CONFORMATIONAL ANALYSIS, SOLVENT-ACCESSIBLE SURFACE AND GEOMETRIC EXTENT OF INHIBITORS AND SUBSTRATES

Matheus FROEYEN[a1], Hans DE WINTER[b] and Piet HERDEWIJN[a2,*]

[a] *Laboratory of Medicinal Chemistry, REGA Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium; e-mail: [1] matheus.froeyen@rega.kuleuven.ac.be, [2] piet.herdewijn@rega.kuleuven.ac.be*

[b] *Silicos NV, Wetenschapspark 7, B-3590 Diepenbeek, Belgium; e-mail: hans.dewinter@silicos.com*

Received March 17, 2005
Accepted April 26, 2006

Dedicated to Professor Antonín Holý on the occasion of his 70th birthday in recognition of his outstanding contributions to the area of nucleic acid chemistry.

Peptide ligands are known to often bind in an extended conformation to maximize the contact surface with the receptor. The complementarity of shape and surface is a very important factor contributing to the stability of a ligand receptor complex. In this communication we try to answer the following questions: What is the influence of conformation on the surface area? Do ligands maximize their contact surface with the receptor to increase stability of the complex?

Keywords: Solvent accessible surface; Enzymes; Active sites; Peptides; Receptors; Conformation; Monte Carlo calculations; Hydrogen bonds; Hydrophobic interactions.

For many small medicinal molecules the three-dimensional structure of the drug target is unknown. This is why the popularity of software tools for de novo design in the absence of information on receptor and predictive model of building of ligands are becoming popular in virtual screening[1,2].

Many attempts have been made to predict the binding conformation of ligands in the active site of enzymes. Major consideration when predicting the docked conformation of a molecule is optimization of the common physicochemical interaction parameters between the receptor and the drug – hydrogen bonding, electrostatic interactions, van der Waals interactions – while changing orientation and position and sometimes dihedral angles of the ligand[3–6].

Recently a number of articles have described the energetics and conformational reorganization of small molecules when binding proteins[7–9]. Small

Collect. Czech. Chem. Commun. 2006, Vol. 71, No. 6, pp. 842–858

doi:10.1135/cccc20060842

molecules per se do not bind in their lowest-energy conformation but they change their conformation in a way to optimize hydrogen bonding interactions with the protein[7] or hydrophobic interactions by extending their conformation[9]. It is generally recognized that the size of the contact surface between the ligand and receptor is very important for the stabilization of the ligand conformation in the binding pocket[10–14]. For instance, peptide ligands are known to bind often in extended conformations[15]. In this communication we report on our efforts to search for a correlation between solvent-accessible surface (or more precisely, the contact surface between the ligand and receptor) and the bound conformation of a ligand to a receptor.

This study originated from the following question. If we want to generate a conformational database of potential medicinal compounds, can we limit the number of conformations by some method? For instance, to keep only the conformations with a maximal solvent-accessible surface, assuming that molecules bind in their extended conformations.

EXPERIMENTAL

Experimental Data

Coordinates of the complexes were obtained from the Protein Data Bank (PDB)[16]. The receptor and ligand type of a first set of complexes are listed in Table I. The ligand structures are shown in Fig. 1. Water molecules were removed from the PDB file. Hydrogen atoms were removed if present. Counter-ions and other ligands were treated as being parts of the receptor. If more ligand conformations were present in the PDB file, the first conformation (not necessarily the most stable) was arbitrarily taken for the calculations. The ligand molecule was separated from the receptor to perform the calculations on the ligand in its bound conformation, with receptor omitted.

Conformational Search

A Monte Carlo conformational search within the modeling program Macromodel[17] was conducted by varying dihedral angles of the ligand randomly[18]. The maximum number of conformations was set to 5000. Structures were first energy-minimized in the Amber* force field[19]. A distance-dependent dielectric permittivity is used with an electrostatic cutoff distance of 12 Å. In the iterative conformational search the program applies random torsional rotations and performs energy minimization. The energy range for stored structures was set to 50 kJ/mol, which means that if the energy of every new conformation has fallen within 50 kJ/mol of the current global minimum, it is kept, otherwise rejected. The final conformation is compared with the set of structures already found, and stored if it is not a duplicate. The number of duplicates is also stored. The conformational search was performed twice: the first without solvent and the second with a water continuum as solvent[20]. The number of different conformations found is indicated in Table I.

TABLE I
Receptor ligand complexes selected from literature

PDB identification	Receptor type	Ligand type	Ligand	n^a	Number of conformations		Ref.
					vacuum[b]	H_2O[c]	
11ba	ribonuclease	nucleotide	uridylyl(2′5′) adenosine	39	715	2221	28
1brn[d]	ribonuclease	nucleotide	d(CGAC)	63	76	582	29
1dbj	antibody	antigen	steroid	21	11	11	30
1ep4	HIV RT	nnrti	S-1153	29	4760	4771	31
1epl	hydrolase	peptide	peptide of 6 amino acids	58	187	306	32
1fmn	RNA	aptamer	flavin mononucleotide	31	372	2040	33
1g3u	TMP kinase	nucleotide	thymidine monophosphate	21	404	500	34
1g3x[e]	DNA	peptide	acridine-peptide	48	375	1116	35
1hef[f]	protease	peptide	HEF	48	585	1386	36
1heg[f]	protease	peptide	HEG	41	669	1614	36
1jtl	DNA	minor groove binder	distamycin	35	4629	754	37
1koc	RNA	peptide	arginine	12	280	699	38
1nem[g]	RNA	aminoglycoside	neoB	44	1263	1390	39
1ola	protein (OppA)	peptide	VKPG	28	353	4934	40
1ppc	hydrolase	inhibitor	NAPAP	37	1763	3148	41
1qd3[g]	RNA (TAR)	aminoglycoside	NeoB	42	701	1088	42
1qs4	integrase	inhibitor	5ClTEP	20	29	23	43
1raw	RNA	nucleotide	AMP	23	400	1323	44
1rnb	ribonuclease	nucleotide	d(GC)	38	755	3248	45
261d	DNA	minor groove binder	netropsin	31	4485	202	46
2er6	hydrolase	peptide	H-256	65	112	269	47
2r04	coat protein	antiviral	w71	25	4790	4731	48
2tob	RNA	aminoglycoside	tobramycin	32	847	920	49
3cpa	hydrolase	peptide	Gly-Tyr	17	128	305	50
4sga	hydrolase	peptide	tetrapeptide	34	2262	3288	51
5tln	hydrolase	peptide	hydroxamic acid inhibitor	23	2837	3270	52
9icd	oxidoreductase	nucleotide	NADP	27	1252	1840	53

