(American Type Culture Collection, Rockville, MD) maintained in RPMI-1640 medium supplemented with 10% fetal bovine serum (Grand Island Biological Co., Grand Island, NY) and 20 mM Hepes buffer. For growth experiments, cells were adjusted to $1 \times 10^5$ cells/mL and distributed to 24-well tissue culture plates (0.5 mL/well). Test compounds were dissolved in growth medium, sterilized by passage through an 0.22-$\mu$m membrane filter, serially diluted, and added to wells (0.5 mL/well). Compounds were tested in duplicate at log concentrations ranging from $1 \times 10^{-10}$ to $1 \times 10^{-4}$ M. Following 48-h incubation at 37 °C, cell counts were determined with a Coulter Model ZF cell counter. Cell growth in the presence of test compounds was expressed as a percentage of growth in untreated control wells and the concentration of compound producing 50% inhibition of cell growth was determined ($ID_{50}$).

**B. B16 Melanoma and Lewis Lung Carcinoma.** Lewis Lung carcinoma was maintained as a monolayer in RPMI-1640 medium supplemented with 5% heat-inactivated fetal calf serum. B16 melanoma was maintained as a monolayer in Eagle's minimum essential medium (MEM) containing 10% heat-inactivated fetal calf serum, 0.1 mM MEM nonessential amino acids, 1 mM sodium pyruvate, and MEM vitamin solution. For determination of cell growth inhibition using either cell line, cells were seeded at $2.5 \times 10^4$ cells per well in 24 tissue culture plates. Cells were grown 24 h at 37 °C in 5% $CO_2$, and then growth medium was replaced with medium containing the compound of interest at log concentrations ranging from $1 \times 10^{-10}$ to $1 \times 10^{-4}$ M. After an additional 72-h incubation, cells were washed twice with phosphate buffered saline, trypsized to single cell suspensions, and counted with a Coulter Model ZF cell counter. Cell growth at each dose level was expressed as a percentage of growth in control wells and the dose resulting 50% inhibition of growth was determined.

**Acknowledgment.** This work was supported in part by Contract DAMD-17-79-C-9046 with the U.S. Army Medical Research and Development Command, Washington, DC. This is Contribution No. 5003 to the USAMRIID Drug Development Program. We thank Ratan B. Bhappu for large-scale syntheses of certain chemical intermediates.

**Registry No.** 1, 30868-30-5; 4, 99298-13-2; 4 (tetraacetyl derivative), 99298-21-2; 4 (5-carbonitrile), 99298-22-3; 5, 99298-15-4; 6, 99298-16-5; 7, 41329-11-7; 8, 99298-14-3; 9, 59252-13-0; 9 (5-carbonitrile), 99298-17-6; 10, 99298-18-7; 11, 99298-19-8; 12, 99298-20-1; 13, 99298-23-4; 14, 99298-24-5; 15, 99298-25-6; 16, 99298-22-3; 17, 99298-26-7; 18, 65446-02-8; 19a, 99298-27-8; 19b, 99298-28-9; 20a, 99298-29-0; 20b, 99298-30-3; 21a, 99298-31-4; 21b, 99298-32-5; 22, 57816-25-8; 23a, 99298-33-6; 23b, 99298-34-7.

**Supplementary Material Available:** Tables of positional parameters (Tables V-VIII), anisotropic thermal parameters of the non-hydrogen atoms (Tables IX-XII), and positional and isotropic thermal parameters of the hydrogen atoms (Tables XIII-XVI) for compounds 1, 8, 7, and 20a, respectively (12 pages). Ordering information is given on any current masthead page.

# Mapping the Turkey Erythrocyte $\beta$ Receptor: A Distance Geometry Approach

Marcel R. Linschoten, Teake Bultsma,* Ad P. IJzerman, and Hendrik Timmerman

*Department of Pharmacochemistry, Free University, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands.*
*Received September 18, 1984*

Extensions and refinements of the receptor mapping method as originally developed by Crippen are presented. In a set of newly developed algorithms measures are taken to reduce the number of required energy parameters to a statistically acceptable degree. The most important measure is the incorporation of lipophilicity as a hydrophobic bonding parameter to describe the binding of parts of the ligands to lipophilic areas on the receptor. In order to test the applicability of our set of programs, we mapped the turkey erythrocyte $\beta$ receptor using a data set of Bilezikian. It was found that the experimentally determined free energies of binding can be reasonably described using a nine-point geometrical representation of the receptor site and only six energy parameters. The deduced model predicts that the phenyl rings of phenylethanolamines and phenoxypropanolamines occupy different parts of the receptor site.

Conformational flexibility of ligands poses a major problem in pharmacophoric pattern search and receptor mapping techniques. It has become clear that it is unrealistic to assume that ligands always bind to the receptor in their conformations of minimal energy as the free energy of association generally outweighs the energy of a conformational change of the ligand. It is therefore necessary to take into account all conformations having energies up to a few kilocalories/mole above the global minimum.

Conformational flexibility can be conveniently condensed in a distance matrix. Considering the total set of possible conformations, the entries in the upper triangle of such a matrix usually are the maximum distances between any pair of preselected atoms or dummy points in the molecule, whereas the entries in the lower triangle are the corresponding minimum distances.

The methodology of distance geometry is a powerful tool in handling the information contained in such matrices. During the last few years especially Crippen[1] has applied the distance geometry approach as a receptor mapping technique.

By comparing the distance matrices of the ligands in the data set, Crippen was able to deduce the common structural features of a set of ligands. Substituent points were subsequently positioned relative to this common base group. Complementary receptor "site points" were then proposed to account for the binding of the structural features of the ligand ("ligand points"). Furthermore, Crippen supposed that the total binding energy of a ligand to its receptor is equal to the sum of the individual interactions between ligand points and site points. Given the experimentally determined free energy of binding, the method enables one to calculate the individual binding energy contributions of any ligand point–site point interaction.

Thus, next to the incorporation of conformational flexibility in the calculations, the strong point of the method is that it enables the researcher to propose new, stronger binding ligands on basis of the geometry and energy parameters of the deduced receptor site model.

A major problem with respect to the general applica-

(1) Crippen, G. *J. Med. Chem.* 1980, *23*, 599.

0022-2623/86/1829-0278$01.50/0 

bility of the method is set by the type of data needed. The data set has to satisfy the following criteria. (1) The geometry of the site can only be deduced if the data set contains various classes of structurally diverse ligands. Only in this way will the deduction of the common features of the ligands lead to meaningful results. (2) The data set must be large enough to allow a statistically sound derivation of the values of the energy parameters which account for the binding strength of the ligand to the receptor. (3) Reliable dissociation constants of the ligand–receptor complex must be available, as only the dissociation constant is directly related to the free energy of binding.

