Journal of Computer-Aided Molecular Design, 1 (1987) 257–264
ESCOM

J-CAMD 020

Distance geometry analysis of ligand binding to drug receptor sites

Gabriëlle M. Donné-Op den Kelder

Department of Pharmacochemistry, Vrije Universiteit, de Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

Received 7 May 1987
Revised 12 October 1987
Accepted 2 November 1987

Key words: Distance geometry; Receptor mapping; QSAR; Receptor interaction

SUMMARY

The method known as 'distance geometry approach' for receptor mapping procedures is discussed. In this method a ligand binding to a certain receptor is considered as a collection of ligand points. Binding sites of the receptor are either 'empty' or 'filled' site points; a ligand point might bind to an empty site point; filled site points indicate that at that point no binding is possible. A binding mode of a ligand is a list of which ligand points coincide with which empty binding sites. The applicability of the method for QSAR studies is discussed; as examples are mentioned the dihydrofolate reductase, β_1- and β_2-receptors. Finally, some ideas on future developments in receptor mapping are discussed.

DEDICATION

This article is dedicated to the late Dr. Teake Bultsma who introduced the distance geometry approach into our department.

INTRODUCTION

When detailed structural information about a receptor is not available, a model can be deduced from the ligands that bind to it. A number of statistical methods are commonly used in this case. These include Hansch analysis [1], discriminant and cluster analysis, principal component and factor analysis, nonlinear regression, pattern-recognition techniques [2–8] and distance geometry methods [9–15] that predict characteristics of a macromolecular binding site by relating the selected structural properties of active compounds to their biological activities. Deducing these so-called quantitative structure activity relationships (QSAR) has long been recognized as an important step in rational drug design. Possibly the best known method is that of Hansch and coworkers [1] where some measure of activity for a series of drugs is empirically correlated with physicochemical properties. However, little information is obtained concerning the size and shape of

the binding site. Only recently the global shape of the molecules was taken into account in the so-called 3-D structure-directed QSAR [9–19]. The distance geometry approach as originally developed by Crippen and co-workers [10] deduces the geometry and chemical nature of a macromolecular binding site, given only the chemical structures and observed free energies of binding for a series of ligands. This method was the first to treat conformational flexibility of ligands explicitly.

GENERAL CONCEPT OF THE DISTANCE GEOMETRY APPROACH

The distance geometry approach [9, 20–21] is a method of general use in conformational analysis. It has been applied to a number of widely different structural problems: the effect of internal and overall shape constraints on the range of allowed conformations of proteins [22], protein tertiary structure prediction [23], ligand binding to drug receptor sites [10], and using incomplete experimental data to determine the structure of ribosomes and antibiotics [24]. Using the distance geometry approach [9], Cartesian coordinates can be calculated for n points given the $n*n$ matrix of interpoint distances. This *distance matrix* contains geometric constraints given in terms of upper and lower bounds on the interpoint distances. A principal component analysis is used to determine the eigenvalues of the diagonalized metric matrix. By choosing the largest three eigenvalues, the coordinate axes are chosen along the largest three principal axes of the collection of points, thus attributing the least possible scatter to the fourth and higher dimensions.

Conformational analysis by *(scaled) energy embedding* [20–21] is an extension of the distance geometry approach in which an energy function of the pairwise distances is optimized in addition to meeting the geometric constraints. This increases the probability of finding low-energy structures or even the global minimum. The projection of the structure onto the three most important axes results in a compression of *all* the distances. This effect can be largely countered by introducing a scaling factor so that the geometric constraints are always satisfied.