[a] Number of ligand non-hydrogen atoms. [b] The number of clusters found in the conformational search with the solvent set to vacuum. [c] The number of different conformations found in water solvent. [d] Incomplete ligand X-ray structure, 3 arginine side-chains missing. [e] Incomplete ligand X-ray structure, C residue missing. [f] Biological dimer of polymerase reconstructed. [g] The same ligand, but different conformation.

Solvent-Accessible Surface and Geometric Extent

Two methods were used to calculate the degree of extension of small molecules.

The first method calculates the solvent-accessible surface (SAS) as defined by Richards[21], for every conformation of all ligands using DMS[22]. A probe of 1.4 Å, rolling on the surface was used in the calculations. The SAS results were subdivided into following contributions: polar atoms P, N, O; hydrophobic atoms C, F, Br, Cl, I, and S[23].

The second method calculates the degree of extension (EXT) of small molecules as the sum of pairwise interatomic distances of bonded atoms in the molecule divided by the total number of bonds. For every conformation, one value defining the extent of the conformation is obtained[9].

For every receptor ligand complex, the calculated solvent-accessible surface values are plotted against their geometric extent value and the SAS (EXT) values for the X-ray conformation are indicated on the plot (Figs 2 and 3). Smooth curves were calculated using the local polynomial regression package Loess[24]. All results were plotted using Gnuplot[25].

RESULTS AND DISCUSSION

Solvent-Accessible Surface and Extendedness

For every conformation output by the conformational search on the ligand structures in Fig. 1, the polar and hydrophobic SAS is plotted against the geometric extent of that conformation for conformational searches performed in a vacuum environment (Fig. 2) and in a water environment (Fig. 3). As expected, in many cases we see a correlation between the solvent-accessible surface (SAS) and the geometric extent (EXT). Also, the reader can verify that several X-ray conformations (indicated by the X symbols in Figs 2 and 3) are often found at the top right of the graphs, confirming our hypothesis that the X-ray conformation has often a high SAS and EXT. However, this hypothesis does not always fit the reality and the relationship is not uniformly straightforward.

Solvent Environment in the Conformational Search Procedure

The parameters used in the conformational search procedure clearly influence the outcome of the search. In the vacuum simulation the number of conformational clusters found is smaller than in the water simulations (Table I). Exceptions are 1jtl, 261d and 2r04, all being long molecules. A first observation is that the conformational distribution (as a function of the extent) is in most cases broader in the water simulations (Fig. 3) than in the vacuum simulations (Fig. 2) and shifts to larger EXT values. The solvation of the molecule in water environment improves the conformational

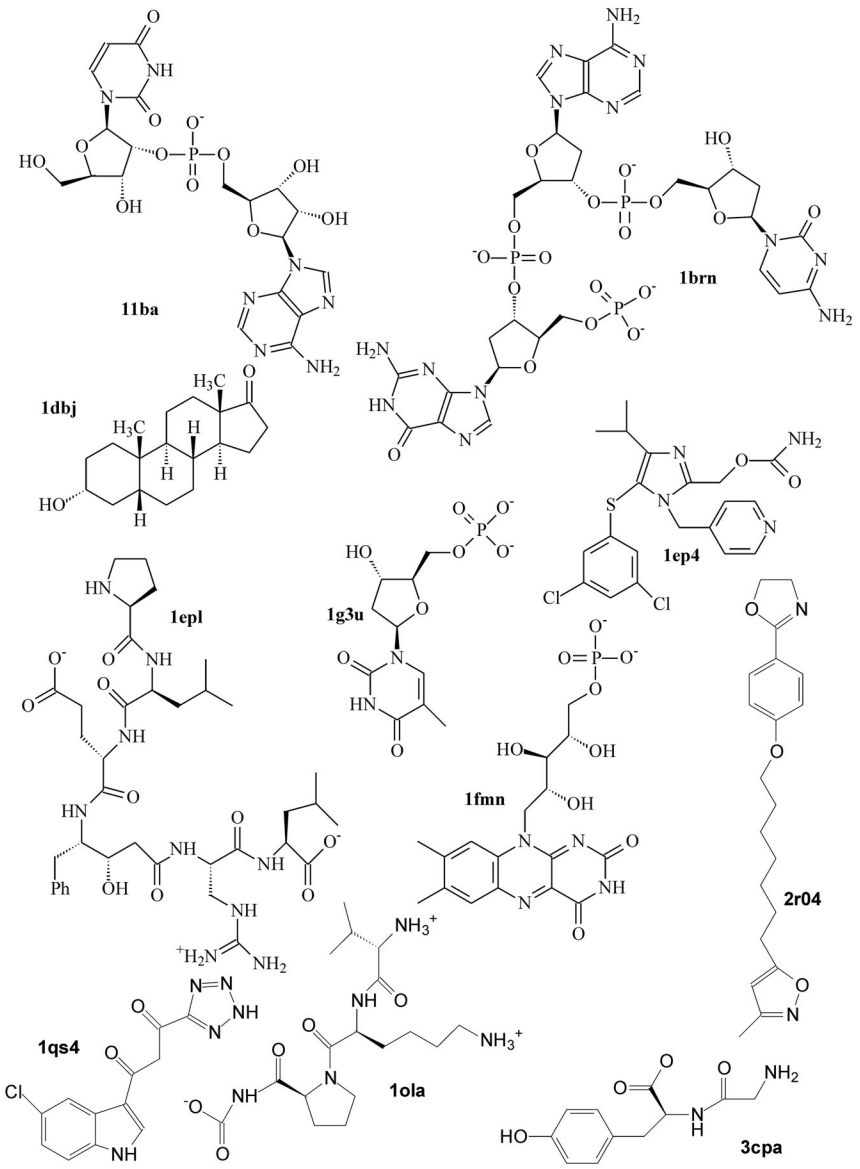

FIG. 1

Ligand structures from protein databank files in Table I. The structures are labeled by the PDB identification code. Phenyl groups are not drawn but replaced by 'Ph'

