Journal of Computer-Aided Molecular Design, 8 (1994) 41–49
ESCOM

J-CAMD 233

The effect of physical organic properties on hydrophobic fields

Donald J. Abraham and Glen E. Kellogg

Department of Medicinal Chemistry, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298-0540, U.S.A.

Received 10 November 1993
Accepted 20 November 1993

Key words: HINT; Hydropathy; Hydrophobic fields; Partition coefficient

SUMMARY

Physical organic structural properties of small molecules and macromolecules such as bond count, branching and proximity between multiple polar fragments contribute significantly to measured hydrophobicity (log P). These structural properties are encoded in the Rekker and Leo methods of calculating log P as structural-dependent factors. Regardless of the size of the atom primitive set, methods predicting log P with only atom primitives can miss subtle structural detail within series of related compounds. The HINT (Hydropathic INTeractions) model for inter- and intramolecular noncovalent interactions calculates *atom-based* hydrophobic constants, but uses all Leo-type factors in the calculation rather than a large set of atom primitives. Two types of applications of HINT are discussed: evaluation of the binding of an inhibitor (A74704) to HIV-1 protease, where it is shown that modeling of the protonation state (i.e., Asp^{25}, Asp^{125}) in the protein can strongly influence perceived substrate binding; and the use of HINT to calculate a third (hydropathic) field for CoMFA can yield a statistically enhanced and predictive model for molecular design.

INTRODUCTION

The nature of hydrophobic interactions and their importance in the binding of drugs to proteins or receptors has long been appreciated. The Interlaken lecture (1993 Meeting of the Molecular Graphics Society) upon which this article is based addressed the question 'Do hydrophobic fields exist?'. Another question posed was 'Do Lennard-Jones interactions account for the majority of energy derived from interacting two hydrophobic side chains or surfaces?'. A summary of perti-nent research was presented to answer these questions and to define a common language for discussing hydrophobic interactions [1–9]. Historical contributions concerning hydrophobicity [10,11], and the development of specialized techniques that represent and use hydrophobic fields [12–25] were addressed. A summary of this portion of the lecture has been

published elsewhere [1]. The remainder of the lecture illustrated the use of our software program, HINT (Hydropathic INTeractions) which will be the focus of the present paper.

HYDROPHOBIC FIELDS

A hydrophobic field has been defined as 'a region of space characterized by atomic partition coefficients, having a determinable value at every point in the region' [1]. This concept describes a region of space characterized by a grid, containing the free energy of transfer of each atom in the molecule(s) under observation. The resulting maps depict hydropathic profiles for individual molecules or interactions between molecules. This methodology, incorporated into a computer program named HINT, provides useful and pragmatic constructs that enable one to understand, explain and in some instances predict hydrophobic and polar biomolecular interactions.

Hydrophobic fields as defined above characterize more than van der Waals attractive forces to the energy of hydrophobic interactions. This is an *empirical approach* to field representation that differs from the steric or electrostatic parameters derived in force field calculations. Also, hydrophobic interactions derived from transfer phenomena include a contribution from entropy [2] as well as hydrogen bonding phenomena and micro-dipolar or electrostatic interactions. In this regard, the affinity of a lipophilic solute for the lipophilic solvent in partitioning ($\Delta G_{transfer}$) can be calculated and this value includes the disorder of water molecules when lipophilic atoms or groups congregate, i.e., what is known classically as hydrophobic bonding.

ERRORS FROM THE DIRECT ADDITION OF HYDROPATHIC PARAMETERS

The direct addition of hydrophobic parameters has not been entirely successful in calculating log P. This is due, in part, to the difficulty in understanding the influence of specific structural features on partitioning. Leo [5] demonstrated that the partitioning of organic molecules into 1-octanol could only be treated adequately if one considered specific structural information of the solutes:

(a) steric factors, i.e., branching and chain length;
(b) proximity factors of nearby polar groups;
(c) bond factors;
(d) stereochemical factors and hydrogen bonding; and
(e) tight water binding.

Later, specifically bound water molecules that are also carried into 1-octanol were believed to have an effect on partitioning [6].