DISTANCE GEOMETRY APPROACH TO RATIONALIZING BINDING DATA

A. Development of the method

The receptor mapping procedures as developed by Crippen [10] aid in the deduction of the nature of a receptor site given experimental data on the binding affinity of a series of ligands. It is tried to obtain a physically reasonable picture of the geometry of the site and plausible deductions as to the chemical character of various parts of the site. The ligands may be conformationally flexible and all are presumed to bind to the same, single site on the receptor molecule. In case the conformation of a ligand changes upon binding to accommodate the binding site, the free energy of such a conformational change is assumed to be small compared to the free energy of binding. The site itself may be slightly flexible, although no major conformational changes are permitted, and the energetic cost of any deformation is assumed to be negligible. The free energy of binding is taken to be approximately equal to the sum of the 'interaction energies' for all 'contacts' between parts of the drug and parts of the receptor site. Not considered are the conformational energy change of the ligand upon binding, nor its loss in translational energy and degrees of rotational freedom.

Basic concepts of the method as developed in 1979 by Crippen [10] are:

(1) Each ligand is represented as a collection of points called ligand points, which correspond to atoms or small groups of atoms. Conformational flexibility of the drug can be treated as upper and lower bounds on the distances between its atoms over all sterically allowed conformations.
(2) A binding site is proposed in terms of a number of 'empty or filled site points' positioned rigidly in space with respect to each other. An *empty* site point is a vacant place positioned where a ligand point may lie when binding takes place. A *filled* site point, however, indicates the position of some steric blocking group, and no ligand point may coincide with it during binding.
(3) A possible binding mode of some ligand amounts to simply a list of which ligand points coincide with which *empty* site points.
(4) The calculated free energy of binding is obtained in a simplified all-or-nothing fashion by adding up the contribution from each contact between a ligand point and a site point. A certain type of site point may be characterized as being a hydrogen bond donor or acceptor, or as a hydrophobic pocket. Each point-point interaction energy is taken to be the $\triangle G$ for the process: solvated ligand point + solvated site point → occupied site point. Thus, solvation, enthalpy and entropy are all included.

The problem of *finding the energetically optimal mode of binding* can be solved in a variable binding mode analysis, in which the energy parameters are evaluated by a quadratic programming algorithm [25].

In 1980, algorithms were developed [11] to aid in the decisions on the number of site points to be employed, their types, relative positions and interaction energies. A *decomposition algorithm* is used to determine the largest number of points common to all ligand molecules. Subsequently, the smallest set of substituent groups needed to account for all remaining points of all ligands is determined. Inherent to this method is that a plausible binding mode for each ligand must be selected in advance. On the basis of these specified binding modes, the site point distance bounds and their coordinates [9] are evaluated.

In 1984, another approach was put forward [26] to deduce the geometry of the site from the structures of the ligands. Site pockets were constructed relative to the ligand molecule on the basis of an 'active conformation' hypothesis. The active conformation can be hypothesized on basis of: (i) conformational analysis of strong and weak binding compounds; (ii) the structure of conformationally rigid bioactive molecules; and (iii) the distance constraints of the 'equivalent' atoms in different ligands having different structure or conformational behaviour.

The problem of finding geometrically allowed binding modes for a given site was reinvestigated by Crippen [26]. The basic idea of the newly developed algorithm is that only the *energetically* (i.e., sterically) allowed conformations are checked for matching with the proposed site (geometrically allowed binding modes) [27, 28]. The method was refined again in 1985 [29] to prevent binding modes being missed. A ligand molecule was no longer represented by *one* distance matrix, but by a collection of distance matrices, each generated from two consecutive conformations when they are sterically or energetically allowed.

In the 1980s Bultsma and co-workers developed receptor mapping procedures [13–15] which essentially are based on the method of Crippen. The possibility of alternating binding modes was excluded from the energy optimization program, while the method was refined at the point of interaction energies between site and ligand points [13]. Physicochemical parameters were introduced to describe the interaction between a site and ligand point. For example, if a site pocket has a hydrophobic character, it is not unrealistic to assume that the energy arising from interaction

with a ligand point can be correlated with the hydrophobic character of this ligand point [30].

