal/mol root mean square error. It is safe to assume that some of these ligands bind with little or no strain. Thus unless all these scoring functions are significantly biased, strain should account for no more than 3–4 kcal/mol in the majority of the protein-ligand complexes used to train these scoring functions.

On the other hand, examination of any structure activity data set will indicate that conformation can make the difference between a tightly bound inhibitor and a weakly bound inhibitor. As a first example consider molecule 1a [19] (see Figure 1). This molecule binds to the vascular endothelial growth factor receptor (VEGFr) with an IC50 of 37 nm. When the C8 atom is changed to a nitrogen (molecule 1b) the compound becomes inactive (IC50 $>$ 10 000 nm) against VEGFr. Solvation effects might account for part of the change but certainly not the entire 5 kcal/mol. The largest difference between the two molecules is that molecule 1b has the potential for an internal hydrogen bond between the amino NH and N8 whereas molecule 1a does not. This hydrogen bond might lock the 4-Cl-phenyl-amino of molecule 1b in an unfavorable conformation and thus prevent this compound from adopting its VEGFr-active conformation. Quantum mechanical calculations using Mopac 6 [20] show that the clearly preferred conformation of molecule 1b has

Table 1.

PDB code	No.	Rot. bonds	PDB code	No.	Rot. bonds
1cbx	1	5	1hfc	34	9
1cps	2	5	1ppc	35	9
1hyt	3	5	1glp	36	10
1lst	4	5	1bbp	37	11
1mcr	5	5	1hpv	38	11
1stp	6	5	1glq	39	12
2cmd	7	5	1htf	40	12
2sim	8	5	4phv	41	12
3cpa	9	5	1tmn	42	13
1acm	10	6	1b5h	43	14
1frp	11	6	1b0h	44	15
1lmo	12	6	1b3h	45	15
1snc	13	6	1b4h	46	15
2cgr	14	6	1b6h	47	15
3tmn	15	6	1epo	48	15
3tpi	16	6	1icn	49	15
1apu	17	7	1ida	50	15
1ett	18	7	1lic	51	15
1hri	19	7	2plv	52	15
1pgp	20	7	1aaq	53	16
1pph	21	7	1b1h	54	16
2r07	22	7	1b2h	55	16
8gch	23	7	1b7h	56	16
1atl	24	8	1apt	57	17
1hvr	25	8	1eed	58	18
1lna	26	8	1hef	59	18
1tka	27	8	1hvi	60	19
4dfr	28	8	5p2p	61	20
1dwd	29	9	1rne	62	21
1eap	30	9	1lyb	63	22
1etr	31	9	1sme	64	22
1ets	32	9	1poc	65	23
1fkg	33	9			

the amino-chlorophenyl in the plane of the central bicyclic core as shown in Figure 1. Due in part to the loss off the internal hydrogen bond and in part to the resulting steric clash between the amino proton and the extra C8 proton (~ 1.8 Å) the planar arrangement for the amino-chlorophenyl is not feasible for molecule 1a. The closest allowed conformation appears to be with the NH proton approximately 30° out of the plane of the central bicyclic core. Even this conformation is approximately 3 kcal/mol above the global minimum energy conformation for molecule 1a. The net result is that molecule 1a has much more conformational freedom than does 1b.

As a second example, consider molecules 1c and 1d [21] (see Figure 1). Against colony stimulating factor-1 receptor (CSF-1r), the IC50 of molecule 1c is 500 nM while molecule 1d is inactive (IC50 $> 50\,000$). The situation for these two molecules is nearly reversed for the epidermal growth factor receptor (EGFr) where the IC50 of compound 1c is 4000 nM compared to 50 nM for compound 1d. The only difference between these two compounds is that the amino is methylated in compound 1d. This methyl could have multiple effects including changing the hydrogen bonding pattern with the protein and altering the conformational preference of the molecule. Bold et al. [19] showed using NMR that this methyl drastically changes the conformational behavior of this molecule in solution which likely produces a significant portion of the change in the activity.

These studies and examples show both that strain is an important factor in determining binding affinity and that our understanding of strain is not yet sufficient to develop a useful model. In lieu of developing a comprehensive strain model, we address the question: 'Can we separate active conformations from random conformations?' The goal is to develop filters that can be used to eliminate conformations as being unlikely to bind tightly to any protein or to develop simple descriptors that can be used to bias conformational search procedures to conformations more likely to be bioactive.