We focused our attention on the second criterion.

It should be clear that a reduction of the number of energy parameters allows the use of smaller data sets.

In this paper we present the receptor mapping algorithms developed in our laboratory. Although the method of Crippen is used in essentials, measures are taken to reduce the number of energy parameters to a degree which is acceptable from a statistical point of view. In evaluating the value of the method, we have mapped the turkey erythrocyte β receptor using the data set published by Bilezikian et al.[2,3] This data set was chosen for two reasons: firstly, our laboratory has been engaged in a program on the $\beta_1/\beta_2$ selectivity of adrenoceptor ligands. Secondly, this data set satisfies the criteria discussed above.

## Methods

As our method deviates from Crippen's method in a few major aspects, it will be outlined to some extent.

With respect to the drug–receptor interactions we use the same assumptions as Crippen. The most important ones are as follows. (1) The experimentally found free energy of binding is approximately equal to the sum of the interaction energies for all contacts between parts of the ligand molecule and parts of the receptor site. (2) The conformation of the ligand may change as the ligand interacts with the receptor site whenever the free energy of such a conformational change is small compared to the free energy of binding.

The conformational energy is not considered in the calculation of the total binding energy nor is entropy loss.[4] We consider these topics as refinements of the procedure. Calculations were performed on a CDC Cyber 170-750 computer. The programs used have all been written in Fortran IV.

The receptor mapping procedure consists of the following steps.

**A. Calculation of Allowed Conformations.** The structures of the ligands are read in using internal coordinates. Interatomic distances and valence angles were taken from tables which were also the basis of CPK models.[5] The bonds which have to be rotated are defined as is the step size over which the torsional angles are to be rotated during the analysis. The energetically allowed conformations are subsequently calculated by program "ENCOR" (ENergies of COnfoRmations) on basis of the energies of the van der Waals interactions and torsional interactions present in the molecule.[6]

All conformations having energies lower than a preselected limiting energy level are allowed. The energy contribution of the van der Waals interactions is calculated by using the potential functions given by Giglio,[7] whereas the contribution of the torsional interactions was calculated on the basis of the usual sinusoidal relationships. Other energy terms were not taken into account. Obviously, this procedure does not lead to exact conformational energies, as strain energies, dipole–dipole interactions, ion–dipole interactions, and, even more important, the effects of the surrounding medium are not taken into account. It should be noted that we do not consider the exact conformational energy of any individual conformer but a whole range of conformers with a sufficiently low energy. The ultimate goal of the conformational analysis is the generation of the distance matrix, which on itself is a rough representation of the energetically allowed conformations. However, at this moment it is the only way to deal with thousands of conformations in a convenient way.

The distance matrix gives both the upper and lower bounds on the distances between preselected atoms or dummy atoms in the molecule (such a dummy could, for example, be the center of a phenyl ring). In general, it is advisable to calculate the maximum and minimum distances between all the atoms in the molecular skeleton. The resulting very large distance matrix can easily be reduced to a matrix only containing information concerning the atoms which are believed to be a good representation of the binding parts of the ligand. This procedure eliminates the need to calculate a new set of matrices using other atoms if it turns out to be necessary to use another representation of the ligands in a later stage of the receptor mapping process.

The conformational analysis is the most time-consuming part of the whole procedure as computer time increases exponentially with the number of rotated torsional angles.

**B. Selection of Ligand Points.** After a set of distance matrices has been obtained for the ligands in the data set, we select the atoms and dummy points which are likely to describe properly the interactions of the drug molecule with the receptor macromolecule. Frequently, other SAR studies indicate the relative importance of the various parts of the ligands. This helps in a proper selection of the ligand points. Following the nomenclature proposed by Crippen, these atoms and dummy points are called *ligand points*. The receptor pockets in which these ligand points are situated when interacting with parts of the receptor are correspondingly called *site points*. As will be shown, the number and positions of the selected ligand points are crucial for the outcome of the procedure.

**C. Decomposition Algorithm.** In order to bind to the site, a ligand has to be "recognized" by the receptor. Those features of the ligands which are indispensable for them to be recognized make up their pharmacophore. In this study it is assumed that the pharmacophore always binds to the receptor in the same orientation. The ligand points not belonging to the common pharmacophoric group of ligand points are referred to as *substituent points*. Once it has been accepted that the pharmacophore which the ligands have in common always binds in the same way, the complementary features of the receptor protein can be thought of as the "backbone" of the receptor site. Correspondingly, site points are defined covering these complementary features. These site points more or less define the coordinate space in which the site points for the substituent points have to be positioned. It is crucial for the success of the method that the pharmacophore is well

(2) Bilezikian, J. P.; Dornfeld, A. M.; Gammon, D. E.; *Biochem. Pharmacol.* 1978, *27*, 1445.
(3) Bilezikian, J. P.; Dornfeld, A. M.; Gammon, D. E. *Biochem. Pharmacol.* 1978, *27*, 1455.
(4) Andrews, P. R.; Craik, D. J.; Martin, J. L. *J. Med. Chem.* 1984, *27*, 1648.
(5) Harte, R. A. "Molecules in Three Dimensions"; American Society for Biological Chemists, Inc.: Bethseda, MD, 1969.
(6) Winter, R. L. de; Bultsma, T.; Nauta, W. Th. *Eur. J. Med. Chem.* 1977, *12*, 137.
(7) Giglio, E. *Nature (London)* 1969, *222*, 339.

defined. To check this the following criteria may be used. (a) There must be enough structural variation in the ligands; otherwise the pharmacophore cannot even be deduced. (b) The mutual spatial orientation of the pharmacophoric ligand points must be well defined. The geometry of the pharmacophoric pattern can only be deduced if strongly binding analogues are known in which the pharmacophore is fairly rigid, e.g. as a consequence of steric requirements or ring closure. The substituent points of the ligands frequently possess considerable conformational freedom, especially in case the number of subsequent rotatable torsional angles is large. However, constraints on this flexibility can often be made: superposition of the ligands from the data set can yield a relatively "sharp" picture of the receptor site.

