Journal of Computer-Aided Molecular Design, 8 (1994) 299–306
ESCOM

J-CAMD 248

Evaluating docked complexes with the HINT exponential function and empirical atomic hydrophobicities

Elaine C. Meng[a,*], Irwin D. Kuntz[a], Donald J. Abraham[b] and Glen E. Kellogg[b,*]

[a]*Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143-0446, U.S.A.*
[b]*Department of Medicinal Chemistry, School of Pharmacy, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298-0540, U.S.A.*

Received 10 October 1993
Accepted 14 February 1994

Key words: Molecular recognition; Ligand binding; Complementarity; Hydropathy

SUMMARY

Methods that predict geometries of ligands binding to receptor molecules can facilitate ligand discovery and yield information on the factors governing complementarity. Here, the use of atomic hydrophobicities in evaluating binding modes has been examined with four ligand–receptor complexes of known structure. In each system, hundreds of hypothetical binding orientations were generated with DOCK and evaluated using the HINT (Hydropathic INTeractions) exponential function and atomic hydrophobic constants. In three of the four systems, the experimental binding mode received the best HINT score; in the fourth system, the experimental binding mode scored only slightly lower than a similar, apparently reasonable orientation. The HINT function may be generally useful as a scoring method in molecular docking.

INTRODUCTION

Ligand binding is a crucial aspect of protein function. Computational approaches to predicting binding geometries are of interest not only for the design of bioactive compounds, but also for insights they may yield on the principles of molecular recognition. In general, such methods 'dock' molecules together in many ways and 'score' or evaluate each orientation. Ligand discovery applications involve searching through a database of compounds, with the idea that compounds which score well should be more likely to bind to the target macromolecule. Once a lead compound has been discovered, docking may again be useful; knowledge of probable binding modes is needed so that one can suggest structural modifications intended to form, enhance or disrupt specific interactions with the receptor.

*To whom correspondence should be addressed.

Assuming there is one true binding mode for given ligand and receptor structures, a correct prediction consists of two parts: generation of the correct orientation, usually among others, and identification of this (family of) orientation(s) as the most favorable by using a scoring function. With the DOCK suite of programs [1–3], it has been shown that when sufficient sampling is performed [4], experimental geometries can be reproduced quite accurately and identified by different scoring functions. The DOCK force field (FF) score has been the most successful of the functions tried in these 'complex regeneration' experiments. It is a simple interaction energy, consisting of a Lennard-Jones van der Waals (vdW) term and a Coulombic electrostatic term [3].

Other scoring methods are of interest, however, particularly those which include some measure of the desolvation that occurs upon complex formation. The FF score does not include desolvation contributions. Although highly effective in identifying the experimental orientation of a single compound, it is only suitable for semiquantitative and qualitative comparisons among different compounds. Whereas the costs of partial desolvation may be roughly equivalent for different orientations of one compound, they may be extremely different for different compounds, especially those dissimilar in terms of hydrophobicity and number of formal charges. Desolvation energies are also likely to be important for identifying binding modes that are dominated by hydrophobic interactions.

An ideal score should incorporate desolvation, including the hydrophobic effect, and be able to identify the correct orientation of a given compound out of hundreds to thousands of possibilities. The score should be atom- or functional group-based, so that desolvation energies depend on molecule orientation as well as structure, according to which portions are buried in the complex.

Thus, the use of atomic hydrophobicity/hydrophilicity descriptors in scoring docked orientations is very attractive. Hydrophobic interactions are related to solvent partitioning phenomena, and the octanol/water partition coefficient P contains significant thermodynamic information. A molecule's partition coefficient is the equilibrium ratio of its concentration in octanol to its concentration in water; log P will be positive for a compound found primarily in the octanol phase and negative for a compound found primarily in the aqueous phase. The program HINT estimates log P from molecular structure and, especially important for the current work, derives additive atomic contributions to this parameter [5,6]. The 'atomic hydrophobic constants' are obtained by reduction of Hansch–Leo fragment constants [7] and contain implicit contributions from bond, chain, branch and polar proximity factors. In general, polar atoms are assigned negative values and nonpolar atoms are assigned positive values. Although the breakdown of a molecular property into atomic contributions is a mental construct, rather than a strict representation of reality, it can be extremely valuable in computational chemistry; partial charge models, for instance, are widely used in molecular simulations. Molecular log P values calculated with HINT are similar to those calculated with the method of Hansch and Leo. For example, the HINT-calculated log P of glucose is –2.99, the Hansch–Leo value is –2.98 [7], and the experimental estimate is –3.24 [7].