Crippen adjusted his receptor mapping procedures in a similar way [31–32]. Whereas Bultsma and co-workers used physicochemical *group* parameters [1, 30], Crippen developed a system containing *atomic* contributions to physicochemical parameters using the fragment values of Hansch [1] and the data set of Martin [33].

In a recent paper of Sheridan et al. [34], an extension of conventional distance geometry techniques was presented: Two or more molecules were treated as an 'ensemble' in order to find a common pharmacophore from a small set of biologically active molecules. The approach can generate, in one step, coordinates for the set of molecules in their 'active' (low-energy) conformations such that their essential groups are superimposed. The energy embedded distance geometry approach [20] is followed to produce docked conformations of the molecules. With the additional information obtained from a superposition of volumes of several agonists the 'active' conformations of the compounds were deduced.

B. Prediction of new leads

When the geometry and nature of the macromolecular binding site are deduced, it is possible to propose new, stronger binding ligands. One can both calculate the binding energy of ligands outside the original data set and predict their mode of binding. The predicted binding mode can be used to suggest conformationally restricted analogues having an improved binding energy and specificity of binding. Crippen [35, 36] describes a computer algorithm which searches for novel chemical classes of compounds with a selected activity. The algorithm consults a library of ring systems and other molecular fragments of known stucture and fits the most likely of these into the site, making chemical modifications where appropriate. The calculations start with a given site, a matrix of energy interaction parameters, and a library of molecular fragments of interest for the particular problem. Rigid groups may be joined to a trial molecule by computer simulating 'substitution reactions'.

C. Accomplishments of 3-D structure-directed QSAR

The work of Crippen is focussed on elucidation of the structural and energetic characteristics of the active site of dihydrofolate reductase isolated from various sources (see, for example, Refs. 10, 36). A study on 68 quinazoline inhibitors of *S.faecium* DHFR led, for example, to a 9-point site model with 6 different site point types. A total number of 20 energy parameters was determined. The calculated free energies of binding fitted the experimental data with a correlation coefficient of 0.955 and a standard deviation of 0.67 kcal/mol. The systematic search method as described in Section B [35] located diaminopteridines with calculated activity comparable to that of the best quinazolines for which the site was constructed. Although in most studies carried out by Crippen and co-workers the predictive power of the proposed site models appeared to be reasonably good [31], and suggestions were put forward for better inhibitors, as far as we know these were never synthesized and tested.

At our laboratory (the group of the late T. Bultsma and co-workers), we investigated some aspects of ligand binding to the β_1/β_2-adrenoceptor from several sources [13–15]. The receptor mapping procedures gave insight in the characteristic differences between the high- and low-affinity

states of the β_2-adrenoceptor. The conformational change which is induced by agonist binding could be modeled, and a comparison of the two states revealed that both steric and energetic interactions with the receptor determine whether a compound acts as an agonist or antagonist. Using a data set of 38 compounds, a 12-point geometric site model was derived with 7 different site point types for the low-affinity state, and an 8-point model for the high-affinity state.

Studying enzyme inhibitors has the clear advantage that in many cases crystallographic information is available either for the enzyme-inhibitor complex under study or for a closely related one. This enables both a check on and refinement of the proposed binding site model. For receptors, crystallographic information is generally not available.

D. Statistics/limitations of the method

A distance geometry study on a large set of molecules consists of guessing the number of site points, guessing a set of binding modes, calculating coordinates for the site points, and then adjusting the energy parameters for a least-squares fit to the binding data. Although the number of adjustable energy parameters can be reduced substantially by correlating interaction energies with physicochemical properties of the ligand points, the Hansch approach is superior in terms of number of variable parameters. For example, in an application of the distance geometry approach to dihydrofolate reductase inhibitors [10], a total of 36 geometric and energetic parameters was used to fit the experimental data of 22 compounds, whereas in a Hansch analysis of 68 dihydrofolate reductase inhibitors, only 6 parameters were used [1].