Following again the nomenclature proposed by Crippen, the algorithm devised to perform the calculation on basis of this rationale is called a *decomposition algorithm*. After the distance matrices have been read in, program "DECOM" performs the following calculations. The distances between all possible pairs of atoms in the distance matrices of the first two molecules are compared. As there are minimum and maximum distances, a certain overlap can be found. This overlap region defines the new maximum and minimum distances for this combination. The final intersection of the two molecules is found when a maximum overlap of ligand points from the first and second ligand has been found. If there are more possibilities, the *types* of the ligand points are considered. These types are defined on basis of the physical chemical characteristics of the atoms or groups of atoms which the ligand points represent. Lipophilic points such as the center of a phenyl group are given a different type than hydrophilic points such as hydroxyl or amino moieties. Of course, the types of amino and hydroxyl groups may also be defined separately. In case more equivalent intersections of the first two ligands are possible (i.e. all intersections have the same number of overlapping ligand points), the intersection giving a maximum type overlap is selected by the program. In other words: the geometric overlap predominates the type overlap. The obtained intersection is then intersected with the following distance matrix and so on. At the end of the first run of the program the ligand points having a common geometric arrangement have been determined. In terms of the foregoing discussion this group may be called the pharmacophoric pattern, necessary for a ligand to be recognized and bind to the receptor site. In the second run of the program the geometrically distinct substituent ligand points have to be deduced and subsequently positioned with respect to the pharmacophoric ligand points. All ligand points of the first ligand which do not belong to the pharmacophoric group obviously are geometrically distinct substituent ligand points. The maximum and minimum distances of these substituent points to the points in the pharmacophoric group are determined from the distance matrix. All substituent points of the second distance matrix and their distances to the pharmacophore are then determined, and it is checked whether these points are geometrically distinct from the already deduced substituent points from the first distance matrix. Whenever two substituent points from the first two ligands overlap, it is postulated that these substituent points occupy the same region of the receptor site. They are consequently merged to only one substituent point. To achieve this, the bounds on the minimum and maximum distances to the pharmacophoric points are constrained such that the largest minimum distance to the pharmacophoric points is taken to be the new minimum distance whereas the smallest maximum distance is taken to be the new maximum distance. When the geometrically distinct substituent points of the first two ligands have been determined, the substituent points of the other ligands from the input file are determined. The ligands have now all been decomposed into a common pharmacophoric group and a set of substituent points. The results are expressed in a distance matrix, which however is incomplete: only the maximum and minimum distances of the substituent points to the pharmacophoric ligand points are known and not the distances between the substituent points themselves. This is no problem if one wants to convert the mutual distances into a set of Cartesian coordinates of all ligand points: the distance geometry algorithm described by Crippen and Havel[8] can be applied to the pharmacophoric group and each of the substituent points separately. However, the flexibility of the ligand points—expressed in different values for the minimum and maximum distances—is a problem in calculating a set of discrete coordinates. The problem might be partly solved using a principal component analysis: the amount of information in the distance matrix which can be explained geometrically using only three coordinate axes can be calculated from the first three eigenvalues of the diagonalized metric matrix. It follows that a distance matrix with different minimum and maximum distances can be geometrically optimized by varying the distances between all sets of two points within the limits indicated by the minimum and maximum distances and subsequently determining which of the possible symmetrical distance matrices has the largest sum of the first three eigenvalues. The corresponding coordinates can then easily be determined.

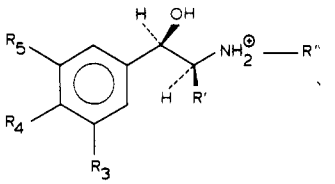
Having obtained a set of coordinates for all the ligand points, site points are proposed to accommodate any occurrence of ligand points. Initially, the site points are given the same type as the corresponding ligand points. In practice there are still cases in which the geometrical freedom of some ligand points is so large that it is impossible to deduce their coordinates. Although no geometrical representation can be given, the fractional binding energies of these parts of the ligands to the receptor are considered in the energy calculations. This, of course, is equivalent to a Hansch approach. We have now determined the geometry of the postulated site. However, no information has been obtained concerning the possible binding modes of each of the ligands on the site. This is the objective of the following algorithm.

**D. Deduction of Feasible Matches.** Given the representation of the site and the distance matrices of the ligands, we want to calculate all binding modes of the ligands to the receptor having a maximum overlap of the site points and ligand points. To achieve this our computer program "FEASM" (FEASable Matches) compares the distance matrix of the site to the matrix of any of the ligands. The fits having a maximum geometric overlap are selected. In case more possibilities are found, the types of the ligand points and site points are compared. Then the binding mode with maximum type overlap is selected. The deduction of the actual binding mode is the task of the energy minimization algorithm described hereafter. A certain tolerance in the site is accepted. This tolerance is expressed in a radius for all the site points. As will be shown in the next section of this paper, it is good practice to take rather large radii initially.

**E. Calculation of Compatible Conformations.** It is not certain whether a conformation of the ligand can be found which has the same distance matrix as the site points

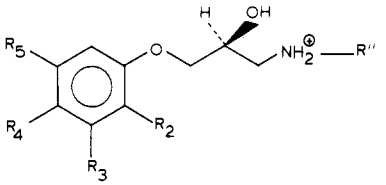
(8) Crippen, G. M.; Havel, T. F. *Acta Crystallogr., Sect. A* 1978, *34*, 282.