The HINT exponential function and atomic hydrophobicities have been used to score docked orientations in four systems of known structure. Since the crystallographic conformations were retained during docking, the important but separate issue of molecular flexibility has not been addressed here. In three of the four systems, the HINT score was able to identify the experimental binding mode, as defined by a root-mean-square deviation (rmsd) of 2 Å or less from the crystal structure position, with a near miss occurring in the fourth system. The success of the

method in 'complex regeneration' experiments suggests that it captures important aspects of molecular recognition.

MATERIALS AND METHODS

Structure preparation

Four well-determined structures of ligand–receptor complexes were selected from the Brookhaven Protein Data Bank [8] (Fig. 1 and Table 1): 3cpa (carboxypeptidase A–glycyltyrosine) [10]; 4 dfr (dihydrofolate reductase–methotrexate) [11]; 2gbp (periplasmic binding protein–glucose) [12]; and 6rsa (ribonuclease A–uridine vanadate) [13]. Different aspects of complementarity are evident in these complexes, including salt-bridge formation, hydrogen bonding and hydrophobic interactions. The same four systems were used in previous tests of scoring [3]. In each case, waters and ions were removed, the ligand and receptor were separated, and hydrogens were added in standard geometries.

DOCK and associated programs

A molecular surface for the receptor region of interest was created with the MS algorithm [14] and used to calculate spheres for docking [1]. Contact grids [2] were created with close-contact limits of 2.3 and 2.8 Å for receptor polar and nonpolar atoms, respectively; hydrogens were ignored. During docking, each nonhydrogen ligand atom was checked for bad contacts with the receptor by examining the nearest grid point; orientations with any 'bumping' atoms were discarded.

The ligand-site matching algorithm is purely geometric [2]. Sets of sphere centers are matched to sets of ligand atoms, based on pairwise internal distances. Least-squares fitting of the matched points produces an orientation. The sphere–sphere and atom–atom distances are first sorted into bins; a sphere–atom pairing is only considered when these points are in the corresponding bins. In the current work, an orientation was generated when the distances among four spheres were found equivalent to the distances among four ligand atoms, within a tolerance of 1.5 Å. All bin widths and overlaps were initially set to 0.8 and 0.2 Å, respectively. In the 2gbp and 4dfr systems, the bin widths and overlaps were reduced to 0.4 and 0.1 Å, to bring the number of sterically acceptable orientations below 1000.

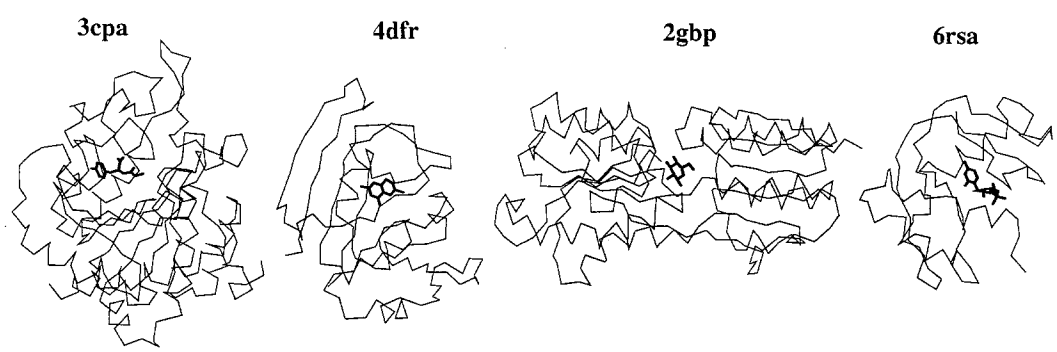

Fig. 1. Receptor C^{α} traces and ligands for the test systems 3cpa, 4dfr, 2gbp and 6rsa. The ligands are shown in heavy lines, omitting hydrogens. The figure was generated with UCSF MidasPlus [9].

