The Ensemble Approach to Distance Geometry: Application to the Nicotinic Pharmacophore

Robert P. Sheridan,* Ramaswamy Nilakantan, J. Scott Dixon,* and R. Venkataraghavan

Lederle Laboratories, Pearl River, New York 10965. *Received August 5, 1985*

We develop an extension of conventional distance geometry techniques that treats two or more molecules as a single "ensemble". This extension can be used to find a common pharmacophore, i.e., the spatial arrangement of essential groups, from a small set of biologically active molecules. The approach can generate, in one step, coordinates for the set of molecules in their "active" conformations such that their essential groups are superimposed. As an example, we show how the nicotinic pharmacophore can be deduced from a set of four nicotinic agonists: nicotine, cytisine, ferruginine methiodide, and muscarone. Three essential groups in each agonist are chosen: the cationic center (A), an electronegative atom (B), and an atom (C) that forms a dipole with B. There is only one pharmacophore possible for the superposition of these essential groups: a triangle with sides 4.8 Å (A-B), 4.0 Å (A-C), and 1.2 Å (B-C). The pharmacophore triangle, which is consistent with previous models in the literature, can also be achieved by the agonist *trans*-3,3′-bis[(trimethylammonio)methyl]azobenzene and the antagonists strychnine, trimethaphan, and dihydro-β-erythroidine. An examination of the common volumes of agonists suggests a specific disposition of molecular volume relative to the pharmacophore triangle. We discuss the relative strengths and drawbacks of the ensemble approach vs. other conformational search methods.

A central concept in medicinal chemistry is that of the pharmacophore, a specific three-dimensional arrangement of essential chemical groups common to active molecules, that is recognized by a single receptor. Although the concept involves many simplifying assumptions (a single binding mode, a single set of important interacting groups), it has proven useful in rationalizing pharmacological data. Much of the effort in modern pharmacology goes into developing and testing hypotheses about which chemical groups are important for a particular biological activity and what the three-dimensional arrangement of those groups is in the receptor-bound or "active" conformation of each molecule. These active conformations may not be the same as conformations in crystals or global minimum energy conformations either in vacuo or in solution. Since nothing is known about the atomic-level properties of the receptor associated with most biological activities, one must deduce the pharmacophore from a set of active and inactive ligand molecules, some or all of which may be conformationally flexible.

The first step in deducing a pharmacophore is to choose the groups essential for activity in each molecule. A group can be an atom, a geometrically defined point within the molecule (e.g., the center of a phenyl ring), or a "receptor point" with a fixed relationship to specific atoms in the molecule (e.g., a hydrogen-bond acceptor along a N-H axis in the molecule). For the purpose of discussion we will assume three such groups A, B, and C. (In principle, there is no restriction on the number of groups; however, three noncolinear groups is the minimum necessary to dock two molecules in three-dimensional space.) The equivalent of A (A_1, A_2, A_3, ..., A_M), B, and C must appear in each of M molecules and a three-dimensional arrangement of A, B, and C common to all molecules must be attainable. The problem then becomes: Find low-energy conformation(s) for each molecule such that A_1 is superimposable with A_2, A_3, etc., B_1 is superimposable with B_2, B_3, etc., and C_1 is superimposable with C_2, C_3, etc. If there is not at least one conformation of each molecule in which all the equivalent groups from all molecules can be superimposed, then, assuming the activity data is correct and the assumptions of the pharmacophore hypothesis hold, the initial choice of A, B, and C in at least one molecule is incorrect and must be revised. If there is more than one way to superimpose the equivalent groups, then not enough constraints have been applied to find a unique solution and additional molecules must be considered.

In this paper, we describe a novel "ensemble" approach to distance geometry that generates three-dimensional coordinates for a set of molecules such that equivalent groups are superimposed. We apply the approach to finding the pharmacophore that is recognized by the nicotinic acetylcholine receptor and discuss the advantages and disadvantages of our approach in the context of previous methods.[1-7]

The nicotinic acetylcholine receptor has been the focus of intense research in recent years. (See the review by Changeux et al.[8]) Progress has been made in developing a model for this transmembrane multimeric protein receptor. The acetylcholine-binding portion has been found to be within a stretch of 20 amino acid residues,[8,9] and hypotheses about the conformation of this stretch have been proposed (for example, Smart et al.[10]) Atomic-level data about this receptor, however, is at least a few years away. Several attempts have been made to deduce the nicotinic pharmacophore for both agonists and antagonists.[11-13] Our primary aim is not to propose a new model for the nicotinic pharmacophore but to demonstrate the use of the ensemble approach in deducing the pharmacophore from a small set of agonists. However, this is the first time that an automated pharmacophore search procedure has been applied to nicotinic agonists and antagonists.