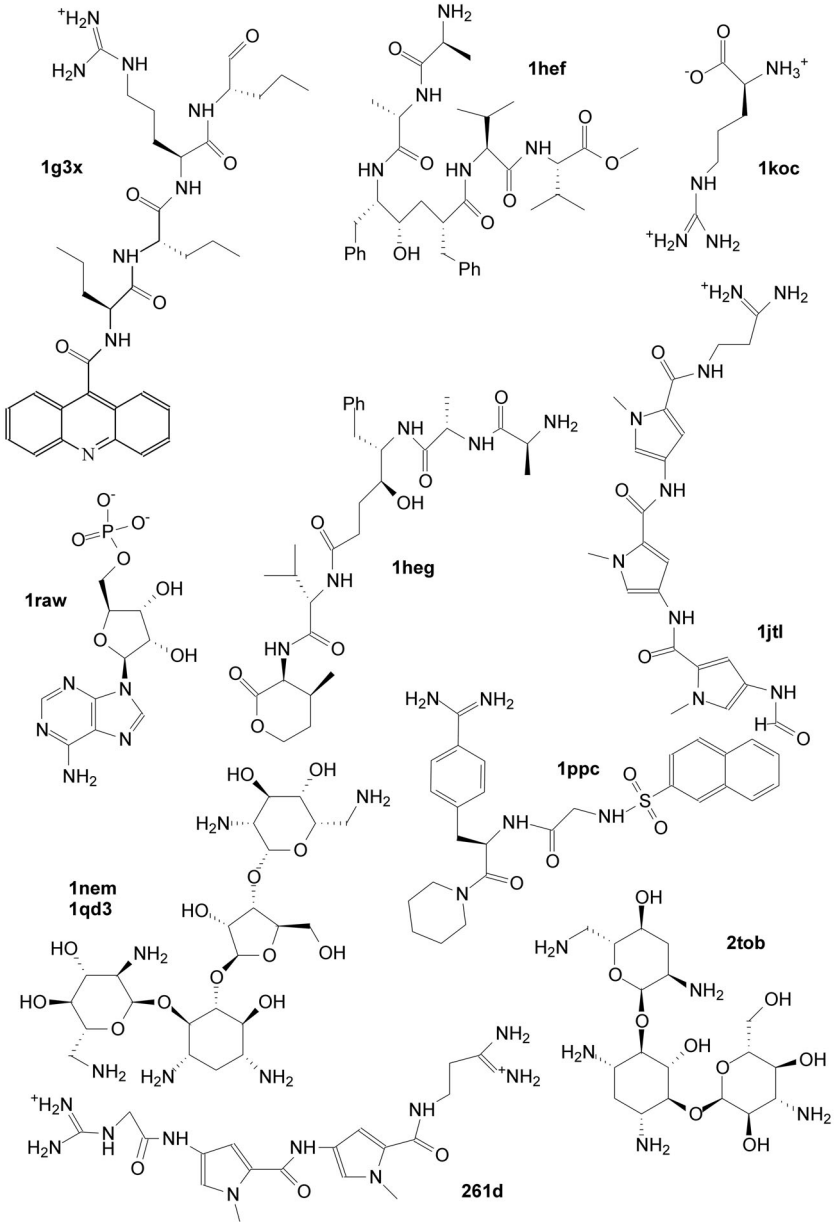

FIG. 1
(*Continued*)

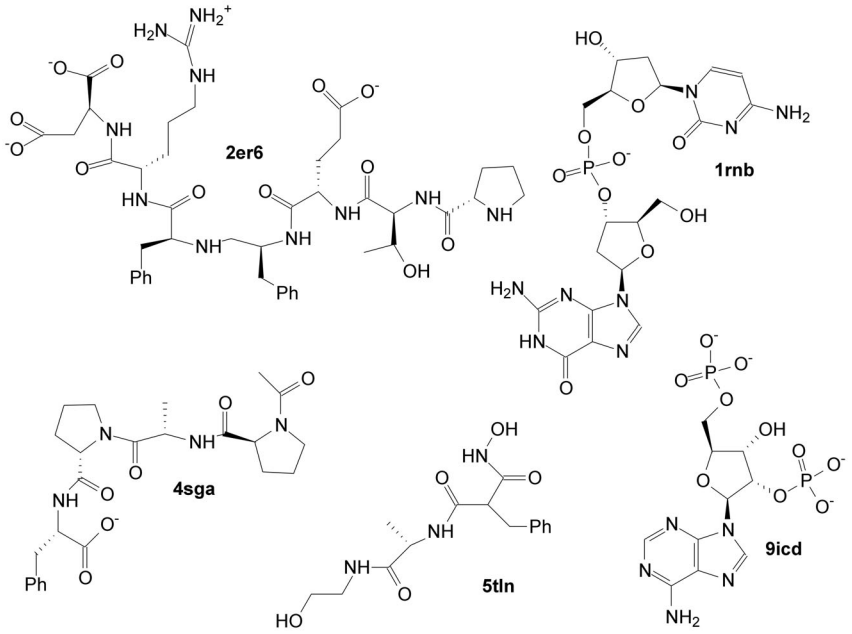

FIG. 1
(*Continued*)

range and shifts the distribution to the right. This is probably a consequence of a damping of the electrostatic interactions by water environment, smoothing out electrostatic attraction and repulsion of intramolecular groups, which eventually increases the number of favorable conformations. A second observation made is that the experimental X-ray conformations depicted on the graphs (symbols X) of the vacuum simulations often fall outside the range on the right side (Fig. 2). This means that the X-ray conformations are very extended. Going to the water plots in Fig. 3 the reader can verify that the X-ray structures are covered by most conformations found in the simulations. An exception is 261d, which is not found by the conformational search in water, but is found in the vacuum conformational search. We may conclude that searching conformational space using only a non-polar environment (vacuum) is inferior to a search in water environment for ligands having both polar and hydrophobic area.