TABLE 1
LIGAND–RECEPTOR TEST SYSTEMS EVALUATED WITH HINT

System	Resolution (Å)	Receptor	Complexed ligand	Docked ligand, formal charge
3cpa	2.0	Carboxypeptidase A	Glycyl-L-tyrosine	Glycyl-L-tyrosine, 0 (zwitterion)
4dfr	1.7	Dihydrofolate reductase	Methotrexate	2,4-Diamino-6-methylpteridine, +1
2gbp	1.9	Periplasmic binding protein	β-D-Glucose	β-D-Glucose, 0
6rsa	2.0	Ribonuclease A	Uridine vanadate	Uridine 3'-phosphate, –2

HINT

Atomic parameter assignments were performed on the ligands with all hydrogens explicit and on the protein receptors with only the 'essential' (potentially hydrogen-bonding) hydrogens. Protein residues were modeled in the charge states expected at neutral pH. Each ligand orientation was scored using an empirical function developed earlier [5,6]:

$$B = \sum_{i=1}^{rec} \sum_{j=1}^{lig} s_i a_i s_j a_j T_{ij} e^{-r_{ij}}$$

This score is considered unitless and consists of a double sum over receptor atoms i and ligand atoms j, where s is the solvent-accessible surface area and a is the hydrophobic constant of an atom, T is a sign-flip function that discriminates between acid–base interactions, which are 'favorable', and acid–acid and base–base interactions, which are 'unfavorable', and r_{ij} is the distance between atoms i and j. The cutoff distance for interactions was 10 Å. Positive scores reflect favorable interactions. A simple Lennard-Jones vdW potential is also available for scoring within HINT; use of this function alone or in a 50:1 combination with the exponential term (to give hydropathy:vdW contributions of approximately 2:1) was investigated as well. Besides being weighted by a factor of 50.0 in the combined score, the Lennard-Jones potential is reversed in sign, in keeping with the positive/favorable, negative/unfavorable convention within HINT.

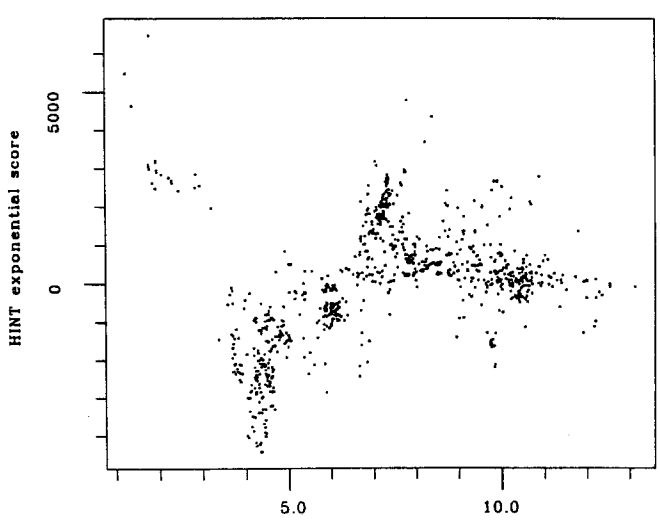

Fig. 2. Rmsd values versus HINT exponential score for the carboxypeptidase A system.

