KEY, LOCK, and LOCKSMITH: Complementary hydropathic map predictions of drug structure from a known receptor–receptor structure from known drugs

Glen Eugene Kellogg and Donald J. Abraham

Department of Medicinal Chemistry and Division of Biomedical Engineering, Virginia Commonwealth University, Richmond, VA, USA

Three new routines (LOCK, KEY and LOCKSMITH) for the program HINT (hydrophobic interactions) are described and demonstrated. The KEY routine uses receptor structure to model the hydropathic profile of the ideal substrate for the receptor. The LOCK routine uses substrate or drug structure to model the hydropathic character of the receptor. LOCKSMITH is an algorithm designed to highlight the significant hydropathic features from a collection of agents. Ten allosteric modifiers of hemoglobin that have been characterized biologically and with X-ray diffraction to determine their protein binding sites/conformations illustrate the KEY and LOCKSMITH routines: The LOCKSMITH composite map correctly identifies the structural features and conformation of the more active modifiers. In addition, many hydropathic features of the "ideal" drug predicted by the KEY map overlap with actual structural features of the most active hemoglobin allosteric modifiers.

Keywords: hydrophobicity, hydropathy, QSAR, allostery, HINT, complementarity

INTRODUCTION

One of the long-standing underlying principles of drug design is that there is a well-defined relationship between the drug and its receptor, and that this relationship is similar to the way that a key fits in a lock.[1] More recently, shape selectivity has become a major factor in computer-assisted drug design methods for QSAR (quantitative structure activity relationships) studies[2,3] and in comparative molecular field analysis (CoMFA).[4,5] We have previously described our modeling program HINT (hydrophobic interactions) as a tool to calculate and visualize the hydrophobic field of small molecules and proteins, and to graphically and numerically quantify hydrophobic interactions between small molecules and proteins.[6] In our earlier work, we presented results from a combined modeling–crystallography study of the binding of allosteric modifiers of hemoglobin. The HINT-calculated hydropathic maps for the allosteric effector molecules (in their X-ray determined bound conformations) and for the native deoxy-hemoglobin (dxHb) matched in a number of key locales, suggesting the importance of the associated interactions.[6] However, the luxury of having knowledge of both the protein and substrate structures is very rare in drug design; most often only the structures of active agents are known, and far less commonly the structure of the protein receptor is known. In the latter case, shape-complementarity algorithms to design substrates from receptor structure using hard sphere repulsions have been described by Kuntz et al.[7–9]

In this contribution we describe the use of HINT to calculate complementary hydrophobic fields, that is, fields that represent the implied interacting species:

(1) From a drug structure and its hydropathic profile, predict the hydropathic profile and elements of the structure of the "ideal" receptor for the drug (termed the *lock*).
(2) From a known receptor structure predict the structure of the "ideal" drug (termed the *key*).

We also propose a scheme (LOCKSMITH) with which a collection of agents, with varying biological activities at the same protein receptor site, can be integrated into a single entity that encodes the activity information of the collection.

Color Plates for this articles are on page 226.

Address reprint requests to Dr. Kellogg at the Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298-0540, USA.

Received 11 July 1991; accepted 2 March 1992

The HINT Model

Hydrophobicity is an empirical, phenomenological property of molecules that encodes significant thermodynamic information about the molecule's interaction with its environment. The experimental measure of hydrophobicity is the solvent partition coefficient (usually expressed as log P) of a molecule between two phases (aqueous and organic). Hydrophobicity has been an important parameter of QSAR and drug design for many years,[10–14] because in its most elemental sense, hydrophobicity indicates the solubility and transport of a drug in the biological system.

The HINT model of molecular interactions is based on the notion that solubility data are just another way of looking at intermolecular interactions between the molecule and solvent. Hydrophobic molecules are attracted to nonpolar solvents, while hydrophilic molecules are attracted to polar solvents such as water. Likewise, in choosing a binding site/receptor, these molecules would be attracted to analogous regions within the biological system. The key step is reduction of molecule-level and fragment-level partition data (termed *hydrophobic fragment constants*[13]) to hydrophobic *atom* constants that represent the empirical hydrophobic and interaction-propensity properties of individual atoms in the system.[6, 14, 15] Signed hydrophobic atom constants model all noncovalent interactions, including polar interactions like hydrogen bonding and Coulombic attractions and repulsions. It is significant that while positive hydrophobic atom constants represent hydrophobic atoms and negative constants represent hydrophilic or polar atoms, charged species have larger magnitude (more negative) atom constants in proportion to their likelihood of involvement in polar noncovalent interactions.