The differences between the distance geometry approach and a simple Hansch analysis lie in their predictions and general predictive power. The Hansch method fails to give much detail on size and shape of the binding site, as the method is based on a fixed position, orientation and conformation of the ligand in the active site. The advantage of a 3-D model is that a molecule can be designed of one class that can possibly interact with a region ordinarily used by members of another class.

Whereas the Hansch approach could just as well correlate structure to very complicated experimental observations, the distance geometry approach is restricted to the much simpler problem of accounting for observed free energies of binding to a single site on a single receptor. For that purpose, reliable affinity constants of the ligand-receptor complex must be available. Furthermore, if the structural diversity of the compounds in the data set is small, the predictive power of the proposed site will be limited. In order to obtain statistically acceptable values for the energy parameters, the number of investigated compounds has to be large (at least 3–5 drugs per parameter estimated).

E. Other 3-D stucture-directed QSAR methods

Besides the distance geometry procedures which were developed by the research groups of Crippen and Bultsma, other 3-D structure-directed QSAR techniques were developed by the research groups of Simon and Hopfinger. The minimal steric (topological) difference (MTD) analysis of Simon et al. [37–38] derives a picture of the site containing only information on steric accessibility but nothing about the chemical nature of parts of the site, such as hydrogen bonding or hydrophobic pockets. The calculated biological activity is assumed to decrease linearly with the steric

misfit parameter, MTD, which is obtained by superimposing each ligand over a hypermolecule. The latter is constructed by superposition of all molecules of the data set aiming at a maximal overlap.

The molecular shape analysis (MSA) developed by Hopfinger et al. [16–18] quantifies the degree of volume overlap between superimposed molecules and uses this as another term in a Hansch analysis. The common overlap steric volume with an idealized structure, the shape reference, is assumed to have a direct relationship to the biological activity of the molecule. In more recent work [19], Hopfinger redefined the shape descriptor with use of a molecular mechanics potential field and results were found to be improved. Both the MTD and MSA analysis are compatible with the lock-and-key interpretation of the active receptor site and are in that way less realistic.

F. Future trends in receptor mapping

Visualization of the proposed receptor site by means of computer graphics techniques and subsequently docking of flexible molecules to the macromolecular receptor is a subject which deserves much attention. Docking means: positioning the probe in the vicinity of a macromolecular region termed active center and observing the conformational and energetic changes in the macromolecule/probe system. The method as developed by Crippen [35] to predict new leads, would be more powerful if parts of the procedures were visualized by applying computer graphics techniques.

Many algorithms [39, 40] have already been developed for interactive docking of molecules to a macromolecule of a well-known 3-D structure followed by molecular mechanics energy minimization procedures for the estimation of binding enthalpy. This process is called receptor fitting and major limitations to these studies are related to possible conformational changes to the crystal structure upon binding of the drug and the nature of the algorithms and parameters used to minimize the receptor alone and the receptor-ligand complex.

Karfunkel [39] developed an algorithm for interactive docking in which the flexibility of both substrate and macromolecule was considered. The central idea is the concept of relevant docking coordinates which reflect the essential features of the macromolecular deformations during the docking maneuver and which substantially diminish the number of considered degrees of freedom. Desjarlais et al. [40] developed an algorithm to explore the interaction of flexible ligands with receptors of known geometry on the basis of molecular shape. The shape of a binding site on a macromolecular receptor is presented as a set of overlapping spheres. The shape characterization step begins with a calculation of the solvent accessible surface. This surface consists of points from which a set of spheres is generated representing the negative image of the receptor volume. Each ligand is divided into a small set of large rigid fragments that are docked separately into the binding site and then rejoined later in the calculation. The division of ligands into separate fragments allows a degree of flexibility at the position that joins them. The rejoined fragments are then energy minimized in the receptor site using AMBER [41].

However, the number of receptors with known crystallographic structures remains extremely small compared to the thousands of biologically important supramolecules. According to the above-mentioned authors [39, 40], a potential use of their method is docking of flexible ligands to a receptor cavity whose shape, in some cases, can be approximated by comparing the shape of several active ligands.