TABLE 2
SUMMARY OF HINT SCORING RESULTS FOR THE LIGAND–RECEPTOR TEST SYSTEMS

System	Orientations[a]	Best score	Rmsd[b] (Å)	Crystal score[c]
Carboxypeptidase A–glycyltyrosine	876	6470	1.774	5380
Dihydrofolate reductase–diaminomethylpteridine	703	4390	0.997	3860
Periplasmic binding protein–glucose	157	562	0.272	505
Ribonuclease A–uridine phosphate[d]	682	1040	4.319	823

[a] Number of sterically acceptable orientations generated by DOCK and evaluated with HINT.
[b] Rmsd of the best scoring orientation, relative to the crystal structure orientation.
[c] Score of the crystal structure orientation.
[d] The second best orientation had a score of 1020 and an rmsd of 0.825 Å.

RESULTS

For each system, the rmsd of every ligand orientation relative to the experimentally observed orientation was calculated. Hydrogens were not included. Rmsd values are plotted versus HINT scores in Figs. 2–5, and the results are summarized in Table 2.

Carboxypeptidase A

Orientations of glycyltyrosine with rmsd values below 2.0 Å describe essentially the experimental binding mode. Relative to the crystallographic orientation, structures with rmsd values just above 2.0 Å are angled slightly, 3.0–5.0 Å structures are barrel-rolled and translated along the long axis of the molecule, and structures with rmsd values greater than 6.0 Å are flipped end-to-end. The HINT score clearly identifies the experimental binding mode; the 3.0–5.0 Å orientations are disfavored, and the closest competitors are the end-to-end flipped orientations (Fig. 2).

Dihydrofolate reductase

Rmsd values below 2.0 Å correspond to orientations of diaminomethylpteridine similar to the observed geometry. These receive the best HINT scores (Fig. 3). The 3.0 Å structures are barrel-rolled and angled slightly relative to the crystallographic orientation, the 4.0 Å structures are angled approximately 90°, and the 5.0 Å structures are flipped end-to-end, so that the methyl substituent is pointing in the opposite direction. Structures with rmsd values greater than 8.0 Å do not overlap the experimental orientation.

Periplasmic binding protein

In the crystal structure, glucose is positioned in the middle of a narrow tunnel traversing the protein (Fig. 1). There are three clusters of rmsd values, corresponding to orientations similar to the crystallographic geometry (rmsd less than 1.0 Å), structures that overlay the observed orientation but are flipped or rotated in several different ways (rmsd 3.0–5.0 Å), and structures located in either end of the tunnel (rmsd greater than 10.0 Å). The orientations with the lowest rmsd values receive the best HINT scores (Fig. 4).

Ribonuclease A

Two groups of uridine phosphate orientations receive scores similar to that of the experimental binding mode. In one family (rmsd 4.0–6.0 Å), the phosphate is positioned essentially correctly, with the rest of the molecule angled 60–90° relative to the crystallographic orientation. A member

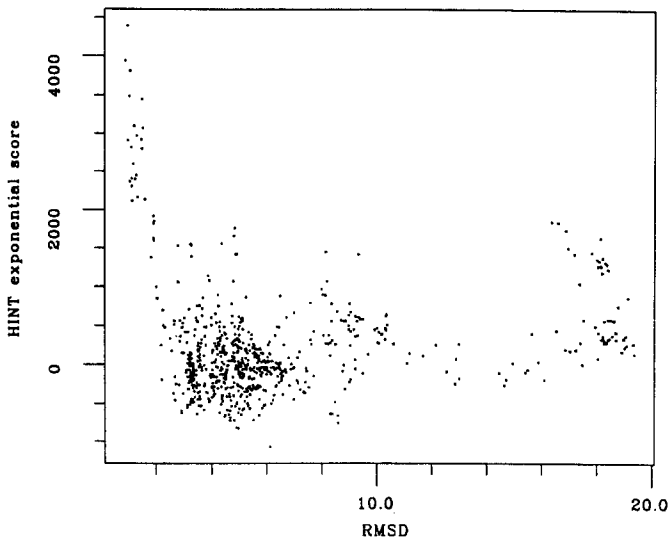

Fig. 3. Rmsd values versus HINT exponential score for the dihydrofolate reductase system.

of this family receives the best HINT score (Figs. 5 and 6). The structures with rmsd values close to 12.0 Å are related to the known orientation by a plane of reflection; the true and image phosphates face each other through the nitrogen of a nearby lysine side chain. Thus, ionic interactions between the phosphate and basic side chains make a significant contribution to the score.