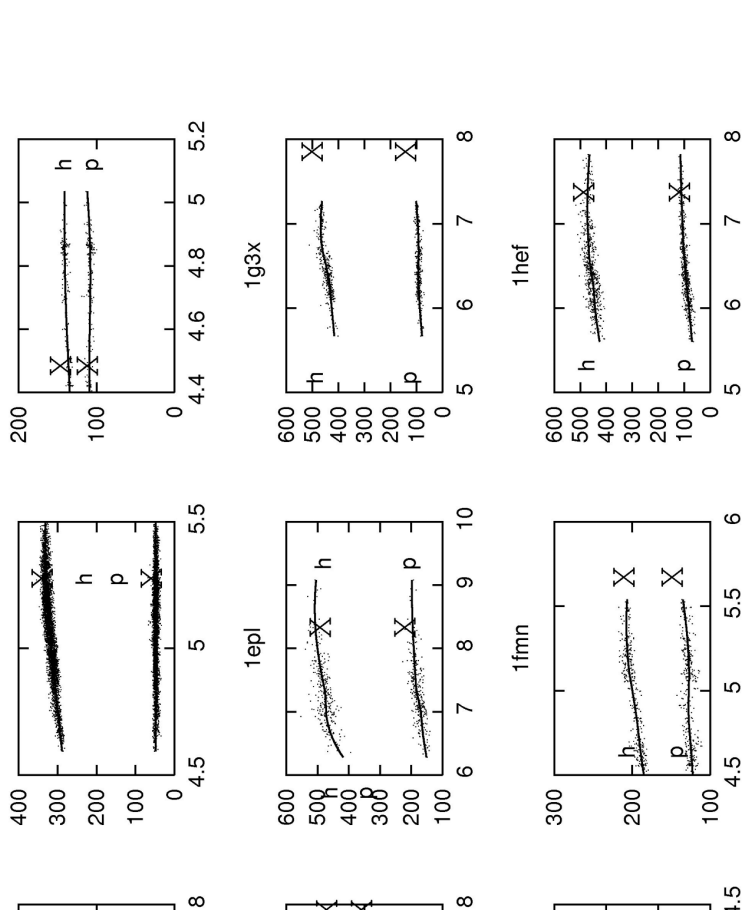

FIG. 2
Graphs showing the polar (p) and hydrophobic (h) solvent-accessible surface (SAS, in Å^2, *y*-axis) versus their geometric extent (EXT, in Å, *x*-axis), for every structure of the conformational search, calculated for all the ligands of the PDB entries of Table I. During the search, molecules are in vacuum environment. In every graph, two points representing the (geometric extent, polar SAS) and (geometric extent, hydrophobic SAS) for the unique ligand conformation in every PDB entry are indicated by the symbol 'X'. There is only one graph for PDB entries 1nem and 1qd3, having the same ligand. This graph shows the 1nem conformational search results. The (EXT, SAS) values for the X-ray conformation are plotted as 1nem (#) and 1qd3 (X)

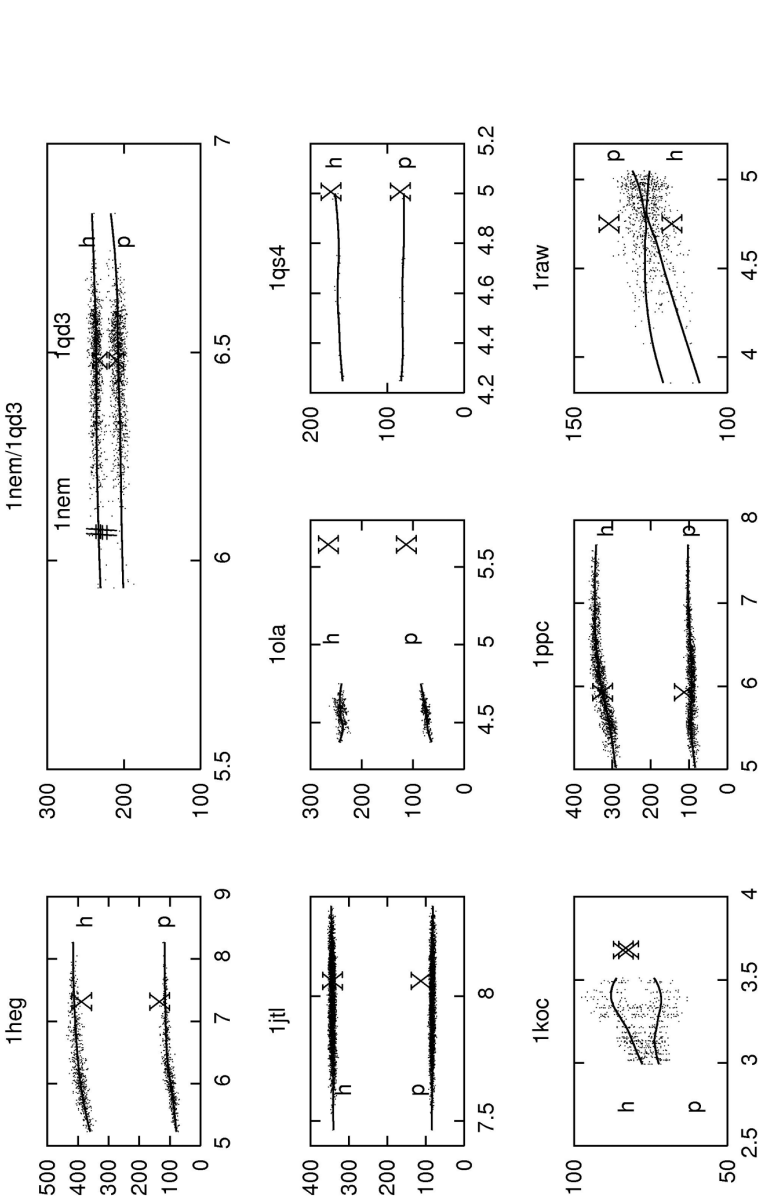

FIG. 2
(*Continued*)