To develop such filters we assemble a small set of active conformations of small molecules extracted from cocrystal complexes deposited in the PDB and consider a number of three-dimensional descriptors to see which of these descriptors best separates the active conformations from random conformations. The descriptors used in this study include polar solvent accessible surface area, apolar solvent accessible surface area, radius of gyration, the number of internal interactions, the ratio of the two principal axes, and the magnitude of the dipole moment. These descriptors were chosen as they are relatively insensitive to small changes in conformation and thus to errors in ligand conformations found in crystal structures. In particular, the calculated conformational force field energy was not used because it is too sensitive to small changes in conformation. This study shows that polar solvent accessible surface area, apolar solvent accessible surface area, radius of gyration, and the number of internal interactions can all be used to separate active from random conformations. These descriptors separate active from random conformations even better for highly flexible molecules. The results with these four descriptors indicate that active conformations are on average significantly less compact than random conformations.

Methods

First, the conformation of 65 small molecules as they are bound to a protein were extracted from the PDB. No molecules with macrocyclic rings or rigid compounds were considered. The molecules have between 5 and 23 rotatable bonds (see Table 1). This data set is not ideal as it contains series of related compounds. For example, there are several aspartyl protease inhibitors and several inhibitors of trypsin.

For each molecule, random conformations were generated in the following manner. The dihedral angles were chosen uniformly at random with bond lengths, bond angles, and rings being held fixed. The conformation was then minimized in dihedral space using just a van der Waals term and a dihedral angle term [22]. Typically, a conformation would minimize to a reasonable energy or get trapped in a very high energy well such as having a bond run through a phenyl ring. With this in mind, any conformation with an extremely high energy (> 100 kcal/mol) after the minimization was discarded. This process was continued until 5000 random conformations were generated for each molecule.

For each molecule, M, and each descriptor, D, the following quantities can be calculated. First is the value of D for the active conformation, a(D,M). Second is the mean value for the descriptor over the 5000 random conformations of the molecule M, μ(D,M). The third quantity is the standard deviation of the descriptor D over the 5000 random conformations of molecule M, σ(D,M). The final quantity of interest