DISCUSSION

The HINT scoring function includes favorable and unfavorable polar and hydrophobic interactions; the toggle function T recognizes polar–polar interactions as favorable when they are of the

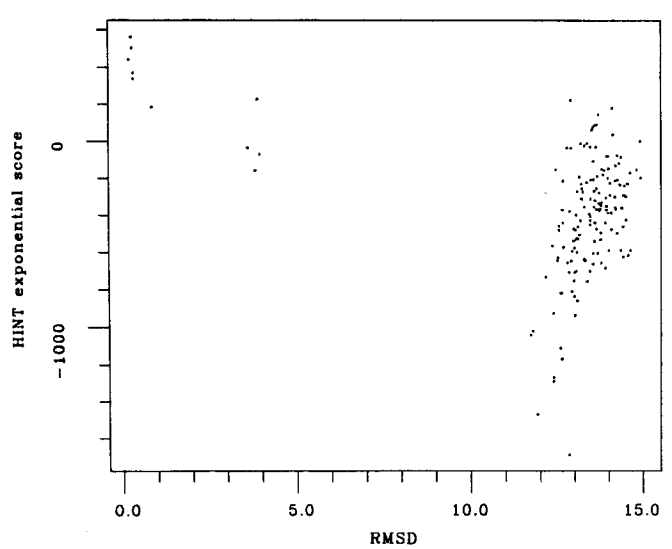

Fig. 4. Rmsd values versus HINT exponential score for the periplasmic binding protein system.

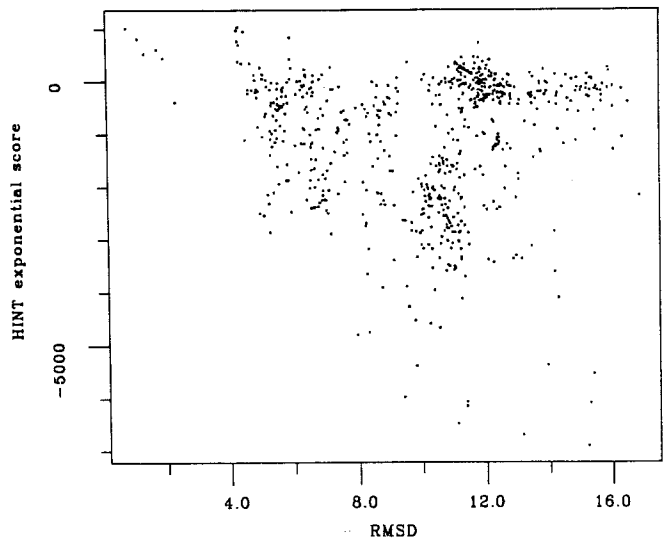

Fig. 5. Rmsd values versus HINT exponential score for the ribonuclease A system.

donor–acceptor type and unfavorable otherwise [5,6]. HINT is quite flexible, in that the user may define different functional forms for the term involving atomic hydrophobic constants, and combine this with a Lennard-Jones vdW term. The relative weights of the terms are also adjustable.

In the present work, a simple contact grid was used within DOCK to screen out ligand orientations that intersect the receptor. It was found that for the remaining orientations, the HINT exponential term alone was more succesful in identifying the known binding mode than the HINT Lennard-Jones term alone or a 50:1 combination of the Lennard-Jones and exponential terms (results not shown). Even with the default radii scaled by 0.9, orientations similar to the known mode received vdW scores comparable to or less favorable than those of members of other orientational families. No other functional forms or combinations of terms were investigated.