A method [42] which has proven to be useful for the determination of energetically favorable

binding sites on biologically important macromolecules (phospholipase A_2, DHFR reductase) might, in the future, be used for the determination of the nature of the amino acids involved in specific drug-receptor complexes and their relative positions and orientations at the drug surface. The computational procedure was developed by Goodford in 1985 and the basic concept is calculation of the interaction between a molecule and a probe at sample positions throughout and around the molecule. Energy calculations are performed with the program CHARMM [42] which considers a distance-dependent dielectric constant in the electrostatic function: the dielectric constant diminishes upon decreasing the relative distance between probe and (macro)molecule. The probes include water, the methyl group, amine nitrogen, carboxy oxygen, and hydroxyl. Contours at negative energy values delineate regions of attraction between probe and molecule.

Without knowing the active (receptor-bound) conformation of a drug, the above approach [42] cannot be applied in a useful way. The distance geometry approach as described in this report can be very helpful in determining this active conformation. Also the 'active analogue approach' can be used. In contrast to the 'distance geometry analysis', this method asks for chosen reference pharmacophoric features. Each compound is allowed to explore its set of conformations while recording the relative positions of the set of pharmacophoric groups. The logical intersection of the sets of patterns thus defined for each ligand must lead to those shared by all of them, and containing all possible candidate pharmacophoric patterns. A single set defines unambiguously the pharmacophore. A null set indicates inconsistency. This method can be used to define the topography of the primary anchoring molecular features and therefore the receptor-bound conformation of each active ligand.

In the near future, computer graphics can play a main role in elucidating the features of receptors in cases where a sound theoretical basis for the algorithms can be guaranteed.

REFERENCES

1 a. Hansch, C. and Fujita, T.J., J.Am.Chem.Soc., 86 (1964) 1616–1626.
b. Hansch, C., Unger, S.H. and Forsythe. A.B., J.Med.Chem., 16 (1973) 1217–1222.
c. Hansch, C. and Leo, A., Substituent Constants for Correlation Analysis in Chemistry and Biology, Wiley, New York, 1979.
d. Hansch, C., J.Med.Chem., 19 (1976) 1–6.
e. Hansch, C., Fukunaga, J.Y., Jow, P.Y.C. and Hynes, J.B., J.Med.Chem., 20 (1977) 96–102.
2 a. Dunn, W.J., III and Wold, S.J., J.Med.Chem., 21 (1978) 1001–1007.
b. Dunn, W.J., III, Wold, S.J. and Stalling, D.L., ACS Symposium Series 292, American Chemical Society, 1985, 2435–2439.
3 Franke, R., Dove, S. and Kuhne, R., Eur.J.Med.Chem., 14 (1979) 363–374.
4 Cammarate, A. and Menon, G.K., J.Med.Chem., 19 (1976) 739–748.
5 Hyde, R. and Lord, E., Eur.J.Med.Chem., 14 (1979) 199–206.
6 Kirschner, G. and Kowalski, B.R., Drug Design, 8 (1979) 73–79.
7 Jurs, P.C., Stouch, T.R., Czerwinski, M. and Narvaez, J.N., J.Chem.Inf.Comput.Sci., 25 (1985) 296–302.
8 Moriguchi, L., Komatsu, K. and Matsushi, Y., J.Med.Chem., 23 (1980) 20–26.
9 Crippen, G.M. and Havel, T.F., Acta.Crystallogr., Sect. A 34 (1978) 282–284.
10 Crippen, G.M., J.Med.Chem., 22 (1979) 988–997.
11 Crippen, G.M., J.Med.Chem., 23 (1980) 599–606.
12 Crippen, G.M., J.Med.Chem., 24 (1981) 198–203.
13 Linschoten, M.R., Bultsma, T., IJzerman, A.P. and Timmerman, H., J.Med.Chem., 29 (1986) 278–286.
14 Donné-Op den Kelder, G.M., Bijloo, G.J. and Bultsma, T., Eur.J.Med.Chem., 21 (1986) 475–485.