Bond factors

It has been shown that it takes more energy to create a cavity (i.e., solubilize a solute) in water, due to disruption of ordered (iceberg) water, than to create a similar-sized cavity in octanol. Consider first straight-chain alkanes. In order to simplify understanding by eliminating polar interactions, the log P increase is larger going from C1 to C2 than that observed for each additional methylene group from C3 to C8 (where it increases linearly with the volume and size of the cavity needed to accommodate each additional methylene moiety). A bond factor that decreases lipid solubility for each methylene group beyond C2 by −0.12 log units adequately

accounts for the increased flexibility of the larger aliphatic chains that reduces the degree of order in the aqueous solvation shell. It has also been noted that charged nitrogens behave in an even more dynamic manner. For example, lysine and arginine methylene groups carry a significant polar contribution that is transmitted through the chain, with the greatest decrease in lipid solubility being nearest the charged ammonium ion. Bond factors that express this graded reduction in lipid solubility for the protonated primary amines are −0.78 (first methylene bond), −0.40 (second methylene bond), −0.26 (third methylene bond), and −0.19 (fourth methylene bond) [5]. It has been demonstrated that lysine methylene protons are sufficiently polar to form novel interactions with polar oxygens in hemoglobin [6,26].

Branching factors

To illustrate the complexity involved in partitioning, consider the steric and partitioning differences between *t*-butanol and *n*-butanol. These molecules have the same number of lipophilic atoms, yet *t*-butanol is soluble in water and *n*-butanol forms a biphasic system with water. At first, one might think that *t*-butanol would be more lipid-soluble than *n*-butanol, since the polar hydroxyl group of *t*-butanol is shielded by nonpolar methyl groups that would be repelled by the polar aqueous medium in partitioning. However, the formation of a water cage around *t*-butanol, that cannot form as readily with *n*-butanol, permits solubilization of *t*-butanol in aqueous media. The other effects, i.e., proximity factors, bond factors and stereochemical factors are also important. Each alkyl chain branching is observed to decrease lipid solubility by −0.13 log units, while a polar group branch, such as OH, decreases lipid solubility by −0.22 log units [5].

Polar proximity factors

Polar proximity factors result from two or more polar groups in close proximity that shield each other from lipid interactions. This is an extremely important phenomenon that has not been accounted for adequately in many methodologies that sum fragments or atom primitives to calculate partition coefficients.

An example of the influence of polar proximity effects on transfer phenomena has been presented by Roseman [7]. It was observed that the hydrophilicity of polar amino acid side chains is markedly reduced by flanking peptide bonds, in agreement with the polar proximity effect described by Leo [5]. This study verifies the need to consider specific structural features that cause deviation from additive hydropathy scales. Roseman suggests that valid hydropathy scales can only be obtained with model peptides. However, it should be pointed out that, if the 3D structure is known, the increase in hydrophobicity due to polar proximity effects can be estimated to a reasonable degree. Roseman also provides evidence that the spontaneous insertion of polypeptides into membranes is likely to occur much more readily than was previously thought, due to this increased hydrophobicity of adjacent polar groups.

HINT

The molecular modeling program HINT was created for the use of atomic transfer energies to evaluate docking and protein folding as stated by Abraham and Leo [6] and was first used to evaluate and visualize the X-ray crystallographic binding interactions (hydrophobic and polar) of allosteric modifiers to hemoglobin [18].