**Table I**



Class I: Phenylethanolamines

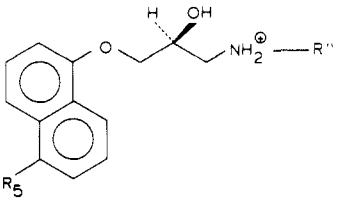
| compd | $R_3$ | $R_4$ | $R_5$ | R′ | R″ | $\Delta G_{obsd}$ | $\Delta G_{calcd}$ | $\Delta G_{obsd} - \Delta G_{calcd}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | OH | OH | H | H | H | −7.26 (±0.02) | −7.50 | 0.24 |
| 2 | OH | OH | H | $CH_3$ | H | −9.64 (±0.04) | −7.79[a] | −1.85 |
| 3 | OH | OH | H | H | $CH_3$ | −7.86 (±0.01) | −7.79 | −0.07 |
| 4 | OH | OH | H | H | $CH(CH_3)_2$ | −9.18 (±0.11) | −8.23 | −0.95 |
| 5 | OH | OH | H | $CH_2CH_3$ | $CH(CH_3)_2$ | −8.02 (±0.03) | −8.74 | 0.72 |
| 6 | OH | OH | H | H | $CH_2CH_3$ | −7.59 (±0.01) | −8.01 | 0.42 |
| 7 | OH | OH | H | H | $CH(CH_3)CH_2C_6H_3$-3,4-$OCH_2O$ | −9.88 (±0.03) | −8.99 | −0.89 |
| 8 | OH | OH | H | H | $CH(CH_3)CH_2C_6H_4$-4-OH | −10.43 (±0.07) | −9.74 | −0.69 |
| 9 | OH | OH | H | H | $C(CH_3)_2CH_2C_6H_4$-4-OH | −10.00 (±0.03) | −9.96 | −0.04 |
| 10 | H | OH | H | H | H | −5.94 (±0.11) | −6.69 | 0.75 |
| 11 | H | OH | H | $CH_3$ | H | −6.27 (±0.09) | −6.98 | 0.71 |
| 12 | H | OH | H | H | $CH_3$ | −6.01 (±0.20) | −6.98 | 0.97 |
| 13 | H | OH | H | $CH_3$ | $CH_3$ | −6.47 (±0.09) | −7.28 | 0.81 |
| 14 | H | OH | H | H | $CH(CH_3)_2$ | −7.54 (±0.13) | −7.42 | −0.12 |
| 15 | H | OH | H | H | $CH(CH_3)CH_2OC_6H_5$ | −7.95 (±0.11) | −7.93 | −0.02 |
| 16 | H | OH | H | $CH_3$ | $CH(CH_3)CH_2CH_2C_6H_5$ | −9.55 (±0.08) | −8.63 | −0.92 |
| 17 | H | OH | H | $CH_3$ | $CH_2CH_2C_6H_5$ | −9.01 (±0.10) | −8.19 | −0.82 |
| 18 | $NHSO_2CH_3$ | OH | H | H | $CH(CH_3)_2$ | −7.36 (±0.01) | −8.55 | 1.19 |
| 19 | $CH_2SO_2CH_3$ | OH | H | H | $CH(CH_3)_2$ | −7.85 (±0.21) | −8.45 | 0.60 |
| 20 | $CH_2OH$ | OH | H | H | $CH(CH_3)CH_2C_6H_4$-4-$OCH_3$ | −7.65 (±0.15) | −8.96[a] | 1.31 |
| 21 | $CH_2OH$ | OH | H | H | $C(CH_3)_3$ | −8.23 (±0.08) | −8.45 | 0.22 |
| 22 | $NHSO_2CH_3$ | OH | H | H | cyclopropyl | −8.19 (±0.09) | −8.39 | 0.20 |
| 23 | $NHSO_2CH_3$ | OH | H | H | $C(CH_3)_2CH_2C_6H_5$ | −9.64 (±0.10) | −9.46 | −0.18 |
| 24 | OH | $NHSO_2CH_3$ | H | H | $CH(CH_3)_2$ | −6.29 (±0.10) | −7.16 | 0.87 |
| 25 | OH | H | $CH_2OH$ | H | $C(CH_3)_3$ | −6.58 (±0.17) | −7.00 | 0.42 |
| 26 | OH | H | H | H | $CH_2CH_3$ | −6.75 (±0.09) | −7.02 | 0.27 |
| 27 | OH | H | $CH_2OH$ | H | $C(CH_3)_2CH_2C_6H_4$-4-OH | −7.61 (±0.04) | −8.51[a] | 0.90 |
| 28 | OH | H | H | H | H | −7.23 (±0.04) | −6.51 | −0.73 |
| 29 | OH | H | H | H | $CH_3$ | −6.72 (±0.10) | −6.80 | 0.08 |
| 30 | OH | H | OH | H | $CH(CH_3)_2$ | −7.13 (±0.03) | −7.03 | −0.10 |
| 31 | OH | H | H | H | $CH(CH_3)_2$ | −7.99 (±0.04) | −7.24 | −0.75 |
| 32 | OH | H | OH | H | $CH(CH_3)CH_2C_6H_4$-4-OH | −8.62 (±0.04) | −8.54 | −0.08 |
| 33 | H | H | H | H | $CH_3$ | −6.79 (±0.04) | −5.99 | −0.80 |
| 34 | H | H | H | H | $C_6H_5$ | −5.76 (±0.05) | −6.47 | 0.71 |
| 35 | H | $NHSO_2CH_3$ | H | H | $CH(CH_3)_2$ | −8.46 (±0.02) | −8.31 | −0.15 |
| 36 | H | $NHSO_2CH_3$ | H | $CH_2CH_3$ | $CH(CH_3)_2$ | −7.41 (±0.03) | −6.87 | −0.55 |
| 37 | Cl | Cl | H | H | $CH(CH_3)_2$ | −8.83 (±0.31) | −8.44 | −0.39 |



Class 2A: Phenoxypropanolamines

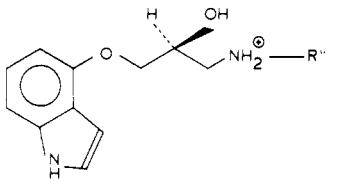
| compd | $R_2$ | $R_3$ | $R_4$ | $R_5$ | R″ | $\Delta G_{obsd}$ | $\Delta G_{calcd}$ | $\Delta G_{obsd} - \Delta G_{calcd}$ |
|---|---|---|---|---|---|---|---|---|
| 38 | I | H | H | H | $CH(CH_3)_2$ | −12.77 (±0.03) | −12.67 | −0.10 |
| 39 | $CH_2CH{=}CH_2$ | H | H | H | $CH(CH_3)_2$ | −12.10 (±0.04) | −12.56 | 0.46 |
| 40 | $OCH_2CH{=}CH_2$ | H | H | H | $CH(CH_3)_2$ | −11.58 (±0.09) | −11.71 | 0.13 |
| 41 | H | OH | H | OH | $C(CH_3)_3$ | −9.86 (±0.03) | −10.53 | 0.67 |
| 42 | H | $CH_2CH{=}CH_2$ | H | H | $CH(CH_3)_2$ | −8.82 (±0.07) | −10.49[a] | 1.67 |
| 43 | H | OH | OH | H | $CH(CH_3)_2$ | −9.13 (±0.03) | −8.23 | −0.90 |
| 44 | H | H | $NH_2$ | H | $CH(CH_3)_2$ | −8.96 (±0.03) | −8.62 | −0.34 |
| 45 | H | H | OH | H | $CH(CH_3)_2$ | −8.91 (±0.43) | −8.31 | −0.60 |
| 46 | H | H | $NHCOCH_3$ | H | $CH(CH_3)_2$ | −7.54 (±0.03) | *b* | |

Table I (Continued)