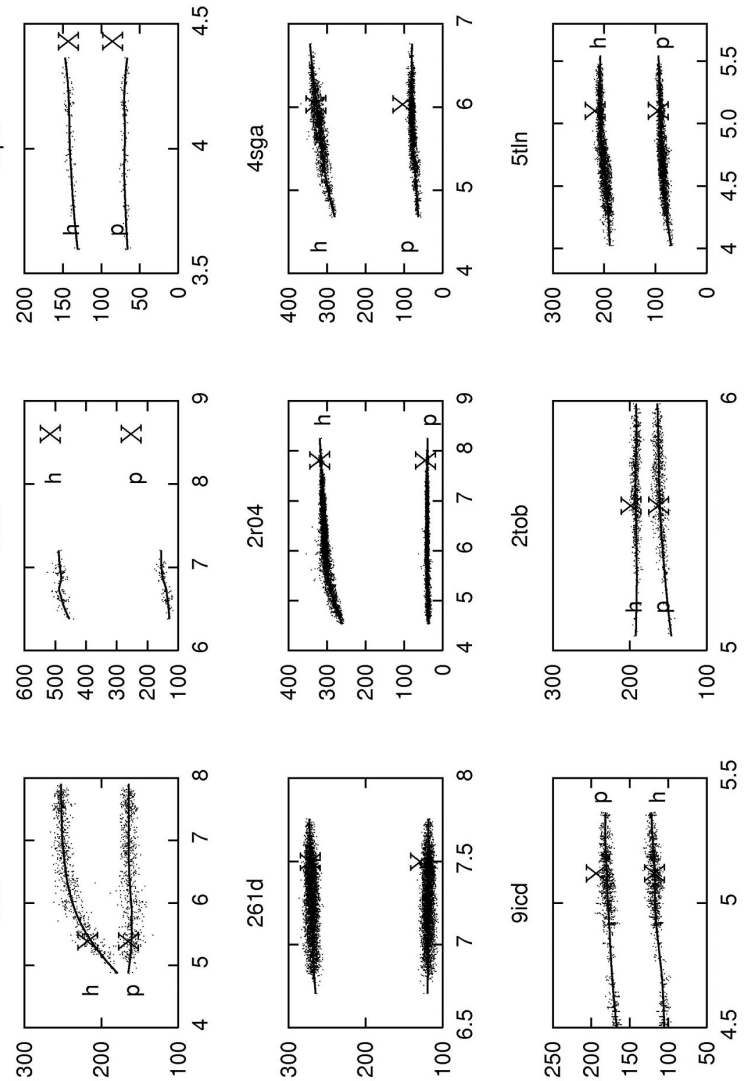

FIG. 2
(*Continued*)

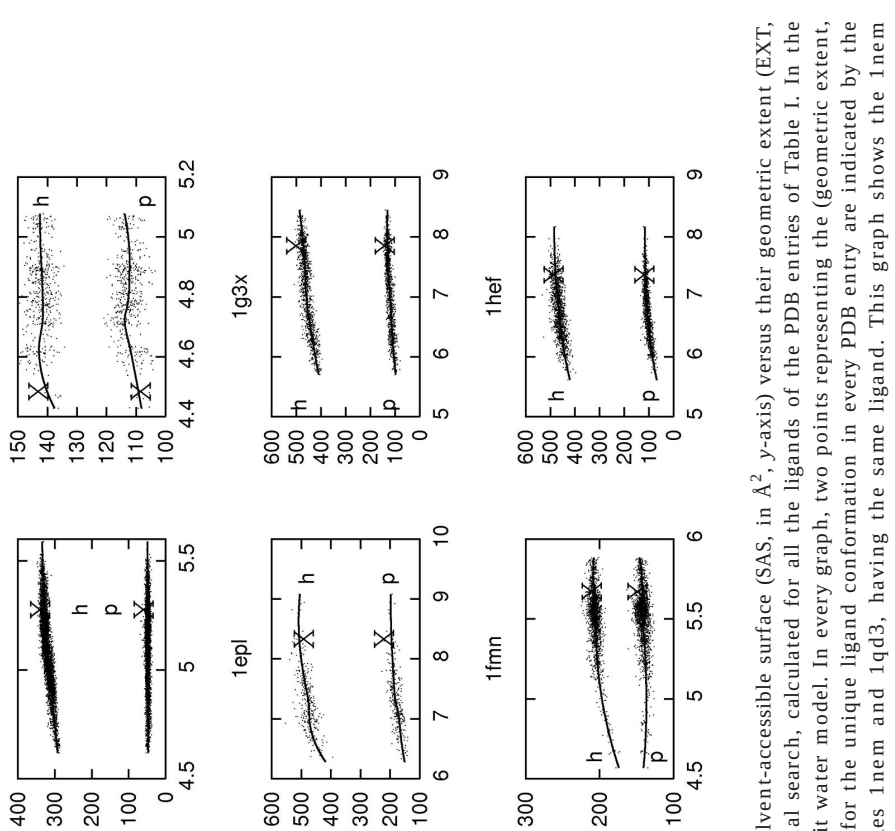

FIG. 3

Graphs showing the polar (p) and hydrophobic (h) solvent-accessible surface (SAS, in Å^2, *y*-axis) versus their geometric extent (EXT, in Å, *x*-axis), for every structure of the conformational search, calculated for all the ligands of the PDB entries of Table I. In the conformational search, the solvent is set to an implicit water model. In every graph, two points representing the (geometric extent, polar SAS) and (geometric extent, hydrophobic SAS) for the unique ligand conformation in every PDB entry are indicated by the symbol 'X'. There is only one graph for PDB entries 1nem and 1qd3, having the same ligand. This graph shows the 1nem conformational search results. The (EXT, SAS) values for the X-ray conformation are plotted as 1nem (#) and 1qd3 (X)