To model the reduced effect of internal shielded atoms and the enhanced importance of terminal frontier atoms with respect to interactions, the hydrophobic atom constants are scaled with atomic solvent accessible surface areas (water probe). HINT models the distance dependence of the hydrophobic effect in the biological environment as a linear combination of two functions: an exponential relationship between the hydrophobic atom constants of interacting atoms,[16] and a Lennard–Jones (6–12) type function for the dispersion contribution.[17, 18]

Thus, the interaction (b_{ij}) between two atoms (i, j) is described as

$$b_{ij} = s_i a_i s_j a_j R_{ij} + r_{ij}$$

where s is the solvent accessible surface area, a is the hydrophobic atom constant, and R and r are functions of the distance between atoms i and j.

$$R_{ij} = T_{ij} e^{-r}$$
$$r_{ij} = A\,\epsilon_{ij}[(r_{vdw}/r)^6 - 2(r_{vdw}/r)^{12}]$$

The sign-flip function T_{ij} examines each atom–atom interaction for external factors that are not otherwise encoded in the hydrophobic atom constants. This function detects and corrects the sign of the hydrophobic-dependent contribution to b_{ij} for interactions like unfavorable polar–polar (e.g., acid–acid or base–base), hydrogen bonding, etc. The quantity A is a scaling factor between the two contributions to b_{ij}, while ϵ_{ij} is a constant representing the energy of the dispersion interaction between atoms i and j.[17, 18] The HINT hydrophobic interaction can be examined as a fieldlike quantity by assuming a grid of test points (t) with unit hydrophobic atom constants and solvent accessible surface areas, and trivial Lennard–Jones interaction constants (i.e., $s_j = a_j = 1$, $\epsilon_{ij} = 0$),

$$A_t = \sum_{i=1}^{\text{atoms}} s_t a_t R_{it}$$

This family of equations represents the hydrophobic field of a small or protein molecule. The maxima for the field are localized at the nucleus of each atom ($r = 0$) as shown in Figure 1a.

KEY and LOCK

To use the hydrophobic field of a known molecule to predict the field of a complementary entity, we assume that the hydropathic nature of the complementary species is the same as the defining atoms; that atoms in the complementary species have the same van der Waals radii as those in the defining species; and that the optimum atom–atom distance between species is the sum of the van der Waals radii. The functional form of this distance behavior is

$$R_{ij} = e^{-|2r_{vdw} - r|}$$

as shown in Figure 1b for two values of r_{vdw}. Thus, the regions with the greatest hydropathic character in KEY or LOCK maps are those $2r_{vdw}$ from the defining atom set. Also, to direct complementary hydropathic density to unoccupied space, the hydrophobicity of grid points within one van der Waals radius of any existing atom are set to zero (see Figure 1b).

The resulting complementary map is an eggshell encasing the protein receptor (for the KEY function) or drug (for the LOCK function). Figure 2a sets out a two-dimensional (2D) contour slice for the LOCK map of acetone. No preference is shown for attack direction, and in fact, this map suggests that the carbonyl carbon may interact with a LOCK atom on the oxygen side of the C=O bond. Adding the asymmetry