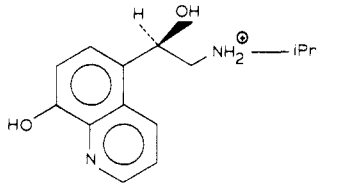
Class 2B: Naphthoxypropanolamines

| compd | $R_5$ | R″ | $\Delta G_{obsd}$ | $\Delta G_{calcd}$ | $\Delta G_{obsd} - \Delta G_{calcd}$ |
|---|---|---|---|---|---|
| 47 | H | H | −10.90 (±0.06) | −9.50 | −1.40 |
| 48 | H | $CH(CH_3)_2$ | −11.70 (±0.11) | −10.24 | −1.47 |
| 49 | H | $CH_2CH_2C_6H_4$-4-OH | −10.93 (±0.02) | −11.53 | 0.60 |
| 50 | H | $C(CH_3)_2CH_2C_6H_4$-4-OH | −11.19 (±0.02) | −11.96 | 0.77 |
| 51 | =O[c] | $CH_2CH_2C_6H_3$-3,4-$(OCH_3)_2$ | −11.35 (±0.03) | −11.95 | 0.60 |
| 52 | OH | $CH(CH_3)_2$ | −11.22 (±0.05) | −11.23 | 0.01 |
| 53 | =O[c] | $C(CH_3)_3$ | −11.91 (±0.03) | −11.27 | −0.64 |



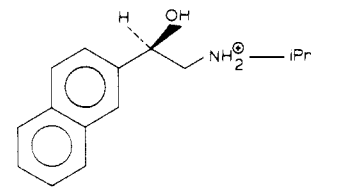
Class 2C: Indoloxypropanolamines

| compd | R″ | $\Delta G_{obsd}$ | $\Delta G_{calcd}$ | $\Delta G_{obsd} - \Delta G_{calcd}$ |
|---|---|---|---|---|
| 54 | $C(CH_3)_3$ | −13.51 (±0.03) | −12.84 | −0.67 |
| 55 | $CH(CH_3)_2$ | −12.99 (±0.04) | −12.62 | −0.37 |
| 56 | $C(CH_3)_2CH_2C_6H_4$-4-OH | −14.72 (±0.15) | −14.35 | −0.38 |



Class 3: Varia

| compd | $\Delta G_{obsd}$ | $\Delta G_{calcd}$ | $\Delta G_{obsd} - \Delta G_{calcd}$ |
|---|---|---|---|
| 57 | −10.36 (±0.12) | −10.91 | 0.55 |



| compd | $\Delta G_{obsd}$ | $\Delta G_{calcd}$ | $\Delta G_{obsd} - \Delta G_{calcd}$ |
|---|---|---|---|
| 58 | −9.06 (±0.29) | −9.17 | 0.11 |

[a] These compounds were omitted from the final energy minimizations for reasons discussed in the text. [b] Compound 46, practolol, was omitted because it is the only compound possessing a $NHCOCH_3$ substituent. [c] Saturated 3,4-benzo ring.

which bind the ligand. This is because the distance matrix of the site may contain high-energy conformations of the ligands. Program "COCON" (COmpatible CONformations) checks whether conformations can be found which are compatible with the distance matrix of the site.

After all the matches indicated by program "FEASM" have been checked for compatible conformations, the calculated binding energies have to be minimized with respect to the observed binding energies over all possible binding modes for the ligands in the data set. This is the objective of the following algorithm.

**F. Energy Minimization Procedure.** As proposed by Crippen, it is postulated that the sum of the energy contributions of the ligand point–site point interactions of a ligand is approximately equal to the total observed binding energy, i.e.

$$\Delta G_{tot} = \sum_{i=1}^{n} \Delta G_i$$

in case of $n$ interactions.

In order to calculate the individual energy contributions, $\Delta G_i$, they have to be varied such that

$$\sum(\Delta G_{obsd} - \sum \Delta G_i)^2$$

reaches a mimimum value.

In many cases purely hydrophobic bonding of parts of ligands is a major factor in the affinity to the receptor. It

is supposed that the affinity contribution of these parts of the ligands can be described by the following Collander equation:

$$\Delta G_i = a\sum f$$

in which $\sum f$ is the calculated lipophilicity contribution of the fragment (according to Rekker's hydrophobic fragmental system[9]). In this way lipophilic–lipophilic and lipophilic–hydrophilic types of interactions can be described by using only one parameter (i.e. the slope a of the Collander equation). Program "RECMA" (RECeptor MApping)—which performs the energy minimization—is based upon a Marquardt algorithm.[10] This procedure combines steepest descent and Gauss–Newton iterative procedures. In preparing the input for this program, an energy interaction table is compiled: the number and types of the energy interactions are proposed as well as the starting values for these parameters. Within a minimization run the binding mode is not changed, as is done in the method of Crippen.[11]

We follow the common practice of Hansch analysis: different sets of parameters are tested until an optimal fit is obtained.

**G. Refinement or Rejection of the Proposed Site.** The described receptor mapping procedure only generates a *model* for the binding site of the receptor protein. It is by no means certain that this model is a good representation of the actual receptor site: a number of factors which cannot be controlled greatly influences the outcome of the procedure. These include the structural variation of the ligands in the data set, the selection of the ligand points, and the order of the ligands in the input file of the decomposition algorithm. As in our procedure only a limited number of energy parameters is used, it is possible to accept or reject a proposed receptor model on basis of the outcome of the energy minimization. Whenever a receptor model is rejected, a new model may be deduced by using another set of ligand points or another order of the ligands in the decomposition algorithm. If, however, the receptor model is accepted on basis of the results of the energy minimization, it is often possible to make further refinements in the representation of the site. A careful study of the structural features of the outliers, for example, may lead to the proposal of additional site points. Moreover, outliers frequently give information about restrictions in the receptor site: the ligands which bind much weaker to the site than expected often possess groups which cannot be accommodated by the receptor site as a result of steric hindrance. It will then be possible to add so called *filled site points* to the site. These filled site points are regions on the receptor which cannot be occupied by ligand points. For a more detailed account on filled site points, the reader is referred to ref 1.

## Results and Discussion

As a test for the applicability of our algorithms, we used the data set of Bilezikian et al.[2,3] This group measured the dissociation constants of a large set of both β-adrenoceptor agonists and antagonists with respect to the turkey erythrocyte β receptor.

As the data set of Bilezikian contains full agonists, partial agonists, and full antagonists; it is important to evaluate in advance whether the affinity state of the receptor is the same for these different classes of compounds. The inclusion of Gpp(NH)p in the assay prevents the formation of the so-called "high-affinity state" of the receptor, as recognized by agonists only.[12] Thus, both for agonists and antagonists, only one conformation, the so-called "low-affinity state" of the receptor, is present. It may therefore safely be concluded that the binding data of both agonists and antagonists can be used in the evaluation of the low-affinity state of the receptor.