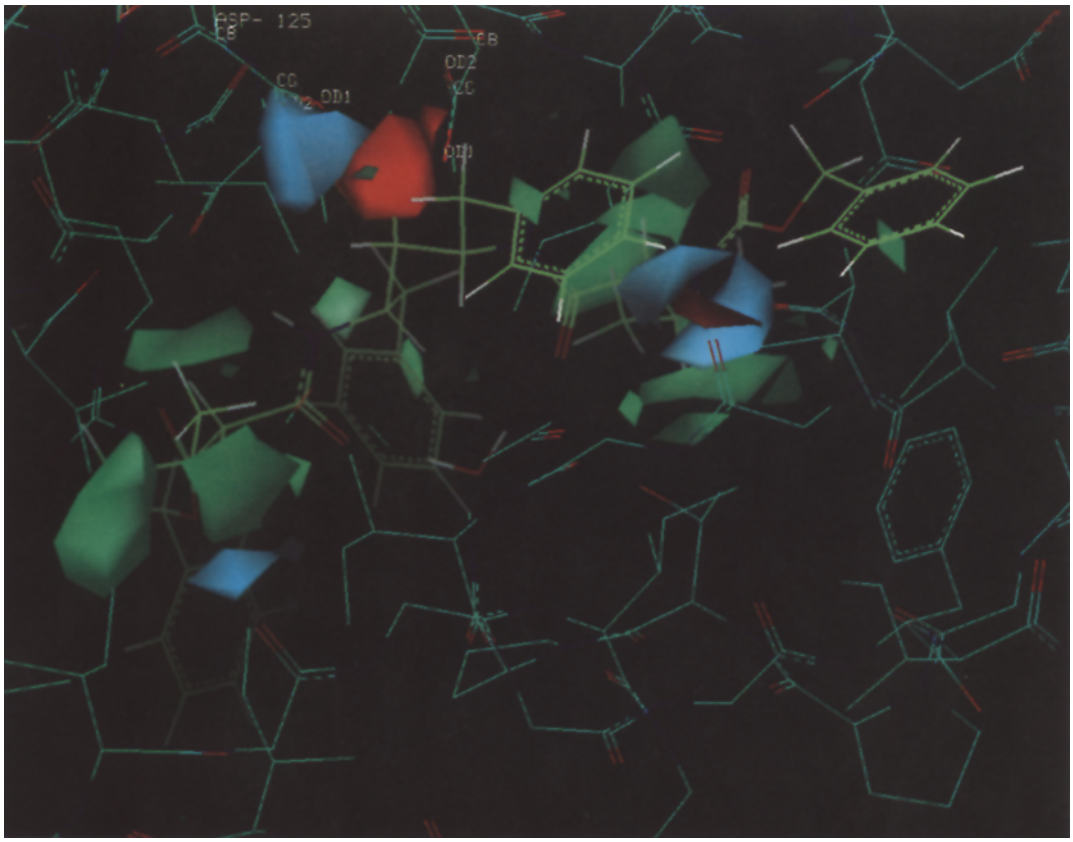

Fig. 1. HINT interaction map between HIV-1 and HIV-1 inhibitor A74704. Green contours represent regions of favorable hydrophobic binding, red contours represent regions of unfavorable polar interaction (e.g., acid–acid or base–base); blue contours represent regions of favorable polar interaction (e.g., hydrogen bonds, Coulombic, etc.). The protein model for HIV-1 here has both Asp^{25} and Asp^{125} ionized.

The original concept in developing a HINT hydropathic field that expresses an interaction between two atoms was to mimic electrostatic interactions by replacing the partial atomic charges by atomic hydropathy constants [27,28], so that the energy of interaction would be:

$$E \propto a_1 a_2 / d^2 \tag{1}$$

where a_1 and a_2 represent atomic hydropathy constants for two interacting atoms, including the Leo correction constants for structural features, and d represents the distance between the two interacting atoms. It was found that using d or d^2 weighted long-range interactions too strongly in the absence of an arbitrary distance cutoff [29]. The modifications described below were made to the equation, resulting in the HINT program.

HINT calculates a double sum for the total protein–substrate or protein–protein interaction:

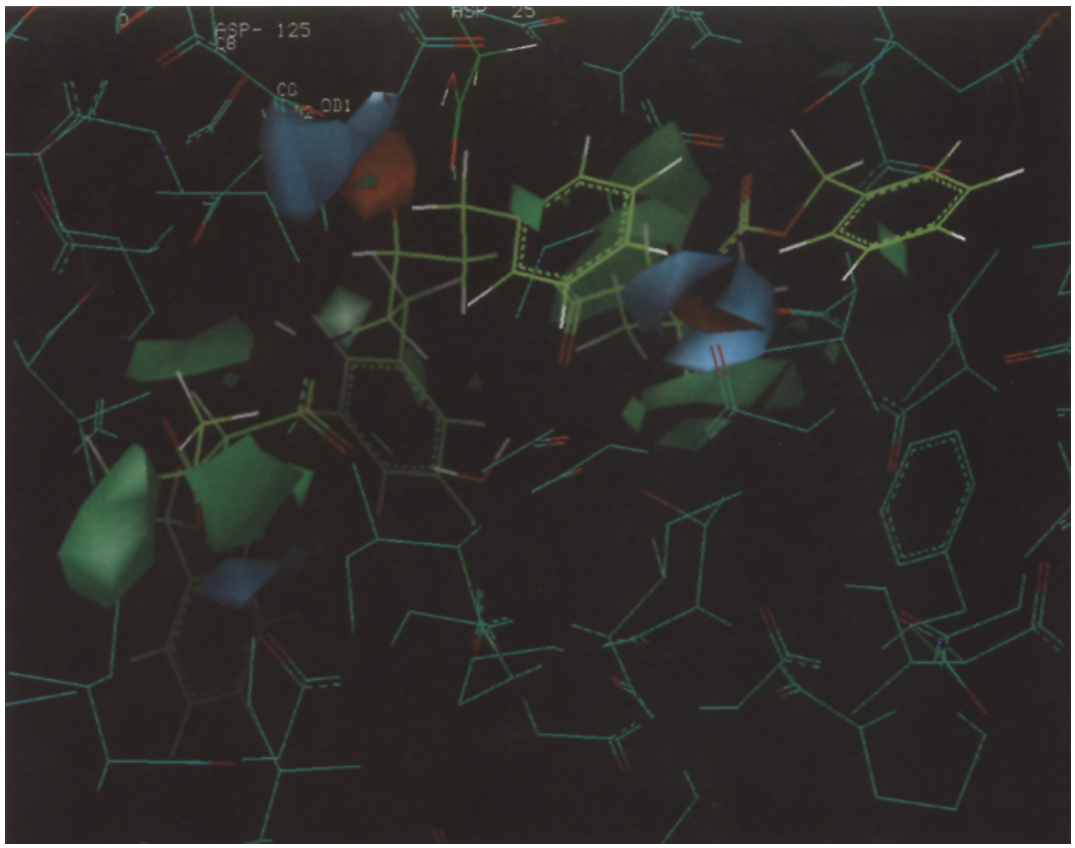

Fig. 2. HINT interaction map between revised HIV-1 model and inhibitor A74704. The model for HIV-1 here has OD2 of Asp^{25} protonated while Asp^{125} is ionized. The enhanced polar interaction between HIV-1 and A74704 is due to the additional hydrogen bond gained in this model.

$$B = \sum\sum b_{ij} = \sum_{j=1}^{protein} \sum_{j=1}^{drug} (s_i a_i s_j a_j R_{ij} + r_{ij}) \tag{2}$$

where b_{ij} is a 'microbinding parameter' specific to the atom pair ij, s is the solvent-accessible surface area, a is the hydropathic atom constant, and R and r are functions of the distance between atoms i and j. The hydropathic atom constants were obtained by further reduction of Leo fragment constants [5] derived from the water/octanol partition coefficient. It is important to note that these atomic constants contain implicit contributions from bond, chain, branch, and polar proximity factors that have been ignored in other approaches.

Positive-signed atom (or fragment) constants indicate hydrophobic atoms (or fragments), while negative-signed constants indicate hydrophilic atoms (or fragments). Atom-based hydropathic constants are constructed from fragment (group) constants such that:

(a) the sum of hydropathic atom constants in a group is consistent with the group's fragment constant;

(b) frontier atoms in a group are more important than shielded (central) atoms and are generally maintained at atom values; and
(c) bond, chain, branch and proximity factors are applied as additive constants; the first three to all eligible atoms, while proximity is applied to the central atom of polar groups.

Partition coefficients (sum of hydrophobic atom constants) for small molecules calculated by HINT are similar to values calculated by Hansch and Leo [5]. Hydropathic atom constants for nonterminal amino acid residues are computed from the corresponding *N*-acetyl amino acid analogs [17].