15 Donné-Op den Kelder, G.M., Bultsma, T., Timmerman, H. and Rademaker, B., submitted.
16 Hopfinger, A.J., J.Am.Chem.Soc., 102 (1980) 7196–7206.
17 Battershell, C., Malhotra, D. and Hopfinger, A.J., J.Med.Chem., 24 (1981) 812–818.
18 Hopfinger, A.J., J.Med.Chem., 26 (1983) 990–996.
19 Mabilia, M., Pearlstein, R.A. and Hopfinger, A.J., Eur.J.Med.Chem., 20 (1985) 163–174.
20 Crippen, G.M., J.Comp.Chem., 3 (1982) 471–476.
21 Crippen, G.M., J.Comp.Chem., 5 (1984) 548–554.
22 a. Crippen, G.M., Int.J.Pept.Protein.Res., 13 (1979) 320–326.
b. Havel, T.F., Kuntz, I.D. and Crippen, G.M., Biopolymers, 18 (1979) 73–81.
23 Kuntz, I.D., Crippen, G.M. and Kollman, P.A., Biopolymers, 18 (1979) 939–946.
24 a. Kuntz, I.D. and Crippen, G.M., Biophys.J., 32 (1980) 677–702.
b. Crippen, G.M., Oppenheimer, N.J. and Connolly, M.L., Int.J.Pept.Protein Res., 17 (1981) 156–161.
25 Kuester, J.L. and Mize, J.H., Optimization Techniques with Fortran, McGraw-Hill, New York, 1973, p. 106.
26 Ghose, A.K. and Crippen, G.M., J.Med.Chem., 27 (1984) 901–914.
27 Burkert, U. and Allinger, N.L., Molecular Mechanics, ACS Monograph 177, American Chemical Society, Washington, D.C., 1982, p. 144.
28 Altona, C. and Faber, D.H., Top.Curr.Chem., 45 (1974) 1–38.
29 Ghose, A.K. and Crippen, G.M., J.Comp.Chem., 6 (1985) 350–359.
30 Rekker, R.F. and de Kort, H.M., Eur.J.Med.Chem., 14 (1979) 479–488.
31 Ghose, A.K. and Crippen, G.M., J.Med.Chem., 28 (1985) 333–346.
32 Ghose, A.K. and Crippen, G.M., J.Comp.Chem., 7 (1986) 565–577.
33 Martin, Y.C., Quantitative Drug Design, Marcel Dekker, New York, 1978.
34 Sheridan, R.P., Nilakantan, R., Dixon, J.S. and Venkataraghavan, R., J.Med.Chem., 29 (1986) 899–906.
35 a. Crippen, G.M., Quant.Struct.-Act.Relat., 2 (1983) 95–100.
b. Kuhl, F.S., Crippen, G.M. and Friesen, D.K., J.Comp.Chem., 5 (1984) 24–34.
36 a. Ghose, A.K. and Crippen, G.M., J.Med.Chem., 25 (1982) 892–899.
b. Ghose, A.K. and Crippen, G.M., J.Med.Chem., 26 (1983) 996–1010.
37 Simon, Z., Badilescu, I. and Racovital, T., J.Theor.Biol., 66 (1977) 485–491.
38 Simon, Z., Dragomir, N., Plauchithiu, M.G., Holban, S., Glatt, H. and Kerek, F., Eur.J.Med.Chem., 15 (1980) 521–527.
39 Karfunkel, H.R., J.Comp.Chem., 7 (1986) 113–119.
40 DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., J.Med.Chem., 29 (1986) 2149–2153.
41 Weiner, P.K. and Kollman, P.A., J.Comp.Chem., 2 (1981) 287–303.
42 Goodford, P.J., J.Med.Chem., 28 (1985) 849–857.