The ethanolamine side chain of β-adrenoceptor agonists and antagonists possesses a chiral β-carbon atom. The (−) configuration seems to be prerequisite for proper binding to the β receptor. The absolute configurations of the (−) stereoisomers have been established to be "*R*" in the case of the ethanolamine class of compounds and "*S*" in the case of the phenoxypropanolamines.[13]

In any occurring case, $K_d$ values were corrected for the (−) isomers, assuming that the activities of the (+) isomers are negligible.

The following formula relates the dissociation constant to the free energy of association:

$$\Delta G = -RT \ln K_a = RT \ln K_d$$

in which $R$ is the molar gas constant and, in this case, $T$ = 310 K. It should be noted that not all compounds from the study of Bilezikian were used in our calculations: some compounds were omitted because no exact binding data were given (e.g. tyramine and normetanephrin); others were omitted as their structures could not be reliably deduced without crystallographic data (e.g. tazolol and timolol). Furthermore, it is important to notice that we have only included compounds with an ethanolamine chain in the decomposition algorithm and the energy calculations. This was done for two reasons: firstly, it is generally accepted that an ethanolamine chain is indispensable for a compound to bind specifically and sufficiently strong to the β receptor. Secondly, Jen and Kaiser[14] have drawn attention to a possible intramolecular interaction between the β-hydroxyl group and the positively charged protonated amino moiety, which are both part of the ethanolamine side chain. If their theory holds, it will be very difficult, if not impossible, to decompose the energetic contribution of the binding of the ethanolamine side chain in *separate* terms for the β-hydroxyl and protonated amino moieties.

The formula of the compounds which were eventually selected along with their free energies of association to the receptor site are given in Table I.

**Mapping the Turkey Erythrocyte β Receptor. Conformational Analysis.** The structural formulae (i.e. their proper absolute configurations) were read in using internal coordinates. As a flexible molecule can assume an infinite number of conformations, it was necessary to select the step size of the rotation of the torsional angles in advance. Whenever the ligand had five relevant torsional angles, it was possible to use 12 steps of 30° for each torsional angle. However, when the number of torsional angles exceeded five, the computing time became excessive. In these cases six steps of 60° were used. Nonetheless, computing the allowed conformations of large ligands like **56** costs about 4000 computer seconds. The conformational analysis of all the ligands required about 50 000 computer seconds.

(9) Rekker, R. F.; de Kort, H. M. *Eur. J. Med. Chem.* **1979**, *14*, 479.
(10) Marquardt, D. M. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431.
(11) Ghose, A. K.; Crippen, G. M. "Quantitative Approaches to Drug Design"; Dearden, J. C., Ed.; Elsevier: Amsterdam, 1983; Vol. 99.
(12) Lefkowitz, R. J.; Caron, M. G.; Michel, T.; Stadel, J. M. *Fed. Proc., Fed. Am. Soc. Exp. Biol.* **1982**, *41*, 2664.
(13) Clarkson, R. *ACS Monogr.* **1976**, *27*, 1.
(14) Jen, T.; Kaiser, C. *J. Med. Chem.* **1977**, *20*, 693.

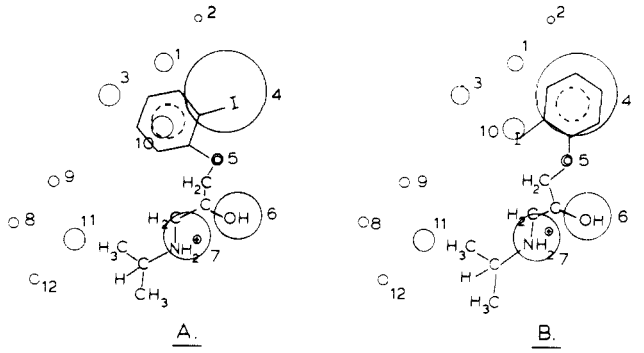


**Figure 1.** Initial receptor map. Program "FEASM" indicates two possible binding modes for the phenoxypropanolamines.

**Selection of Ligand Points.** As the ethanolamine side chain is present in all ligands in our data set, each atom of this side chain was initially taken to be a ligand point. In this way we hoped to obtain a rather large base group. This is important as the geometry of the substituents is determined by their distances to each point of the pharmacophore. Thus, the larger the pharmacophore, the more reliable the geometry of the substituents will be. All rings were represented by only a single dummy atom at the center of the ring. Furthermore, each non-hydrogen substituent of the ring was represented by a ligand point.

After all distance matrices had been reduced to matrices containing only the distance information of the selected ligand points, types were assigned to the ligand points according to the physical chemical nature of the groups or atoms which they represent. Initially, different types were taken for phenyl ring dummy's, hydroxyl groups, amino groups or substituent amino groups, and the hydroxyl group and nitrogen atom of the ethanolamine side chain.

**Geometrical Decomposition and Deduction of Feasible Matches.** In the next step the ligands were decomposed to deduce the common pharmacophoric ligand points and the substituent points. The chosen selection of the ligand points, however, gave rise to an unprobable representation of the receptor site: when each non-hydrogen atom of the ethanolamine chain was taken to be a ligand point, the algorithm always superponed all atoms of the ethanolamine chains of both the phenoxypropanolamines and phenylethanolamines because in this way a maximum geometric overlap was obtained. As a consequence a very large set of substituent points was found. A strong preponderance of ligand points in a small part of the ligand should thus be avoided. It was then decided to represent the ethanolamine chain by only two ligand points, one for the $\beta$-hydroxyl group and one for the positively charged nitrogen atom. As a consequence, however, the pharmacophoric group of the ligands deduced by the decomposition algorithm contained only three ligand points, thus impairing the geometric accuracy of the site representation. This is one of the two reasons that we took rather large radii for the site points initially, the other reason being that, according to our experience, the deduction of feasible matches of the ligands on the site is more flexible and generates more plausible binding modes when the radii of the site points are not taken too small. The deduced initial site model (shown with compound **38** superimposed) is depicted in Figure 1. Whereas only one geometrically optimal binding mode was found for the phenylethanolamines, two plausible binding modes were found in the case of the phenoxypropanolamines: both site points 4 and 10 which both possess a lipophilic character