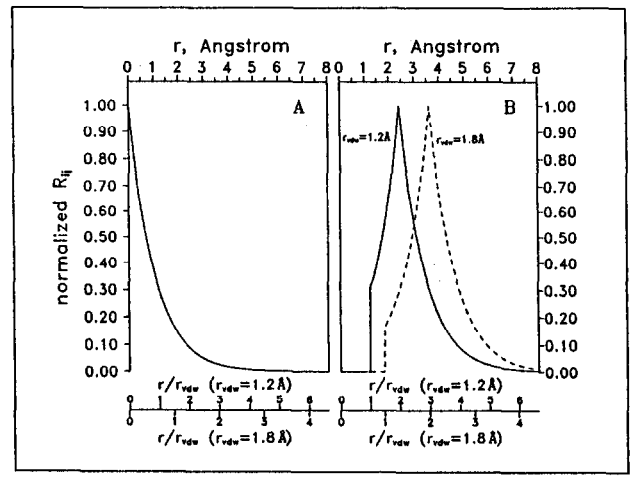
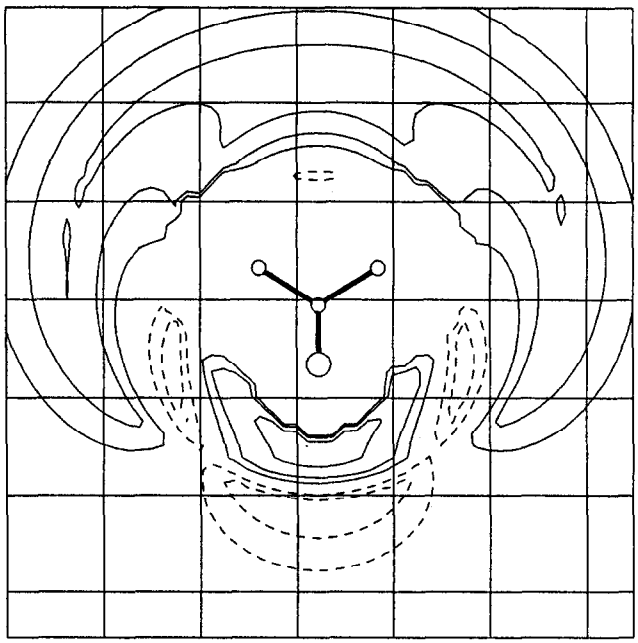
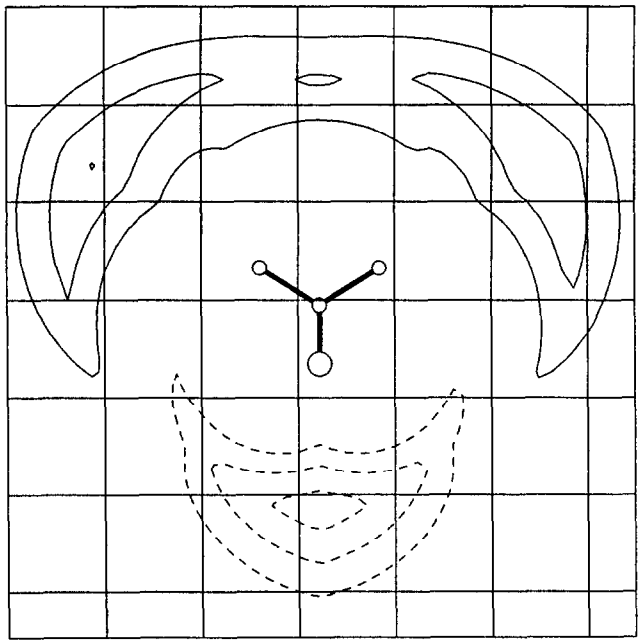


Figure 1. HINT distance functions for (a) molecular hydrophobicity; and (b) complementary hydrophobicity.



(a)



(b)

Figure 2. Two-dimensional LOCK maps of acetone (heavy atom plane): (a) f = 0; (b) f = 1. Contour levels ±12, ±24, ±48; solid lines—hydrophobic, dashed lines—polar, grid lines at 2 Å.

function,

$$F_{it} = e^{-f\theta_{it}}$$

where f is a nonnegative input parameter defining the extent of asymmetry, and θ_{it} (Figure 3) is an angle between the two vectors—the vector from atom i to the atoms bonded to it (a bond vector), and the vector from atom i to grid point t, yields the field definition equation

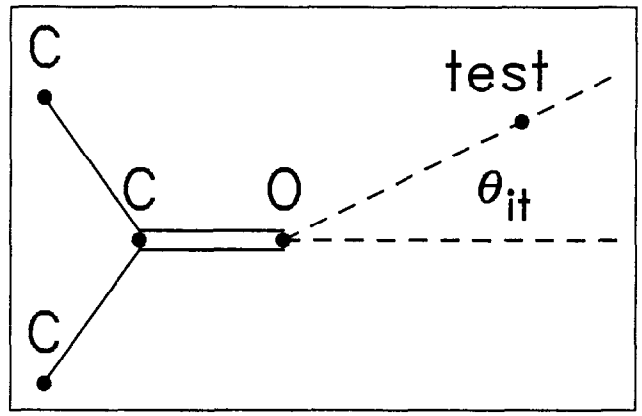


Figure 3. Definition of θ_{it}.