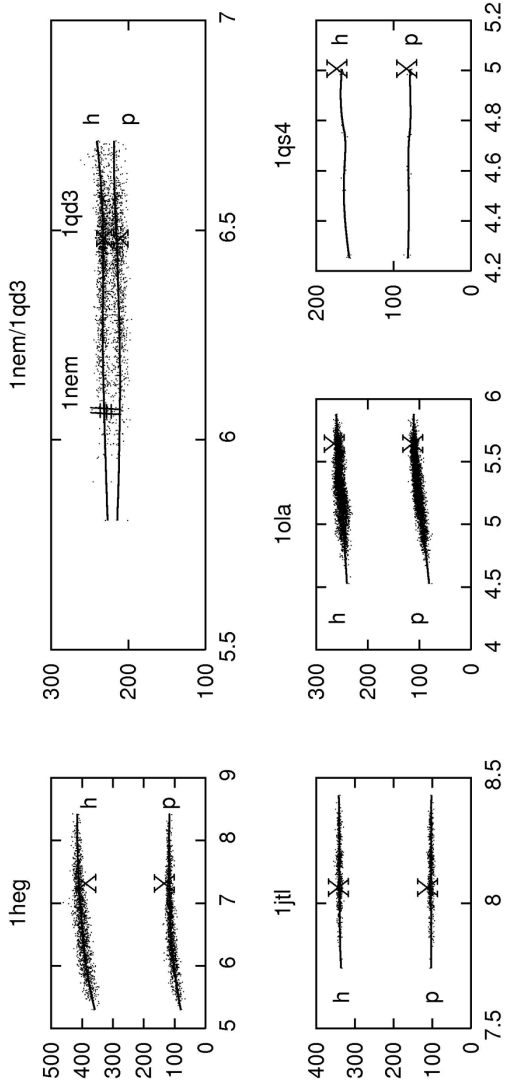

FIG. 3
(*Continued*)

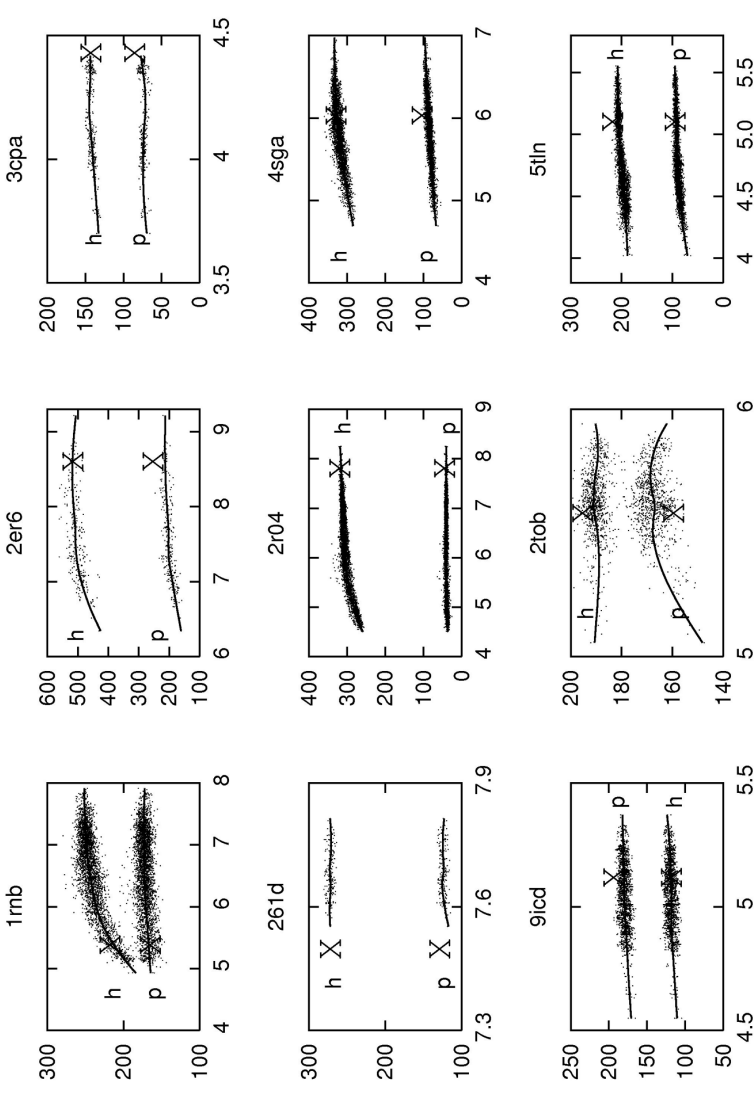

FIG. 3
(*Continued*)

Solvent-Accessible Surface and Geometric Extent

In Fig. 3 we see a clear increase in accessible surface when extending the molecules, but some of the X-ray conformation parameters plotted, clearly show that the ligands do not preferentially bind in their most extended conformation. Not all curves in Fig. 3 have a smooth increasing curvature. There are many exceptions like the less flexible ligand (1dbj), see Fig. 1. Also less clear are some nucleotides (1g3u) and others like 1jtl and 1koc. The curves of 1hef and 1heg, which are HIV-1 protease peptide inhibitors, do show the expected pattern, with the X-ray conformation at the high end of the curves. It is well established that peptide mimetics, are binding in the active site in an extended conformation[26].