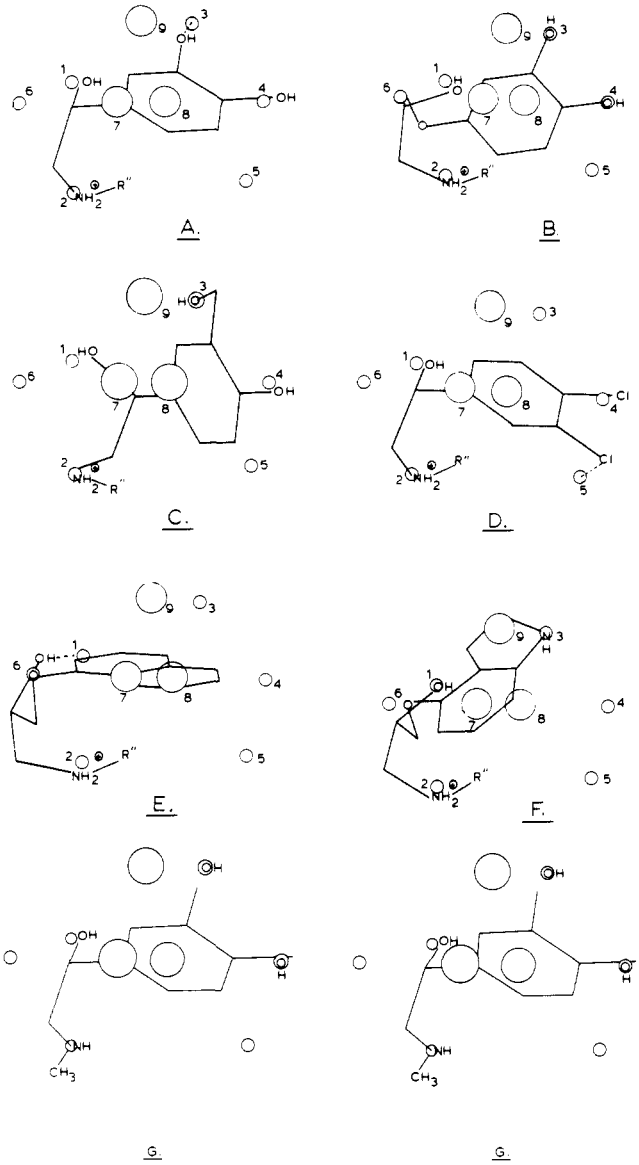


**Figure 2.** Binding modes of various classes of adrenergics and antiadrenergics to the deduced final site: (A) isoprenalin type, (B) RO 363 type, (C) salbutamol type, (D) dichloroisoproterenol type, (E) propranolol type, (F) pindolol type, (G) stereo view of adrenalin on the site. Line of view along the $x$ axis.

can be occupied by the phenyl rings of these compounds (see Figure 1A,B).

**Energy Minimization.** We use the energy minimization algorithm to evaluate which of the two binding modes gives binding energies closest to the observed binding energies. It is possible to exclude binding modes on the basis of the energy calculation because we have taken three measures to reduce the number of required energy parameters.

(1) The inclusion of lipophilicity contributions for those fragments which bind to lipophilic areas on the receptor site allows the incorporation of only one interaction parameter as hydrophobic bonding strength is proportional to the lipophilicity of the fragment (using the hydrophobic fragmental system). Bonding of lipophilic as well as hydrophilic parts of the ligands can be treated in this way: the sign of $\sum f$ then determines whether the interaction is attractive or repulsive. An example may be instructive: Whereas Crippen has to incorporate two parameters for say an iodo substituent and an allyl substituent, we only use one parameter as the ratio of their $\sum f$ values is equal

Table II. Coordinates, Radii, and Types of Site Points of the Refined Model

| | X | Y | Z | R | |
|---|---|---|---|---|---|
| 1 | −4.05 | 0.07 | 0.51 | 0.15 | Accommodates β-OH of ethnaolamine side chain |
| 2 | −3.86 | 0.12 | −2.20 | 0.15 | Accommodates $N^+$ of ethanolamine side chain |
| 3 | 0.49 | 3.04 | 1.95 | 0.15 | Hydrophilic, type 1<br>Accommodates hydrophilic 3 substituents of phenylethanolamines and ring nitrogen of pindolol analogues |
| 4 | 0.01 | 4.79 | 0.00 | 0.15 | Hydrophilic, type 2<br>Accommodates hydrophilic 4 substituents of phenylethanolamines |
| 5 | −1.75 | 4.34 | −1.95 | 0.15 | Weakly lipophilic type<br>Accommodates 5 substituents of phenylethanolamines |
| 6 | −0.70 | −1.20 | 0.00 | 0.15 | Hydrophilic, type 2<br>Accommodates oxygen atom of phenoxy group of the phenoxypropanolamines |
| 7 | 0.69 | 1.20 | 0.00 | 0.40 | Strongly lipophilic type<br>Accommodates phenyl ring of phenoxypropanolamines |
| 8 | −1.38 | 2.39 | 0.00 | 0.40 | Weakly lipophilic type<br>Accommodates phenyl ring of phenylethanolamines and 3,4-benzo ring of propranolol analogues |
| 9 | −0.38 | 1.81 | 2.05 | 0.40 | Strongly lipophilic type<br>Accommodates heterocyclic ring of indole analogues, lipophilic ortho substituents of ortho-substituted phenoxypropanolamines and the 3,4-benzo ring of **51** and **53** |
| 10 | no coordinates deduced | | | | Weakly lipophilic type<br>Accommodates part I of the terminal amino chain (see Figure 3) |
| 11 | no coordinates deduced | | | | Weakly lipophilic type<br>Accommodates part II of terminal amino chain (see Figure 3) |
| 12 | no coordinates deduced | | | | Hydrophilic, type 1<br>Accommodates part III of terminal amino chain (see Figure 3) |

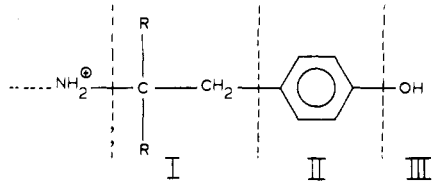



**Figure 3.** Binding of the terminal amino substituent of the ethanolamine side chain was represented by one, two, or three energy contributions in the minimization.

to the ratio of their respective bonding strengths to the considered lipophilic site point according to the used Collander equation (see Methods).

(2) Crippen uses a very large number of ligand point types and site point types. As a result the number of energy parameters is equivalently large. We do not use so many different ligand point types and site point types. The $NHSO_2CH_3$ group, for example, is considered to have the same type as an $NH_2$ group. Of course, this is a rough approximation, but in our philosophy it is preferable to deduce a rough model in the first run and to make the refinements afterwards.

(3) Whenever a certain ligand point type can interact with more site point types, we try to describe these interactions using only two types of interaction, as will be described hereafter.