The functional form of the range dependence is described by two terms (R_{ij} and r_{ij}, Eq. 2). The former scales the hydrophobic atom constant/solvent-accessible surface area product with distance, while the second is independent of hydrophobicity and responds only to distance variations. Typically R_{ij} is

$$R_{ij} = T_{ij}e^{-r} \tag{3}$$

where T_{ij} is a sign-flip function which rewards acid–base and penalizes acid–acid or base–base interactions, and r is the distance between atoms i and j. Interestingly, e^{-r}, where r is the distance in Angstroms, is a reasonable fit [17] to the Leo polar proximity factors [5] for intramolecular through-bond polar–polar interactions if they are plotted on a distance scale rather than a bond-count scale. r_{ij}, which adds a Lennard-Jones-type potential, is set as

$$r_{ij} = A\varepsilon_{ij}[(r_{vdW}/r)^{12} - 2(r_{vdW}/r)^{6}] \tag{4}$$

Here A is an empirical scaling factor between the hydropathic term (i.e., exponential in Eq. 3) and Lennard-Jones terms, defined by Levitt and Perutz [30,31], ε_{ij} is the depth of the Lennard-Jones potential well, and r_{vdW} is the sum of van der Waals radii for atoms i and j.

Superimposition of a 3D grid of test atoms over the molecule or region of interest allows the calculation of a hydropathic field. The test atoms are assumed to have solvent-accessible surface areas, hydrophobic atom constants, and other parameters set to unity. Maps, displaying graphically the polar and hydrophobic interactions between the protein and substrate/drug, are calculated by an adaptation of Eq. 5:

$$A_t = \sum\sum[(s_i a_i R_{it})(s_j a_j R_{jt}) + r_{it} r_{jt}] \tag{5}$$

where A_t is the map value at the test point t, and R_{it}, R_{jt}, r_{it} and r_{jt} are the distance functions between atoms i (or j) and the test point.

As HINT has developed over the last few years, unique and useful capabilities have been added to the program for facilitating drug design and for understanding biomacromolecular structure. Specifically, HINT provides tools for modeling biological/chemical interactions in terms of hydrophobicity and hydropathy [19]. HINT can be used to map the hydropathic nature of unknown species from their complements (i.e., the 'key' from a receptor structure and the 'lock' from a ligand structure) [20]. Three-dimensional maps of interactions between ligands and

receptors, or between subunits of a single protein can be calculated, contoured and displayed by HINT. These maps are particularly useful for visualization of the effects of structural perturbation of the ligand by functional group changes or of the protein by site-directed mutagenesis. An example of this use will be presented below with an examination of the binding of a designed substrate (A74704) [32] in HIV-1. HINT estimates the relative strength of individual interactions between interacting species, and sums the individual interactions to obtain a binding score estimation. In series of related molecules, where crystallographic structural data for both the ligand and receptor are known, this calculated binding score correlates with the measured affinities [19]. Another use of HINT is to provide an empirical hydrophobic field for use with CoMFA. Further discussion of the HINT/CoMFA interface is presented below.

HINT examination of the binding of A74704 with HIV-1

The designed inhibitor A74704 [32] binds with an affinity of 4 nM to HIV-1. In order to understand the specific interactions that contribute to this binding, a HINT study was undertaken on this and several other related inhibitors. Here we describe the binding of A74704 to the protease. A later work will completely detail this study [33]. The HINT interaction map for these two species is shown in Fig. 1. The HINT interaction binding score for this system is 642.9. The green contours represent the regions where strong hydrophobic–hydrophobic interactions occur. In particular, these are associated with the methyl groups of the two valine-like groups in the inhibitor that interact with a pair of strong hydrophobic pockets in the HIV-1 protease. Blue contours are associated with favorable polar interactions between the species; the largest of these interactions are hydrogen bonds, e.g., between the core hydroxyl of the inhibitor and Asp^{125} of the protease, which is the largest single interaction in this system. Red contours, while not especially prominent in this system, represent unfavorable polar interactions. In many cases, these interactions are among the most interesting, because they suggest where changes can be made in either the substrate or receptor (through site-directed mutagenesis) to improve binding.