Methods

Summary of Distance Geometry. Distance geometry addresses the problem: Given a set of N points and a matrix describing the distance between each pair of points,

(1) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; Dunn, D. A. In *Computer-Assisted Drug Design*; Olson, E. C., Christoffersen, R. E. Eds.; ACS Symp. Series 112; American Chemical Society: Washington, DC, 1979; pp 205-226.
(2) Schulman, J. M.; Sabio, M. L.; Disch, R. L. *J. Med. Chem.* **1983**, *26*, 817-823.
(3) Crippen, G. M. *J. Med. Chem.* **1979**, *22*, 988-997.
(4) Crippen, G. M. *J. Med. Chem.* **1980**, *23*, 599-606.
(5) Crippen, G. M. *J. Med. Chem.* **1981**, *24*, 198-203.
(6) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1982**, *25*, 892-899.
(7) Ghose, A. K.; Crippen, G. M. *J. Med. Chem.* **1984**, *27*, 901-914.
(8) Changeux, J.-P.; Devillers-Thiery, A.; Chemouilli, P. *Science* **1984**, *225*, 1335-1345.
(9) McCormick, D. J.; Atassi, M. Z. *Biochem. J.* **1984**, *224*, 995-1000.
(10) Smart, L.; Meyers, H.-W.; Hilgenfeld, R.; Saenger, W.; Maelicke, A. *FEBS Lett.* **1984**, *178*, 64-68.
(11) Pauling, P.; Petcher, T. J. *Chem.-Biol. Interact.* **1973**, *6*, 351-365.
(12) Beers, W. H.; Reich, E. *Nature (London)* **1970**, *228*, 917-922.
(13) Palmer, R. A.; Tickle, J. H.; Tickle, I. J. *J. Mol. Graphics* **1983**, *1*, 94-106.

0022-2623/86/1829-0899$01.50/0

generate coordinates for the points in three-dimensional space such that the distances are best satisfied. In most applications of distance geometry to chemistry, where molecular flexibility is important, one uses upper and lower bounds to the distance between each pair of points (with most points being atoms) rather than a single fixed distance. The original practical computer algorithm for generating final coordinates for molecules was developed by Crippen and co-workers and we refer the reader to his 1981 monograph[14] and a later review[15] for details. The algorithm can be summarized as follows:

(1) For each pair of points generate an upper and a lower bound for the distance between them. Note the absolute handedness of chiral centers.

(2) Smooth the bounds. That is, lower some of the upper bounds and raise some of the lower bounds by exhaustive application of the triangle inequality. This generates new upper and lower bounds for each pair of points. If a violation of the triangle inequality is found, indicating the bounds are contradictory, stop. Otherwise...

(3) For each pair of points randomly select a trial distance between the new upper and lower bounds.

(4) Construct coordinates in $N - 1$ space, where N is the number of points in the problem, from the randomly selected distances.

(5) Project the coordinates into three-dimensional space such that the distances in $N - 1$ space are best preserved. If the projection cannot be made or if the distance bounds cannot be obeyed in three dimensions (i.e., if the coordinates violate distance bounds too severely), go back to step 3. Otherwise...

(6) Refine the three-dimensional coordinates using a potential function that penalizes violations of the upper and lower distance bounds and penalizes "chirality" violations. Write a set of coordinates for the final refined structure.

Steps 3–6 are repeated for each structure to be obtained. Each final structure represents a solution to the problem of finding atom positions such that the distance constraints are best obeyed. Because the selection of distances in step 3 is random, each final structure represents a Monte Carolo sampling of conformation space within the constraints of the distance bounds. Since there is coupling between various distances due to the triangle inequality, the selection for any one distance cannot be truly random, and conformation space cannot be uniformly sampled. However, the final conformations are sufficiently diverse for most purposes.[14,16]

Molecular Ensemble Treatment. The key to our distance geometry application is in the treatment of two or more molecules as a single ensemble. This requires modifications to the standard distance geometry algorithm in the construction of the distance matrix and assignment of chiral centers (step 1 above). We developed a program ENSEMBLE for this purpose. Steps 2–6 are performed by the subroutine EMBED, which was developed by Havel et al.[15] and which we have licensed from the University of California at San Fransisco. As input for ENSEMBLE, we use one or more previously generated coordinate sets for the molecule(s) of interest. These sets may be transcribed from crystal coordinates or be generated by any model-building program. The coordinate sets must contain proper bond distances, bond angles, and proper dihedral angles for nonrotatable bonds (e.g., C—N bonds in amides, C=C double bonds, bonds in aromatic rings, etc). However, since the final structures are independent of the conformation in these coordinate sets, no attention need be paid to the dihedral angles around rotatable bonds. For each molecule we store a bond matrix, note the chiral centers and note the rotatable bonds.