Initially, we have used four energy parameters (two parameters for hydrophilic–hydrophilic interactions, one parameter for both hydrophobic–hydrophobic and hydrophobic–hydrophilic types of interactions, and one parameter for the interaction of the whole ethanolamine side chain with the receptor). However, it soon turned out that a second hydrophobic bonding parameter had to be used for interactions with site point 4 as these were found to be much stronger than interactions with the other hydrophobic points on the receptor site (i.e. site points 10 and 11). This is in accordance with results of other investigations in our institute, which indicate that the hydrophobic bonding strength of the phenyl ring of the phenoxypropanolamines is greater than hydrophobic bonding of the phenylethanolamine ring.[15] The energy minimization was undoubtedly in favor of binding mode B in Figure 1. The other model mispredicted the energies of the indole analogues and the compounds alprenolol **39** and the 2-I-substituted *N*-isopropylphenoxypropanolamine **38** by an unacceptable amount of 3 kcal/mol. On the basis of the given geometric and energetic considerations, the following conclusions were drawn. (1) The phenyl nuclei of the phenylethanolamines and phenoxypropanolamines occupy different parts of the receptor site (site points 10 and 4, respectively). (2) The 2,3-fused benzo ring of the propranolol analogues occupies the same site point as the phenyl ring of the phenylethanolamines (site point 10). (3) If the radius of site point 10 is taken somewhat larger, the heterocyclic ring of the indole moiety of the pindolol analogues occupies site point 10. The nitrogen atom of the heterocyclic ring is then able to occupy site point 1 which, in this preliminary model, is also occupied by the 4-OH of the phenylethanolamines.

**Refinement of the Hypothesized Receptor Site.** In this stage of the receptor mapping process, it is possible to refine the model to a more acceptable degree in a new run of our decomposition algorithm. The types and coordinates of the site points of the final model are depicted in Table II. No coordinates are given for site points representing substituents attached to the terminal nitrogen atom of the ethanolamine chain as the corresponding substituent ligand points are too flexible to allow a sufficiently accurate geometric interpretation.

For most ligands more than one conformation was compatible with the binding modes indicated by "FEASM". The torsional angles of the allowed conformation having the lowest energy were supposed to reflect the conformation in which a ligand binds to the hypothesized receptor site.

Initially it was tried to fit the data using only five energy parameters, i.e. one energy contribution for the ethanolamine chain, two hydrophilic–hydrophilic types of interaction, and one strong and one weak type of hydrophobic bonding. Although acceptable results were obtained, a careful examination of the differences between the ob-

(15) IJzerman, A. P.; Aue, G.; Bultsma, T.; Linschoten, M. R.; Timmerman, H. *J. Med. Chem.* **1985**, *28*, 1328.

Table III. Optimized Energy Parameters (kcal/mol) (95% Confidence Intervals in Parentheses)

| type ligand point | type site point | refined value |
|---|---|---|
| ethanolamine chain | specific interaction | −4.81 (±0.68) |
| −O− containing | hydrophilic type 1 | −0.78 (±0.33) |
| −O− containing | hydrophilic type 2 | −0.86 (±0.40) |
| −NH− containing | hydrophilic type 1 | −1.18 (±0.54) |
| any | weakly lipophilic | −0.51[a] (±0.17) |
| any | strongly lipophilic | −1.98[a] (±0.2o) |

[a] To be multiplied with lipophilicity contribution ($f$ system) of considered fragment. The obtained value is the binding energy of the fragment in kilocaloric/mole.

served and calculated binding energies indicated that phenylethanolamines which possess a 3-$NHSO_2CH_3$ group were predicted to have a more favorable binding energy than they have in reality. Therefore we split the hydrophilic–hydrophilic types of interactions in three types: two OH-hydrophilic types of interaction and one NH-hydrophilic type of interaction. Three binding energy contributions are taken into account for the terminal amino substituent, i.e. two hydrophobic–hydrophobic/hydrophilic interaction contributions and one contribution for a 4-substituent (Figure 3). Substituents attached to the $\alpha$-C atom of the ethanolamine chain were supposed to contribute to hydrophobic bonding. The final energy minimization then gave rise to the energy interaction parameters given in Table III.

The calculated vs. observed values are depicted in Table I. Attempts to get still better results by adding more parameters failed as it turned out that these parameters were not significant using 95% confidence limits. The plots in Figure 2 visualize the binding modes of the different classes of ligands when bound to the receptor.

The most important predictions of the model are as follows.

(1) The phenylnuclei of the phenoxypropanolamines and phenylethanolamines occupy adjacent parts of the receptor surface. The naphthyl head of the propranolol analogues occupies both regions.

(2) The 3-hydroxy substituent of the catecholamines and the nitrogen atom of the indole nucleus of the pindolol analogues are accommodated by the same binding group on the receptor.

(3) The binding of the classical antagonist dichloroisoproterenol (DCI), **37**, differs from that of the catecholamines in that its phenyl nucleus is flipped over its 1,4-axis.

(4) The 5-hydroxy substituent of 5-hydroxypropranolol **52** occupies the same region on the receptor as the 4-hydroxy substituent of the phenylethanolamines.

(5) Insertion of a $OCH_2$ group in the phenylethanolamines of the isoproterenol type does not result in another binding mode.

(6) Salbutamol **21** (C in Figure 2) binds to the receptor in its $S$ configuration. The introduction of an additional site point would allow for proper binding of the $R$ configuration. The lack of data prevents us from drawing a definite conclusion.

Although the general agreement between the calculated and observed binding energies is quite acceptable considering the inherent variation in biological data, a few outliers are present; AH2923 **20** is predicted to bind stronger than has been found experimentally. As the hydroxyl group of the 3-$CH_2OH$ substituent binds to site point 3, the phenyl ring is flipped over its 1,4-axis when compared to the binding mode of the catecholamines (compare the binding modes of salbutamol and isoproterenol in Figure 2). It is possible that in this case the substituent attached to the nitrogen atom of the ethanolamine chain cannot be properly accommodated by the site: this would mean that the ends of long terminal amino substituents are located somewhere near site point 8.

There is no explanation for the extraordinary strong binding of cobrefin **2** and W10773 **27**.

H64/52 **42** is the only 3-substituted phenoxypropanolamine in the data set. This compound binds much weaker than predicted. Inclusion of more 3-substituted phenoxypropanolamines in the data set might lead to the proposal of a filled site point in the complementary region of the receptor.

**Significance.** In an evaluation of the significance of our study in the field of receptor mapping, it may be best compared to the method of Crippen, from which it has been derived. The major difference between our method and Crippen's is the decoupling of the deduction of feasible matches and the energy calculations. The inclusion of a Collander type equation to describe binding of parts of the ligands to hydrophobic parts of the receptor gives rise to a reduction of the number of energy parameters. A substantially reduced number of ligand point types and site point types gives statistically significant results. According to our opinion the distance geometry approach to receptor mapping is an important new tool in three-dimensional QSAR techniques.