In other cases either the polar or hydrophobic SAS component decreases while the other one is constant or even increases: 1g3u, 1raw, 1qs4. Entries 1nem and 1qd3 contain the same ligand neomycin B in a different conformation. The 1nem ligand conformation has a higher hydrophobic SAS (227 Å^2) than 1qd3 (209 Å^2) but a lower polar component (222 and 230 Å^2, respectively). The curves of 1qd3/1nem do not show a good discriminating effect. Figure 3 shows a slight increase in the polar component of the solvent-accessible surface when extending the molecule. The hydrophobic accessible surface is smaller and changes are small. An explanation for this may be that 19 of 42 non-hydrogen atoms are defined as polar atoms, lying at the surface of the molecule, while the hydrophobic atoms (carbon) are more in the interior of the ligand molecule (Fig. 1). However, the 1nem (#) and 1qd3 (X) ligand conformations show a geometric extent and accessible molecular surface, which are not maximal, and not comparable with each other (Fig. 3). This means that the ligand in this example clearly does not seek maximal extension or contact surface with the receptor.

An interesting approach is the construction of a composite binding pocket (CBP)[27]. Here the authors superimpose existing RT inhibitor X-ray structures onto each other and calculate an overall molecular surface. This surface is then colored to reflect hydrophobic regions and H-bond possibilities. This CBP pocket could then be used for further analysis and design of new compounds maximizing the contact surface or improving certain interactions. This approach might be a useful starting point for new inhibitor design. It is based on similar ideas as presented here, i.e. that optimizing contact surface would lead to a better inhibitor.

CONCLUSIONS

This analysis reveals the difficulty of carrying out predictive studies for the binding of ligands to receptors and enzymes. It sounds logic that optimizing and increasing surface complementarity between the low-energy conformation of a ligand and the receptor or active site will lead to more potent compounds. Calculation of SAS will certainly help in this selection procedure. However, this can only be successfully carried out (resulting in a clear structure-activity relationship) using a well-defined series of structurally related molecules, which were tested in calibrated conditions. Too many other parameters are involved in the formation of a stable complex, not at least, dynamic factors that are difficult to generalize.

ABBREVIATIONS AND SYMBOLS

5ClTEP	1-(5-chloroindol-3-yl)-3-hydroxy-3-(2*H*-tetrazol-5-yl)propen-1-one (Fig. 1, 1qs4)
AMP	adenosine monophosphate (Fig. 1, 1raw)
DMS	"Distributed MS", a version of the solvent-accessible surface computation program
DNA	deoxyribonucleic acid
d(GC)	short DNA chain with guanine and cytosine bases (Fig. 1, 1rnb)
d(CGAC)	short DNA chain (Fig. 1, 1brn)
EXT	geometric extent
H-256	Pro-Thr-Glu-Phe-Phe-Arg-Glu peptide (Fig. 1, 2er6)
HEF	Ala-Ala-Phe-Ohe-Phe-Val-Val-OMe peptide (Fig. 1, 1hef)
HEG	Ala-Ala-Phe-Ohe-Gly-Val-Val-OMe peptide (Fig. 1, 1heg)
HIV	human immunodeficiency virus
NADP	nicotinamide-adenine-dinucleotide phosphate (Fig. 1, 9icd)
NAPAP	1-{N^2-[*N*-(naphthalene-2-sulfonyl)glycyl]-4-carbamimidoylphenylalanyl}piperidine (Fig. 1, 1ppc)
NeoB	Neomycin B (Fig. 1, 1qd3)
OppA	oligopeptide binding protein
PDB	protein data bank
RNA	ribonucleic acid
RT	reverse transcriptase
S-1153	{5-[(3,5-dichlorophenyl)sulfanyl]-4-isopropyl-1-(pyridin-4-ylmethyl)-1*H*-imidazol-2-yl}methyl carbamate (Fig. 1, 1ep4)
SAS	solvent accessible surface
TAR	trans-activation-response element
TMP	thymidine monophosphate (Fig. 1, 1g3u)
VKPG	Val-Lys-Pro-Gly peptide (Fig. 1, 1ola)
w71	5-{7-[4-(4,5-dihydrooxazol-2-yl)phenoxy]heptyl}-3-methylisoxazole (Fig. 1, 2r04)