$$A_t = \sum_{i=1}^{\text{atoms}} s_i a_i R_{it} F_{it}$$

The effect of adding the F_{it} term to the field definition equation is shown by the 2D contour slice in Figure 2b, where $f = 1$ in F_{it}.[19]

LOCKSMITH

The combination of biological data and structural parameters from a series of molecules active at the same receptor site is the basis of QSAR. Thus, a statistical model of activity (which may be difficult to measure) as a function of a structural index (which is easy to measure or predict) can be derived and utilized to select or design biologically active molecules. The extension of QSAR to three dimensions with CoMFA and other field-based techniques has recently been employed in drug design.[4, 5, 20, 21] CoMFA is a three-dimensional (3D) QSAR technique that utilizes calculated steric and electrostatic fields as the input structural parameters to derive a model of biological activity.[4] Partial least-squares (PLS)[22] and cross-validation methodology enable the CoMFA model derivation, due to the very large ratio of independent variables to dependent variables in the CoMFA equation. Previously, we have shown that the HINT hydrophobic field may be a valuable addition for drug design purposes to the steric and electrostatic fields of standard CoMFA.[5] The empirical hydrophobic field was found to be potentially useful for drug design due to the chemically relevant information content of the hydrophobic field.

The LOCKSMITH scheme is a 3D nonstatistical method to exploit the hydropathic maps of a series of compounds for insight on: possible structural modifications within the series to obtain the most active compound; and hydropathic structure of the receptor. LOCKSMITH requires some form of molecular overlap rule of the involved molecules, much as does CoMFA.[4] The procedure is to sum the activity-scaled HINT small molecule or LOCK maps for each molecule in the series at each grid point and obtain a hydropathic map for

the composite,

$$A_t^* = \frac{\sum_{k=1}^{\text{drug}} A_t(k)\sigma(k)}{\sum_{k=1}^{\text{drug}} \sigma(k)},$$

where $\sigma(k)$ is the functionalized biological activity of drug k. The success of this scheme depends on two factors: molecular overlap and appropriate functionalization of the biological activity. The former factor is common to most, if not all, 3D QSAR techniques, and will not be discussed here. Functionalization of biological activity refers to construction of an appropriate scale for the measured data that allows map interpretation:

(1) All $\sigma(k)$ should be positive.
(2) $\sigma(k)$ for "active" compounds should be numerically greater than for "inactive" compounds.
(3) the dynamic range of biological activities should be within one or two orders of magnitude.

These functionalization "rules" are designed to give appropriate weighting to the biological data for composite map construction. The simplest functionalizations would be $\log \sigma$ or $-\log \sigma$, depending on Rules 1 and 2.[23] Note that functionalization is not as foreign a concept as it sounds—partition coefficients and binding constants are usually reported as logarithmic values, and these are often the forms imported into QSARs.

RESULTS

The LOCKSMITH and KEY routines have been applied to the set of crystallographically determined allosteric modifiers of hemoglobin (see structural classes 1–4).[6, 24–26] This set of molecules permits evaluation of the techniques because the receptor structure and bound conformations of the effectors are known. In this example, the native protein atomic coordinates and the X-ray determined small molecule atomic coordinates at the binding site were used; i.e., the molecular overlap rule is from actual diffraction results. The coordinate set for the drug molecules used was for those bound at the primary site associated with LYS 99α_1. Color Plate 1 is a stereo view of three of the modifiers (2b, 3b, and 4b), color coded by atom type (white—carbon, red—oxygen, blue—nitrogen, green—chlorine, and cyan—hydrogen) in their bound conformations superimposed on the native deoxy (*T*) hemoglobin coordinates. Molecules bound at the LYS 99α_2 primary site are related to the LYS 99α_1 molecules by a symmetry two-fold axis. Some of the allosteric effectors (3a, 4b and 4c) have additional (secondary) binding sites around ARG 104β_1 or ARG 104β_2.[6] For reference, four of the protein residues surrounding the modifier binding site (PHE 36α_1, LYS 99α_1, ASN 108β_1, ARG 141α_2) have been color-coded purple in Color Plates 1–3.