The success of the exponential term is encouraging, and suggests that further trials should be performed in single-molecule docking and in making comparisons among different ligands. While the DOCK FF score identified the experimental binding mode in all four systems (results not shown), the two kinds of scores were not highly correlated. This is due partly to the inclusion of vdW energy in the FF score and its omission in the HINT exponential function; furthermore, the HINT score is expected to place more emphasis on hydrophobic interactions. Ionic interactions

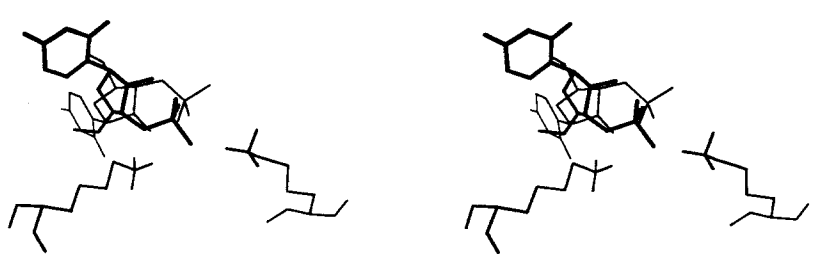

Fig. 6. Stereoview of the best scoring (heavy lines) and crystallographic orientations of uridine phosphate in the ribonuclease A system, shown with nearby lysine residues. The figure was generated with UCSF MidasPlus [9].

make significant contributions to both scores, as illustrated by the orientations of uridine phosphate that both consider favorable: all have the phosphate group in roughly the same location, interacting with lysine side chains, but the positions of the uracil and ribose moieties vary.

Two final comments on HINT scoring are in order. First, atoms of the same type often receive different hydrophobic constants, based on their differing structural surroundings. For example, the five hydroxyl oxygens in glucose receive five different values, ranging from −1.828 to −2.322. This is in contrast to other approaches where a few atom types are defined and always make the same solvation energy contributions per unit surface area. Second, partial atomic charges are not required. This can be a major practical advantage in docking, particularly when large, heterogeneous databases are to be searched. Derivation of the necessary atom-type information is generally a much simpler task than calculating partial charges.

CONCLUSIONS

In test systems with known ligand–receptor binding geometries, hundreds of ligand orientations from DOCK were evaluated with the HINT exponential function and atomic hydrophobic constants. The HINT score identified the experimental binding mode in three of the four systems, with a near miss occurring in the fourth. Although very dissimilar in functional form and in the parameters used, the HINT score and the DOCK force-field score both perform well in identifying crystal structure orientations. The two approaches will likely exhibit different strengths and weaknesses, and one may be preferable to the other, depending on the situation. The HINT results are encouraging and suggest that the method may be generally useful in conjunction with molecular docking.

ACKNOWLEDGEMENTS

We gratefully acknowledge support from NIH grants GM-31497 (I.D.K.) and HL-32793 (D.J.A.).

REFERENCES

1 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., J. Mol. Biol., 161 (1982) 269.
2 Shoichet, B.K., Bodian, D.L. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 380.
3 Meng, E.C., Shoichet, B.K. and Kuntz, I.D., J. Comput. Chem., 13 (1992) 505.
4 Meng, E.C., Gschwend, D.A., Blaney, J.M. and Kuntz, I.D., Proteins, 17 (1993) 266.
5 Wireko, F.C., Kellogg, G.E. and Abraham, D.J., J. Med. Chem., 34 (1991) 758.
6 Kellogg, G.E., Joshi, G.S. and Abraham, D.J., Med. Chem. Res., 1 (1992) 444.
7 Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, NY, 1979, pp. 1–339.
8 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
9 Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R., J. Mol. Graphics, 6 (1988) 13.
10 Rees, D.C. and Lipscomb, W.N., Proc. Natl. Acad. Sci. USA, 80 (1983) 7151.
11 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J., J. Biol. Chem., 257 (1982) 13650.
12 Vyas, N.K., Vyas, M.N. and Quiocho, F.A., Science, 242 (1988) 1290.
13 Borah, B., Chen, C.-W., Egan, W., Miller, M., Wlodawer, A. and Cohen, J.S., Biochemistry, 24 (1985) 2058.
14 Connolly, M.L., Science, 221 (1983) 709.