REFERENCES

1. Lloyd D. G., Buenemann C. L., Todorov N. P., Manallack D. T., Dean P. M.: *J. Med. Chem.* **2004**, *47*, 493.
2. Jain A. N.: *J. Med. Chem.* **2004**, *47*, 947.
3. Jones G., Willett P., Glen R. C., Leach A. R., Taylor R.: *J. Mol. Biol.* **1997**, *267*, 727.
4. Morris G. M., Goodsell D. S., Halliday R. S., Huey R., Hart W. E., Belew R. K., Olson A. J.: *J. Comput. Chem.* **1998**, *19*, 1639.
5. Kuntz I. D., Meng E. C., Shoichet B. K.: *Acc. Chem. Res.* **1994**, *27*, 117.
6. Jain A. N.: *J. Med. Chem.* **2003**, *46*, 499.
7. Nicklaus M. C., Wang S., Driscoll J. S., Milne G. W. A.: *Bioorg. Med. Chem.* **1995**, *3*, 411.
8. Bostrom J., Norrby P. O., Liljefors T.: *J. Comput.-Aided Mol. Des.* **1998**, *12*, 383.
9. Perola E., Charifson P. S.: *J. Med. Chem.* **2004**, *47*, 2499.
10. Sobolev V., Wade R. C., Vriend G., Edelman M.: *Proteins* **1996**, *25*, 120.
11. Norel R., Petrey D., Wolfson H. J., Nussinov R.: *Proteins* **1999**, *36*, 307.
12. Hou T., Wang J., Chen L., Xu X.: *Protein Eng.* **1999**, *12*, 639.
13. Gabb H. A., Jackson R. M., Sternberg M. J.: *J. Mol. Biol.* **1997**, *272*, 106.
14. Chen R., Weng Z.: *Proteins* **2003**, *51*, 397.
15. Stanfield R. L., Wilson I. A.: *Curr. Opin. Struct. Biol.* **1995**, *5*, 103.
16. Berman H. M., Westbrook J., Feng Z., Gilliland G., Bhat T. N., Weissig H., Shindyalov I. N., Bourne P. E.: *Nucleic Acids Res.* **2000**, *28*, 235.
17. Mohamadi F., Richards N. G. J., Guida W. C., Liskamp R., Lipton M., Caufield C., Chang G., Hendrickson T., Still W. C.: *J. Comput. Chem.* **1990**, *11*, 440.
18. Saunders M., Houk K. N., Wu Y., Still W. C., Lipton M., Chang G., Guida W. C.: *J. Am. Chem. Soc.* **1990**, *112*, 1419.
19. Weiner P., Kollman P. A., Nguyen D. T., Case D. A.: *J. Comput. Chem.* **1986**, *7*, 230.
20. Still W. C., Tempczyk A., Hawley R. C., Hendrickson T.: *J. Am. Chem. Soc.* **1990**, *112*, 6127.
21. Richards F. M.: *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
22. Ferrin T. E., Huang C. C., Jarvis L. E., Langridge R.: *J. Mol. Graphics* **1988**, *6*, 13.
23. Wang R., Fu Y., Lai L.: *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 615.
24. Cleveland W. S., Grosse E.: *Stat. Comput.* **1991**, *1*, 47.
25. Broeker H. B., Campbell J., Cunningham R., Denholm D., Elber G., Fearick R., Grammes C., Hart L., Hecking L., Koenig T., Kotz D., Kubaitis E., Lang R., Lehmann A., Mai A., Steger C., Tkacik T., Van der Woude J., Van Zandt J. R., Woo A.: Gnuplot 3.7; http://gnu.org.
26. Huang X., Xu L., Luo X., Fan K., Ji R., Pei G., Chen K., Jiang H.: *J. Med. Chem.* **2002**, *45*, 333.
27. Vig R., Mao C., Venkatachalan T. K., Tuel-Ahlgren L., Sudbeck E. A., Uckun F. M.: *Bioorg. Med. Chem.* **1998**, *6*, 1789.
28. Vitagliano L., Adinolfi S., Riccio A., Sica F., Zagari A., Mazzarella L.: *Protein Sci.* **1998**, *7*, 1691.
29. Buckle A. M., Fersht A. R.: *Biochemistry* **1994**, *33*, 1644.
30. Arevalo J. H., Taussig M. J., Wilson I. A.: *Nature* **1993**, *365*, 859.
31. Ren J., Nichols C., Bird L. E., Fujiwara T., Sugimoto H., Stuart D. I., Stammers D. K.: *J. Biol. Chem.* **2000**, *275*, 14316.
32. Bailey D., Cooper J. B.: *Protein Sci.* **1994**, *3*, 2129.

33. Fan P., Suri A. K., Fiala R., Live D., Patel D. J.: *J. Mol. Biol.* **1996**, *258*, 480.
34. de la Sierra L., Munier-Lehmann H., Gilles A. M., Barzu O., Delarue M.: *J. Mol. Biol.* **2001**, *311*, 87.
35. Malinina L., Soler-Lopez M., Aymami J., Subirana J. A.: *Biochemistry* **2002**, *41*, 9341.
36. Murthy K. H. M., Winborne E. L., Minnich M. D., Culp J. S., Debouck C.: *J. Biol. Chem.* **1992**, *267*, 22770.
37. Uytterhoeven K., Sponer J., Van Meervelt L.: *Eur. J. Biochem.* **2002**, *269*, 2868.
38. Yang Y., Kochoyan M., Burgstaller P., Westhof E., Famulok M.: *Science* **1996**, *272*, 1343.
39. Jiang L., Majundar A., Hu W., Jaishree T. J., Xu W., Patel D. J.: *Structure* **1999**, *7*, 817.
40. Tame J. R. H., Murshudov G. N., Dodson E. J., Neil T. K., Dodson G. G., Higgins C. F., Wilkinson A. J.: *Science* **1994**, *264*, 1578.
41. Bode W., Turk D., Sturzebecher J.: *Eur. J. Biochem.* **1990**, *193*, 175.
42. Faber C., Sticht H., Schweimer K., Rosch P.: *J. Biol. Chem.* **2000**, *275*, 20660.
43. Goldgur Y., Craigie R., Cohen G. H., Fujiwara T., Yoshinaga T., Fujishita T., Sugimoto H., Endo T., Murai H., Davies D. R.: *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13040.
44. Dieckmann T., Suzuki E., Nakamura G. K., Feigon J.: *RNA* **1996**, *2*, 628.
45. Baudet S., Janin J.: *J. Mol. Biol.* **1991**, *219*, 123.
46. Nunn C. M., Garman E., Neidle S.: *Biochemistry* **1997**, *36*, 4792.
47. Sali A., Veerapandian B., Cooper J. B., Foundling S. I., Hoover D. J., Blundell T. L.: *EMBO J.* **1989**, *8*, 2179.
48. Badger J., Minor I., Oliveira M. A., Smith T. J., Rossmann M. G.: *Proteins* **1989**, *6*, 1.
49. Jiang L., Patel D. J.: *Nat. Struct. Biol.* **1998**, *5*, 769.
50. Christianson D. W., Lipscomb W. N.: *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 7568.
51. James M. N. G., Sielecki A. R., Brayer G. D., Delbaere L. T. J., Bauer C. A.: *J. Mol. Biol.* **1980**, *144*, 43.
52. Holmes M. A., Matthews B. W.: *Biochemistry* **1981**, *20*, 6912.
53. Hurley J. H., Dean A. M., Koshland D. E., Stroud R. M.: *Biochemistry* **1991**, *30*, 8671.