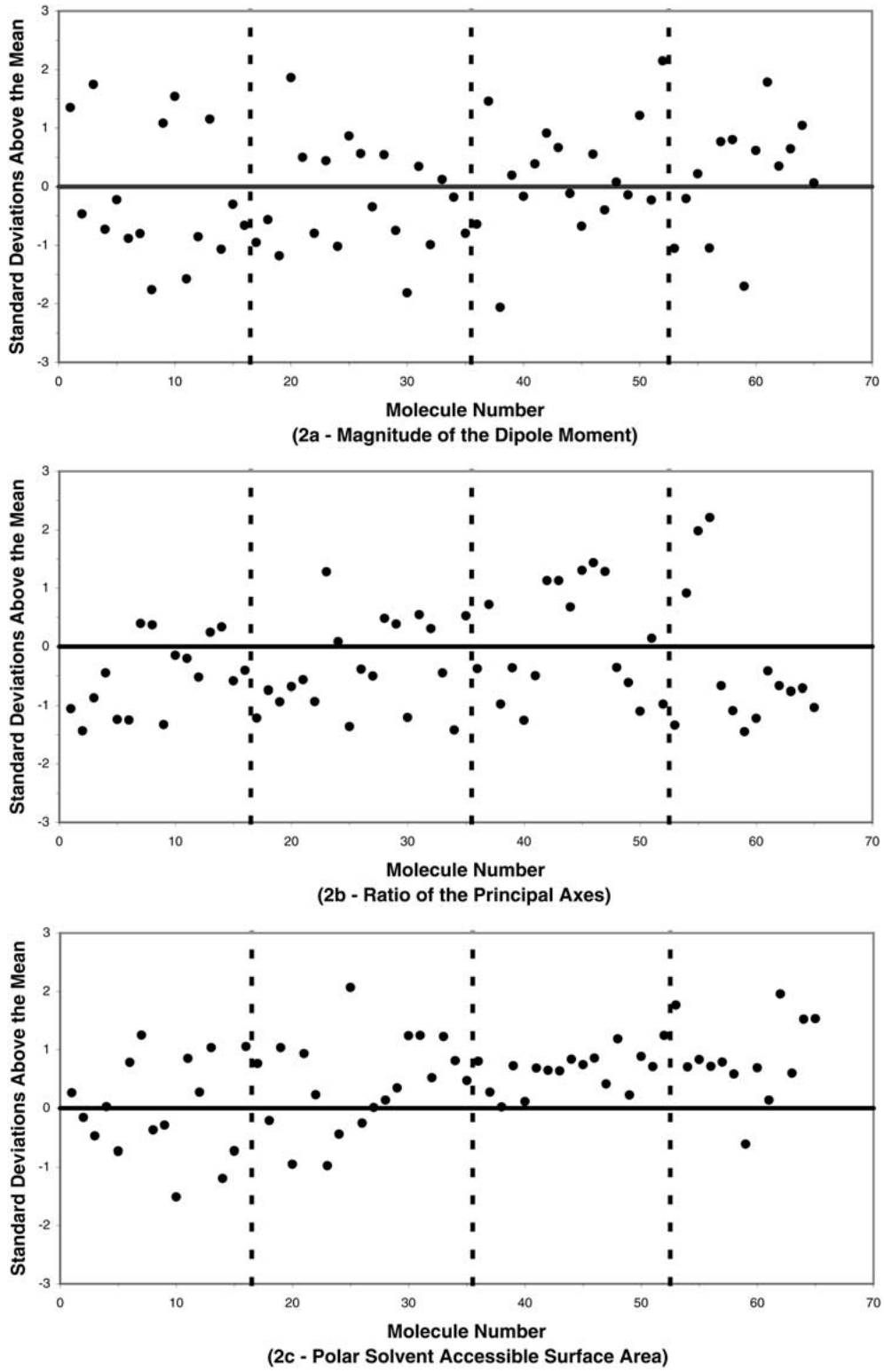
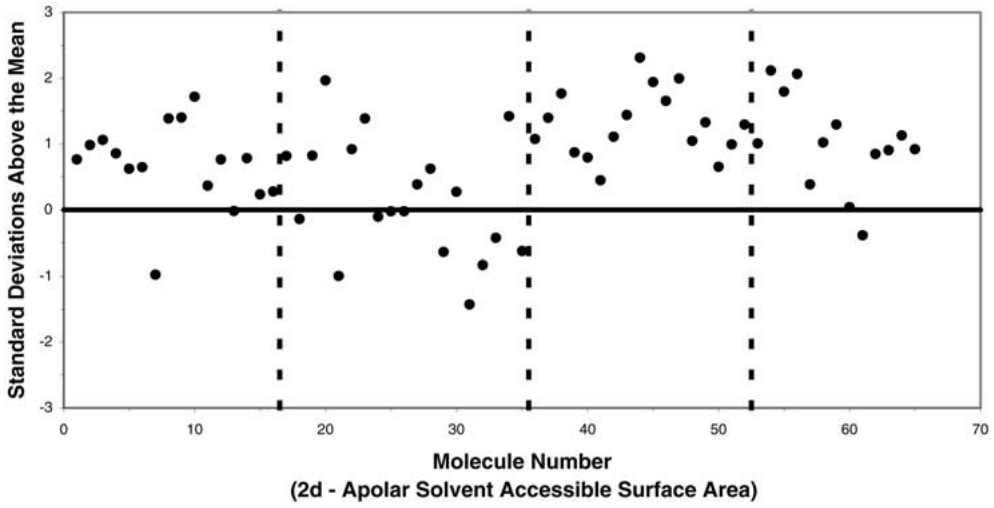
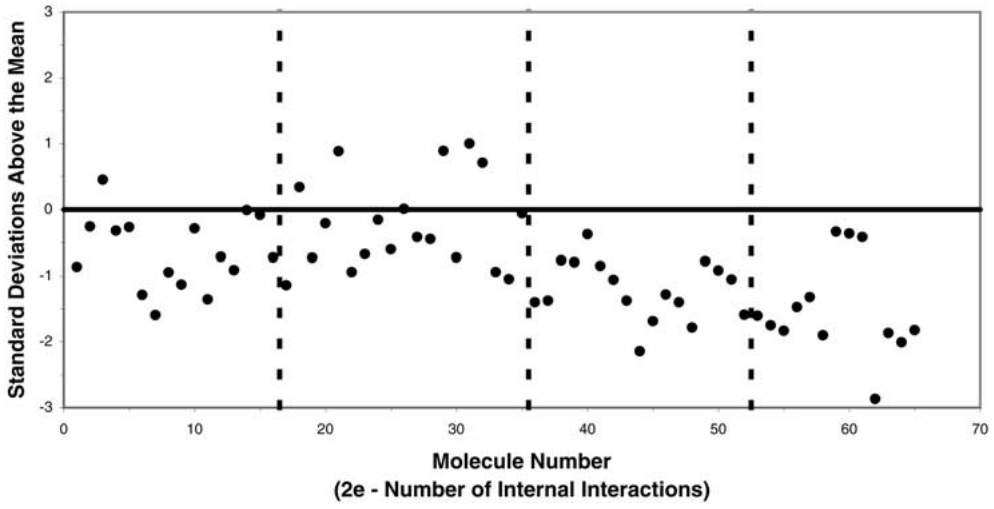