Construction of the ensemble distance matrix is as follows: Let the atoms of molecule 1, molecule 2, molecule 3, ... collectively represent a set of points 1, 2, ..., i, ...N. Consider the distance matrices **L** and **U** for these points. Let **U**(i,j), be the upper bound distance between atom i and j. Let **L**(i,j) be the lower bound distance. (Because the matrices are symmetric, only the upper triangle of each matrix is considered.) By definition of distance

$$\mathbf{U}(i,i) = \mathbf{L}(i,i) = 0$$

Initially set

$$\mathbf{U}(i,j) = 100\ \text{Å}$$

(an arbitrarily large distance) for all i and j. The lower bound is

$$\mathbf{L}(i,j) = \text{sum of the van der Waals radii of atoms i and j}$$

if i and j are from the same molecule. That is, intramolecular hard-sphere close contacts are forbidden. On the other hand

$$\mathbf{L}(i,j) = 0$$

if i and j are from two different molecules and these two molecules are to be superimposed. That is, different molecules are allowed to pass through each other unless otherwise specified. If two molecules are to interact sterically, **L**(i,j) is set to the sum of the van der Waals radii if i and j are from those specified molecules. When either atom i or j is a "dummy" atom, i.e., a point that is included to help define a geometry but which has no volume, **L**(i,j) is always set to 0.

Certain distances defining the covalent structure of each molecule replace the previously defined bounds:

$$\mathbf{U}(i,j) = \mathbf{L}(i,j) = d(i,j),$$

where d is the distance between atoms i and j. Atoms i and j are from the same molecule and form the bond i–j. Similarly

$$\mathbf{U}(i,k) = \mathbf{L}(i,k) = d(i,k)$$

where atoms i and k form the bond angle i–j–k, and

$$\mathbf{U}(i,m) = \mathbf{L}(i,m) = d(i,m)$$

where atoms i and m form the dihedral angle i–j–k–m and j–k is not a rotatable bond. In all cases d is taken directly from the input coordinate sets.

Any additional upper and lower bounds are then added to the matrices **U** and **L**. For example, additional bounds would be set for the atoms of each molecule to be superimposed. If atoms i and j from different molecules are to be superimposed

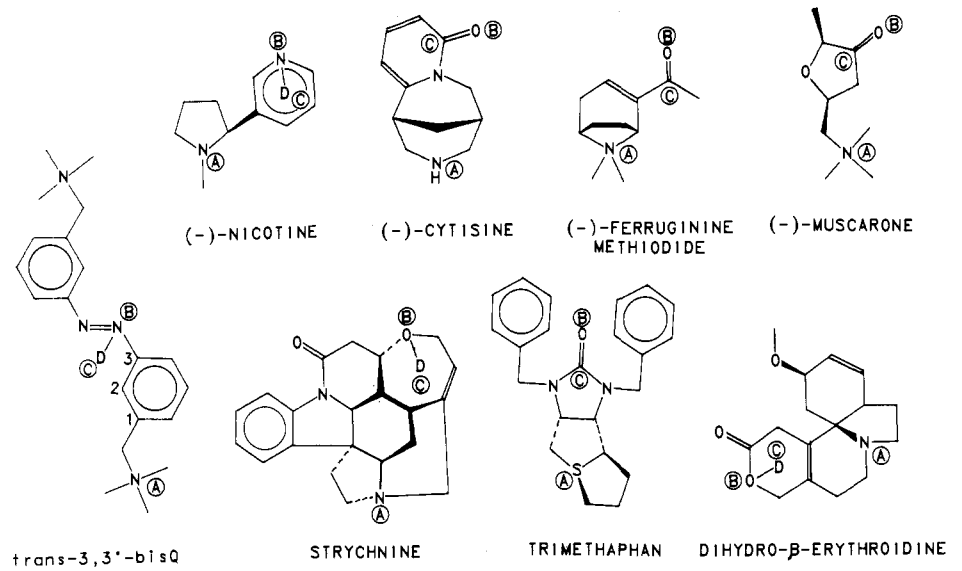
$$\mathbf{L}(i,j) = 0$$
$$\mathbf{U}(i,j) = t$$

where t is a "tolerance" parameter for superposition; that is, no pair of equivalenced atoms is to be farther apart than t in the final structure.