For the purpose of LOCKSMITH calculations, the allosteric activities (P_{50}(drug)/P_{50}(control), Table 1) were used for $\sigma(k)$ without further functionalization. The individual HINT hydropathic maps were calculated for each modifier with explicit polar hydrogens, united-model nonpolar hydrogens, $R_{ij} = e^{-r}$ and $r_{ij} = 0$, on a 1-Å grid. Color Plate 2 shows the allosteric activity-scaled hydrophobic composite map of the ten-member allosteric effector series calculated using the LOCKSMITH algorithm and contoured/displayed using[27] SYBYL 5.41 on an ESV-3 workstation. Green contours indicate regions of hydrophobicity, and red contours indicate polar regions. This map would be expected to display the hydropathic profile of a molecule possessing the best activity-enhancing structural features of the molecules in the collection. The structures for compounds 2b, 3b and 4b are shown superimposed on the map in Color Plate 2. Of the compounds listed in Table 1, 2b and 4b are the most effective allosteric modifiers,[28] and correspond most closely to the hydropathic features of the composite map. Compound 3b has the carbonyl directed in an opposite sense to that of 2b and 4b, and has significantly lower allosteric activity.

The complementary KEY map for hemoglobin in the region surrounding the allosteric effector binding site is shown in Color Plate 3. This map attempts to predict, based only on the molecular structure of native deoxy hemoglobin,

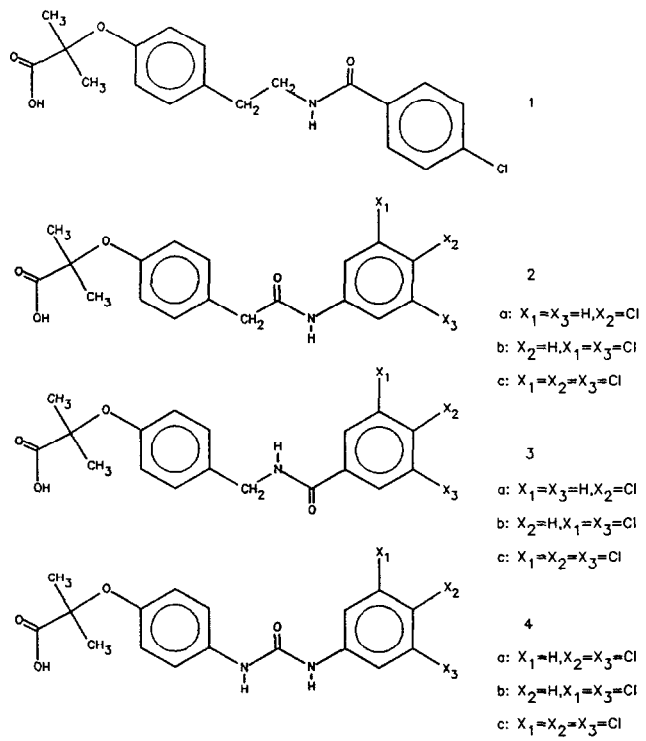


Table 1. Allosteric activity for right shifting agents.[6]

Compound	P_{50}(drug)/P_{50}(control)
1	1.74(6)
2a	2.79(4)
2b	4.32(8)
2c	3.17(9)
3a	2.58(9)
3b	2.48(5)
3c	2.13(8)
4a	1.77
4b	3.62(9)
4c	2.57(7)

the ideal drug for the site. In this map, red contours denote where the KEY should be polar, and green contours denote where the KEY should be hydrophobic. The map identifies the major hydrophobic regions of the effectors. There is a (green) contour encompassing the halogens (of the lower ring) and another contour spanning the acid (upper) ring and the methyls of the isobutyric acids. Two other major contact points between hemoglobin and the allosteric effectors are due to ARG 141α_2 interacting with the acid group, and LYS 99α_1 interacting with the amide carbonyl of the drug.[6] Both of these interactions are contingent on flexibility of the arginine and lysine residue sidechains. The KEY map indicates this flexibility in terms of the comparatively large areas "swept out" by the polar complementarity contours for both of these residues (in Color Plate 3, ARG 141α_2, upper right to middle right; LYS 99α_1 wrapping around modifiers from rear). It is important to emphasize that, because the protein coordinates used for the KEY calculation are from the *native* deoxy-hemoglobin structure, the likely optimization of side chains in the protein to maximize interactions with the bound drugs are not taken into account in the calculation presented in Color Plate 3. Also interesting to note is that two of the weaker effectors have their acid turned away from the ARG 141α_2 region[24, 26] and that the amide carbonyls of the compounds in series 3 (also weak effectors) are oriented away from the area accessible to LYS 99α_1.[6]