Figure 2. The mean centered and scaled values for the three dimensional descriptors of the active conformations. In all cases, the y-axis is the adjusted value (see Equation 1) for the three dimensional descriptor and the x-axis is the molecule number. The molecules are numbered from least flexible to most flexible (see table 1). The molecules to the left of the first dashed line have either 5 or 6 rotatable bonds. The molecules between the first and second dashed lines have 7 to 9 rotatable bonds. The molecules between the second and third dashed lines have 10 to 15 rotatable bonds. The molecules to the right of the third dashed line have more than 15 rotatable bonds. (a) Magnitude of the dipole moment. (b) Ratio of the principal axes. (c) Polar solvent accessible surface area. (d) Apolar solvent accessible surface area. (e) Number of internal interactions. (f) Radius of gyration.





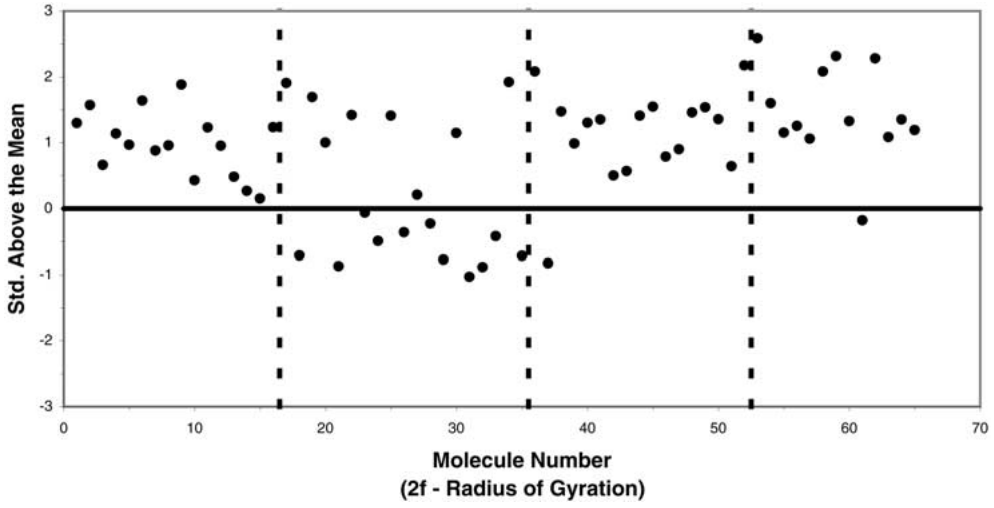


Figure 2. Continued.

is the adjusted value of the descriptor for the active conformation, which is given by

$$A(D,M) = \frac{a(D, M) - \mu(D, M)}{\sigma(D, M)}. \quad (1)$$

If the active conformations are indistinguishable from random conformations the values of adjusted descriptors should be evenly distributed about 0 over the molecules in the data set.

The following descriptors were used in this study: polar solvent accessible surface area (PSASA), apolar solvent accessible surface area (ASASA), number of internal interactions (NI), radius of gyration (RG), ratio of the two principal axes (RPA), and the magnitude of the dipole moment (MDM). The solvent accessible surface areas [23] were calculated using the van der Waals radii of the atoms plus 1.4 Å. No hydrogen atoms were used in the calculation. A nitrogen or oxygen was treated as polar if it had a hydrogen or it had a lone pair capable of accepting a hydrogen bond. All other atoms were treated as apolar. The quantity NI is a simple count of the number of pairwise interactions in a given molecule. It is given by

$$NI(C) = \sum_{i<j} f(d_{ij}) \quad (2)$$

where the sum is over all pairs of atoms i,j except for 1–2 and 1–3 atoms, d_{ij} is the distance between the ith and jth atoms, and

$$f(r) = \begin{cases} 1 & \text{if } r < 4.0 \\ 0.5(6.0 - r & 4.0 < r < 6.0 \\ 0 & r > 6.0 \end{cases} \quad (3)$$

with all the units in Å. The radius of gyration of a conformation is the root mean square of the distance between the atoms and the center of mass of the conformation. The ratio of the principal axes is the second largest eigenvalue of the covariance matrix of the atomic coordinates of the conformation divided by the largest eigenvalue. A value of RPA near 0 indicates a long extended conformation whereas a value near 1 indicates a round and compact conformation. Finally, the dipole moment was calculated using atom point charges calculated using the methods of Rappe and Goddard [24] available through Cerius2 [25].