Unfavorable HINT polar interactions also may suggest where the molecular model can be improved. In the HIV-1/A74704 system, it may be appropriate to consider more closely the protonation state of the aspartic acids (positions 25 and 125) that interact strongly with the core hydroxyl moiety. The initial model of Fig. 1 assumed that both acids were ionized, as were all acids in HIV-1. However, consider protonation of OD2 of Asp^{25}. This would add a second hydrogen bond between the core hydroxyl and the protein. HINT maps for this modification of the model are shown in Fig. 2 (same color scheme as described above) and the HINT interaction binding score for this modification to the HIV-1/A74704 system is 873.7.

HINT hydropathic fields and CoMFA

The Comparative Molecular Field Analysis technique (CoMFA) has, in the few years since it was introduced by Cramer [21], become one of the pre-eminent tools for integrating QSAR studies with molecular design of new lead molecules in drug discovery. Basic CoMFA combines 3D structural information in the form of steric and electrostatic fields with a partial least-squares statistical evaluation of physical, chemical and biological properties. In many documented cases this has provided significant information, aiding in the design of new therapeutic agents.

A logical extension of CoMFA that we have addressed with HINT was to include a hydrophobic field along with the steric and electrostatic fields, since hydrophobicity is known to

play an important role in increasing biological activity in a large number of systems. In 1991, Kellogg et al. [22] first combined the HINT hydropathic fields with the CoMFA steric and electrostatic fields, using the same set of steroid structural and biological data used by Cramer et al. [21]. This work showed that the visual representation of the hydrophobic field offered increased interpretability of CoMFA models, even though there was no significant improvement in the statistical measures of the CoMFA model. One notable point is that steroid activity does not appear to correlate with hydrophobicity in traditional QSAR studies.

Since then, more convincing examples of the use of the HINT field to supplement CoMFA steric and electrostatic fields have been presented. DePriest [34] performed CoMFA on 52 inhibitors of thermolysin with MNDO (MOPAC 6.0) charges. The addition of the hydrophobic field increased the cross-validated r^2 from 0.5 to 0.66 and reduced the press from 1.56 to 1.32. Additionally, the ability of the derived model to predict the pK_i values of an additional set of 11 inhibitors was significantly improved. Nayak and Kellogg [35] calibrated the HINT hydropathic field by varying some of the HINT parameters in a three-field CoMFA study of the binding of barbiturates to cyclodextrins (CD). The best three-field model for binding to α-cyclodextrin had a cross-validated r^2 of 0.775. In addition, hydrophobic (one-field) CoMFA models show significantly higher correlation for binding to α-CD than to β-CD. This agrees with suggestions that the smaller α-cyclodextrin can only accommodate the hydrophobic side chain of the barbiturates, but the cavity of β-cyclodextrin is larger and attempts to fit the entire barbiturate molecule.

CONCLUSIONS

The driving force behind hydrophobic interactions is encoded in the partition coefficient. Therefore, the potential use of hydropathy-based interaction models, derived from experimental transfer data in drug design, protein folding and substrate interactions with macromolecules, may be a productive approach to understanding noncovalent interactions in biological systems. In this regard, HINT and similar programs, based upon atomic contributions to transfer energies, could play a significant role. However, there is much that we do not yet know about the details of partitioning. For example, field effects through space are essential and must be included. Generalized parameters for atom types alone will not be sufficient. The correction factors that Leo introduced for calculating log P are based upon thousands of measurements; their experimental validity is well established. They are indicators of specific atom–atom interactions and the influence of these interactions on the degree of increase or decrease of the transfer energy. Chain branching that decreases or the polar proximity effects of polar groups across space that increase the energy of hydrophobic interactions must be considered in any program that is designed to produce information that correlates with observed transfer phenomena. The effect of specifically bound water molecules to drugs or amino acid side chains involved in binding also must be considered in mapping out the forces involved in partitioning. We have tried to accommodate all of these features into HINT.

In conclusion, it should be pointed out that most quantum and molecular mechanics calculations are based on first principles of wave mechanics or neutron physics. By using experimentally determined transfer phenomena, HINT provides molecular interactions largely based on empirical observations.