DISCUSSION

The HINT complementary (KEY and LOCK) maps may be useful predictors of an implied interacting species: either the "ideal" drug in the case of the KEY map or the receptor in the case of the LOCK map. This examination of allosteric effectors of hemoglobin is a particularly rigorous test of the KEY technique. Very few other biological systems have a known receptor/binding site structure and experimentally determined conformations for a variety of bound molecules. Clearly, however, the optimization of protein side-chain orientations concomitant with the substrate binding would likely lead to better agreements in KEY map/drug comparisons.

The LOCKSMITH summation of the activity-weighted effector hydrophobic maps correctly indicates the structure and orientation of the most active compounds. It also indicates, in terms of relative hydrophobicity, the important structural features of the series. This algorithm should be of general utility for drug design purposes. The two considerations for meaningful application of the scheme are molecular alignment and appropriate activity functionalization. It is important that the functionalized activities not prejudice the LOCKSMITH result by overweighting the more active compounds relative to the less active compounds. This is a matter for experimentation. LOCKSMITH maps may be considered as a nonstatistical hydrophobicity-based analogue of CoMFA coefficient maps. We are currently investigating algorithms to "score" map overlaps. It would be useful to have numerical, in addition to the graphical, predictors of drug or substrate structure.

Hydrophobicity is a key parameter of biological molecular systems.[12, 29–31] Because the biological environment is water-based, knowledge of the aqueous solubility of proposed substrates is of significant value. Even more important for drug design purposes is developing understanding of the intermolecular and intramolecular interactions within the system. Hydrophobicity, as an experimental measure of intermolecular interactions, provides valuable insight into biologically important polar and hydrophobic interactions. The design process for active agents must recognize and consider these factors. The KEY, LOCK, and LOCKSMITH functions of HINT may be useful in drug design for these reasons.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the National Institutes of Health (NIH HLBI-1 RO1 H1 32793) for this research. The SYBYL software has been made available through a University Software grant from Tripos Associates.

REFERENCES

1 Erlich, P. Chemotherapeutics: scientific principles, methods, and results. *Lancet* 1913, 445
2 Motoc, I., Dammkoehler, R.A., and Marshall, G.R. Three-dimensional structure–activity relationships and biological receptor mapping. in *Mathematical and Computational Concepts in Chemistry* (N. Trinajstic, Ed.) Horwood, Chichester, U.K., 1986, pp. 221–251
3 Marshall, G.R. Computer-aided drug design. *Ann. Rev. Pharmacol. Toxicol.* 1987, **27**, 193–213
4 Cramer, R.D., III, Patterson, D.E., and Bunce, J.D. Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 1988, **110**, 5959–5967
5 Kellogg, G.E., Semus, S.F., and Abraham, D.J. HINT—A new method of empirical hydrophobic field calculation for CoMFA. *J. Comput.-Aided Mol. Design.* 1991, **5**, 545–552
6 Wireko, F.C., Kellogg, G.E., and Abraham, D.J. Allosteric modifiers of hemoglobin. 2. Crystallographically determined binding sites and hydrophobic binding/interaction analysis of novel hemoglobin oxygen effectors. *J. Med. Chem.* 1991, **34**, 758–767
7 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Landridge, R., and Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* 1982, **161**, 269–288
8 DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D., and Venkataraghaavan, R. Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.* 1986, **29**, 2149–2153
9 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D., and Venkataraghaavan, R. Docking flexible ligands to macromolecular receptors by molecular shape. *J. Med. Chem.* 1988, **31**, 722–729
10 Meyer, H. Zur Theorie der Alkoholnarkose. Erste Mittheilung. Welche Eigenschaft Anästhetica bedingt ihre narkotische Wirkung? *Arch. Exp. Pathol. Pharmakol.* 1899, **42**, 109–118