Results

The adjusted values (see Equation 1) of each of the descriptors for the active conformations are plotted in Figure 2 versus the molecule number with the molecules being ordered by number of rotatable bonds. Since the adjusted values are evenly distributed about zero, the magnitude of the dipole moment (see Figure 2a) and the ratio of the principal axes (see Figure 2b) do not appear to separate the active conformation from the random conformations. The remaining four descriptors, PSASA (see Figure 2c), ASASA (see Figure 2d), NI (see Figure 2e), and RG (see Figure 2f), do appear to be useful for separating the active conformations from the random conformations particularly for the large and flexible molecules. These four descriptors are discussed in some detail below.

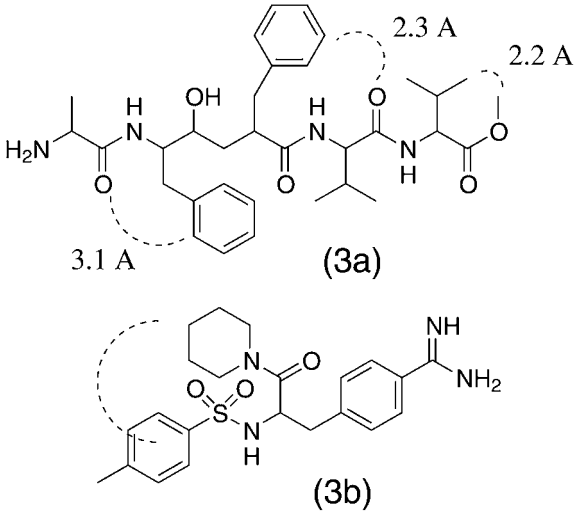


Figure 3. Two examples of outliers. (a) The case 1hef. The dashed lines indicate the steric clashes in the crystallographically observed bound conformation. (b) The case 1pph. The dashed line indicates the rings that are packed together in the bound conformation.

Of the 65 molecules only 14 have an active conformation with an adjusted PSASA below zero and of the 37 molecules with more than eight rotatable bonds, only one has an active conformation with an adjusted PSASA below zero. Thus bioactive conformations appear to have on average more PSASA than random conformations. In this respect, active conformations are similar to solution conformations. In addition, the case 1hef, which is the only case with more than eight rotatable bonds and an adjusted PSASA below zero, appears to have some problems. The conformation exhibits several serious internal clashes (see Figure 3a) including a carbonyl oxygen clashing with a phenyl ring (C-O distance ~ 2.3 Å). This clash accounts in part for the lower than average PSASA. This molecule also makes some undesirable contacts with the pro-

tein and appears to have a more reasonable alternate binding mode [3].

Only 10 of the 65 cases have an active conformation with an adjusted ASASA below zero. This result might seem surprising. One would expect that in solution the low energy conformations would be those that buried as much apolar surface area as possible. Unlike water, however, a protein will compete effectively for both the apolar and polar interactions. The cases with an active conformation with a negative adjusted ASASA are primarily those that have two large hydrophobic groups capable of interacting. Many of these are inhibitors of trypsin that contain an aromatic ring and a piperidine that pack against one another (see Figure 3b). Thus while this result indicates that a protein can compete effectively for the apolar interactions, there are circumstances when the intramolecular apolar interactions are sufficiently strong to be maintained upon binding to a protein.

The number of internal interactions is the descriptor that best separates the active from the random conformations. In this case, only 5 of the active conformations have a positive adjusted NI indicating that the active conformations have far fewer internal interactions than random conformations. The outliers are primarily the trypsin inhibitors discussed in the previous paragraph.

The final descriptor that has some potential for separating the active from the random conformations is the radius of gyration. In this case, 13 of the 65 cases have an active conformation with a negative adjusted RG indicating that the radii of gyration of the active conformations are greater than those for the random conformations. Again the outliers are similar to those for the apolar solvent accessible surface area.

Conclusions

The conformations of small molecules as they bind to proteins can be separated from random conformations using a variety of descriptors. These descriptors include the polar solvent accessible surface area, the apolar solvent accessible surface area, the number of internal interactions and the radius of gyration. Not all conformationally dependent descriptors are useful for separating active from inactive conformations. Neither the magnitude of the dipole moment nor the ratio of the two principal axes appears to be useful for this purpose.