REFERENCES

1 Abraham, D.J. and Kellogg, G.E., In Kubinyi, H. (Ed.) 3D QSAR in Drug Design: Theory, Methods and Applications, ESCOM, Leiden, 1993, pp. 506–522.
2 Tanford, C., The Hydrophobic Effect: Formation of Micelles and Biological Membranes, Wiley, New York, NY, 1980, pp. 1–233.
3 Fauchere, J.L. and Pliska, V.E., Eur. J. Med. Chem., 18 (1983) 369.
4 Creighton, E., Proteins – Structures and Molecular Properties, Freeman, New York, NY, 1993, pp. 157–163.
5 Hansch, C. and Leo, A.J., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, NY, 1979, pp. 1–339.
6 Abraham, D.J. and Leo, A.J., Proteins, 2 (1987) 130.
7 Roseman, M.A., J. Mol. Biol., 200 (1988) 513.
8 Eisenberg, D., Wilcox, W. and McLachlan, A.D., J. Cell. Biochem., 31 (1986) 11.
9 Nicholls, A., Sharp, K.A. and Honig, B., Proteins, 11 (1991) 281.
10 Fujita, T., Iwasa, J. and Hansch, C., J. Am. Chem. Soc., 86 (1964) 5175.
11 Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
12 Abraham, D.J., Intra-Science Chem. Rep., 8 (1974) 1.
13 Ghose, A.K. and Crippen, G.M., J. Comput. Chem., 7 (1986) 565.
14 Eisenberg, D. and McLachlan, A.D., Nature, 319 (1986) 199.
15 Audry, E., Dubost, J.P., Colleter, J.C. and Dallet, Ph., Eur. J. Med. Chem., 21 (1986) 71.
16 Furet, P., Sele, A. and Cohen, N.C., J. Mol. Graphics, 6 (1988) 182.
17 Fauchere, J.L., Quarendon, P. and Kaetterer, L., J. Mol. Graphics, 6 (1988) 203.
18 Wireko, F.C., Kellogg, G.E. and Abraham, D.J., J. Med. Chem., 34 (1991) 758.
19 Kellogg, G.E., Joshi, G.S. and Abraham, D.J., Med. Chem. Res., 1 (1992) 444.
20 Kellogg, G.E. and Abraham, D.J., J. Mol. Graphics, 10 (1992) 212.
21 Cramer III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
22 Kellogg, G.E., Semus, S.F. and Abraham, D.J., J. Comput.-Aided Mol. Design, 5 (1991) 545.
23 Goodford, P.J., J. Med. Chem., 28 (1985) 849.
24 Folkers, G., Merz, A. and Rognan, D., In Wermuth, C.G. (Ed.) Trends in QSAR and Molecular Modelling 92 (Proceedings of the 9th European Symposium on Structure–Activity Relationships: QSAR and Molecular Modelling), ESCOM, Leiden, 1993, pp. 233–244.
25 Norinder, U., J. Comput.-Aided Mol. Design, 4 (1990) 381.
26 Perutz, M.F., Fermi, G., Abraham, D.J., Poyart, C. and Bursaux, E., J. Am. Chem. Soc., 108 (1986) 1064.
27 Abraham, D.J., (1983) unpublished work.
28 Abraham, D.J. and Lesk, A., (1985) unpublished work.
29 Kellogg, G.E. and Abraham, D.J., (1989) unpublished work.
30 Levitt, M., J. Mol. Biol., 168 (1983) 595.
31 Levitt, M. and Perutz, M.F., J. Mol. Biol., 201 (1988) 751.
32 Erickson, J., Neidhart, D.J., VanDrie, J., Kempf, D.J., Wang, X.C., Norbeck, D.W., Plattner, J.J., Rittenhouse, J.W., Turon, M., Wideburg, N., Kohlbrenner, W.E., Simmer, R., Helfrich, R., Paul, D.A. and Knigge, M., Science, 249 (1990) 527.
33 Kellogg, G.E., Burt, S., Gussio, R., Erickson, J.E. and Abraham, D.J., manuscript in preparation.
34 DePriest, S.A., American Crystallographic Association Meeting, Albuquerque, NM, May 23–28, 1993, Abstract G002.
35 Nayak, V.R. and Kellogg, G.E., Med. Chem. Res., (1994) in press.