11 Hansch, C. and Fujita, T. ρ-σ-π analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* 1964, **88**, 1616–1626
12 Akamatsu, M., Yoshida, Y., Nakamura, H., Asao, M., Iwamura, H., and Fujita, T. Hydrophobicity of di- and tripeptides having unionizable side chains and correlation with substituent and structural parameters. *Quant. Struct.-Act. Relat.* 1989, **8**, 195–203
13 Hansch, C. and Leo, A.J. *Substituent Constants for Correlation Analysis in Chemistry and Biology*. J. Wiley and Sons, New York, 1979, pp. 1–339
14 Abraham, D.J. and Leo, A.J. Extension of the fragment method to calculate amino acid zwitterion and side chain partition coefficients. *Proteins: Struct. Funct. Genetics* 1987, **2**, 130–152
15 Kellogg, G.E., Joshi, G.S. and Abraham, D.J. New Tools for Modeling and Understanding Hydrophobicity and Hydrophobic Interactions. *Med. Chem. Res.* 1992, **1**, 444–453
16 Fauchere, J.-L., Quarendon, P. and Kaetterer, L. Estimating and representing hydrophobicity potential. *J. Molec. Graphics* 1988, **6**, 203–206
17 Levitt, M. Molecular dynamics of native protein. I. Computer simulation of trajectories. *J. Mol. Biol.* 1983, **168**, 595–620
18 Levitt, M. and Perutz, M.F. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* 1988, **201**, 751–754
19 The F_{it} function directs complementary hydropathic density along bond vectors. A more sophisticated model would recognize the orbital hybridization of the defining atom set and optimize the LOCK or KEY maps with this information.
20 Allen, M.S., Tan, Y.-C., Trudell, M.L., Narayan, K., Schindler, L.R., Martin, M.J., Schultz, C., Hagen, T.J., Koehler, K.F., Codding, P.W., Skolnick, P., and Cook, J.M. Synthetic and computer-assisted analyses of the pharmacophore for the benzodiazepine receptor inverse agonist site. *J. Med. Chem.* 1990, **33**, 2343–2357
21 Norinder, U. Experimental design based 3-D QSAR analysis of steroid-protein interactions: Application to human CBG complexes. *J. Comput.-Aided Mol. Design* 1990, **4**, 381–389
22 Wold, S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 1978, **20**, 397–405
23 Other functional forms that should be considered include: $n - \log \sigma$, $n + \log \sigma$, σ^x, $(\log \sigma)^x$, etc.
24 Perutz, M.F., Fermi, G., Abraham, D.J., Poyart, C., and Bursaux, E. Hemoglobin as a receptor of drugs and peptides: X-ray studies of the stereochemistry of binding. *J. Am. Chem. Soc.* 1986, **108**, 1064–1078
25 Lalezari, I., Rahbar, S., Lalezari, P., Fermi, G., and Perutz, M.F. LR16, a compound with potent effects on the oxygen affinity of hemoglobin, on blood cholestrol, and on low density lipoprotein. *Proc. Natl. Acad. Sci. U.S.A.* 1988, **85**, 6117–6121
26 Lalezari, I., Lalezari, P., Poyart, C., Marden, M., Kister, J., Bohn, B., Fermi, G., and Perutz, M.F. New effectors of human hemoglobin: Structure and function. *Biochemistry* 1990, **29**, 1515–1523
27 Available from Tripos Associates, 1699 South Hanley Road, St. Louis, Missouri 63144 (USA)
28 Randad, R.S., Mahran, M.A., Mehanna, A.S., and Abraham, D.J. Allosteric modifiers of hemoglobin. 1. Design, synthesis, testing, and structure-allosteric activity relationship of novel hemoglobin oxygen affinity decreasing agents. *J. Med. Chem.* 1991, **34**, 752–757
29 Tanford, C. *The hydrophobic effect: Formation of micelles and biological membranes*. Wiley: NY, 1980, pp. 1–233
30 Kollman, P.A. The nature of the drug-receptor bond. In Wolff, M.E., Ed., *Burger's Medicinal Chemistry. Part 1. The Basis of Medicinal Chemistry, 4th Edition*. J. Wiley and Sons, New York, 1980, pp. 313–329
31 Franke, R. *Theoretical Drug Design Methods*. Elsevier, Amsterdam, 1984, pp. 1–412