Active conformations have on average more polar and apolar solvent accessible surface area, fewer internal interactions and a larger radius of gyration than random conformations. These results indicate that on average active conformations are less compact than random conformations. These descriptors would be useful weights for biasing conformational search procedures to include fewer compact conformations thereby improving the results of modeling techniques, such as pharmacophore searching, molecular docking, and 3D-QSAR.

While for most cases these descriptor work to separate active from random conformations, there are still some pathological misclassifications: the most notable being the inhibitors of trypsin. The most apparent way to circumvent these pathological misclassifications would be to combine these three-dimensional descriptors into a higher dimensional classification model. Clearly, more three-dimensional descriptors could also be included. Beyond three-dimensional descriptors, two-dimensional descriptors could also be included to flag certain moieties that are likely to cause pathological misclassifications. To do this more data will be necessary and this will be a topic of future study.

References

1. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
2. Kuntz, I.D., Science, 257 (1992) 1078.
3. Diller, D.J. and Merz, K.M., Proteins Struct. Funct. Genet., 43 (2001) 113.
4. Kearsley, S.K., Underwood, D.J., Sheridan, R.P. and Miller, M.D., J. Comput. Aid. Mol. Des., 8 (1994) 565.
5. Catalyst4.5., Molecular Simulations, I., San Diego, CA.
6. Greene, J., Kahn, S., Savoj, H., Sprague, P. and Teig, S., J. Chem. Inf. Comput. Sci., 34 (1994) 1297.
7. Cramer, R.D., III, Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
8. Vieth, M., Hirst, J.D. and Brooks, C.L., III, J. Comput. Aid. Mol. Des., 12 (1998) 563.
9. Nicklaus, M.C., Wang, S., Driscoll, J.S. and Milne, G.W.A., Bioorg. Med. Chem., 3 (1995) 411.
10. Liljefors, T., Bostrom, J. and Norrby, P.-O., Conformational energies of protein-bound ligands. In Rational Molecular Design in Drug Research, Alfred Benzon Symp. 42, 1998, Munksgaard, Copenhagen.
11. Bostrom, J., Norrby, P.-O. and Liljefors, T., J. Comput. Aid. Mol. Des., 12 (1998) 383.
12. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., Nucleic Acids Res., 28 (2000) 235.
13. Berman, H.M., Bhat, T.N., Bourne, P.E., Feng, Z., Gilliland, G., Weissig, H. and Westbrook, J., Nat. Struct. Biol., 7 (2000) 957.

14. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J., J. Comp. Chem., 19 (1998) 1639.
15. Bohm, H.-J., J. Comput. Aid. Mol. Des., 8 (1994) 243.
16. Bohm, H.-J., J. Comput. Aid. Mol. Des., 12 (1998) 309.
17. Baxter, C.A., Murray, C.W., Clark, D.E., Westhead, D.R. and Eldridge, M.D., Proteins: Struct. Funct. Genet., 33 (1998) 367.
18. Jain, A.N., J. Comput. Aid. Mol. Des., 10 (1996) 427.
19. Bold, G., Altmann, K.-H., Frei, J., Lang, M., Manley, P.W., Traxler, P., Wietfeld, B., Brueggen, J., Buchdunger, E., Cozens, R., Ferrari, S., Furet, P., Hofmann, F., Martiny-Baron, G., Mestan, J., Roesel, J., Sills, M., Stover, D., Acemoglu, F., Boss, E., Emmenegger, R., Laesser, L., Masso, E., Roth, R., Schlachter, C., Vetterli, W., Wyss, D. and Wood, J.M., J. Med. Chem., 43 (2000) 2310.
20. Stewart, J.J., J Comput. Aid. Mol. Des., 4 (1990) 1.
21. Myers, M.R., Setzer, N.N., Spada, A.P., Persons, P.E., Ly, C.Q., Maguire, M.P., Zulli, A.L., Cheney, D.L., Zilberstein, A., Johnson, S.E., Franks, C.F. and Mitchell, K.J., Bioorg. Med. Chem. Lett., 7 (1997) 421.
22. Diller, D.J. and Verlinde, C.L.M.J., J. Comp. Chem., 20 (1999) 1740.
23. Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
24. Rappe, A.K. and Goddard, W.A., III, J. Phys. Chem., 95 (1991) 3358.
25. Cerius2, Molecular Simulations, Inc, San Diego, CA.