The function we use to refine the three-dimensional coordinates is similar to that used previously[14,15] and contains terms for distance bounds violations and for chirality violations. The latter deserves some comment. The absolute handedness of chiral centers cannot be addressed directly in terms of distances. Instead, the signed

(14) Crippen, G. M. *Distance Geometry and Conformational Calculations*; Research Studies Press: New York, 1981.
(15) Havel, T. F.; Kuntz, I. D.; Crippen, G. M. *Bull. Math. Biol.* **1983**, *45*, 665–720.
(16) Havel, T. F.; Wuthrich, K. *Bull. Math. Biol.* **1984**, *46*, 673–698.

Chart I. Nicotinic Agonists and Antagonists Used in This Study[a]



[a] The four molecules in the top row were used to derive the common nicotinic pharmacophore via the ensemble treatment. All are agonists. The molecules in the bottom row were checked for consistency with the pharmacophore. *trans*-3,3'-bisQ is an agonist; the remaining molecules are antagonists. Atoms labeled D are "dummy" atoms along the bond angle bisector and 1.2 Å from the atom to which they are attached. In each molecule the circled letters A, B, and C indicate the final choice of pharmacophore atoms.

volume f is calculated for each tetrahedron formed by the four atoms attached to each chiral center (usually tetravalent carbon). If there are only three explicit neighbors, the tetrahedron is formed by the three neighbors and the central atom. Deviations from the initial value of f are penalized. Sometimes it is useful to treat *chemically* nonchiral atoms as chiral. For instance, so treating carbonyl carbons improves the planarity of carbonyl groups.

In addition to chirality terms, we include bond dihedral terms. A signed volume is calculated for the tetrahedron formed by atoms i, j, k, and m, which form a dihedral angle if the j–k bond is not rotatable. Deviations from the initial value of f are penalized as for chirality violations. Including such dihedral terms greatly improves the planarity of aromatic rings, amides, esters, etc. in the final structures.

In our current program, refinement is done in three steps. First the chirality function alone is minimized for a preset number of iterations; then the chirality function and the distance-bounds violation functions are minimized together. Finally, the chirality function, the distance-bounds violation function, and the dihedral function are minimized together.

Application to Nicotinic Agonists and Antagonists. We illustrate the use of the ensemble application by finding a common pharmacophore of four semirigid potent nicotinic agonists: (-)-nicotine, (-)-cytisine, (-)-ferruginine methiodide,[17] and (-)-muscarone. These appear in the top row in Chart I. These molecules were chosen because they are small and structurally diverse. In addition, the assignment of essential groups is fairly straightforward in these molecules. All of the molecules are conformationally restricted, but none is so rigid that the pharmacophore can be deduced from that molecule alone. (+)-Nicotine is somewhat less active than (-)-nicotine[18] and (+)-muscarone is marginally less active than (-)-muscarone.[19] While we do not use these less active enantiomers in the original ensemble, we consider them later.

We check the pharmacophore derived from the four agonists against a set that includes an agonist and three antagonists. This set consists of, respectively, *trans*-3,3'-bis[(trimethylammonio)methyl]azobenzene (*trans*-3,3'-bisQ),[20] strychnine, trimethaphan, and dihydro-β-erythroidine. These molecules are shown in the bottom row of Chart I.

We built sets of three-dimensional coordinates for each molecule in Chart I with the program CHEMGRAF[21] using standard bond lengths and angles. For the purposes of this study, we represented =CH—, —CH_2—, and —CH_3 groups as "united atoms" with increased van der Waals radii. Dummy atoms were added, where needed, along bond bisectors 1.2 Å from the appropriate atom.

The selection of superimposable groups from each molecule is the critical step in finding common pharmacophores no matter what method is used to explore conformational space. Three sets of atoms, labeled A, B, and C in Chart I, were selected as the groups. The cationic centers (quaternary nitrogen or protonatable nitrogen) in each molecule are obviously equivalent. They form set A. Nicotinic agonists almost always contain an electronegative atom or center B which may act as a hydrogen-bond acceptor. For nicotine, cytisine, and ferruginine methiodide there is only one such atom, a carbonyl oxygen. For muscarone, there is a choice of two oxygens. To be consistent across the set, we chose the carbonyl oxygen. As the third atom C, we chose the carbonyl carbon for all molecules except nicotine, for which we used a dummy atom. The line connecting B and C defines the direction of the local dipole moment along which a hydrogen bond is likely to form.

Treating the four agonist molecules collectively as an ensemble and using t = 0.3 Å for the tolerance for superimposing equivalenced atoms, we generated the upper

(17) Spivak, C. E.; Waters, J.; Witkop, B.; Albuquerque, E. X. *Mol. Pharmacol.* **1983**, *23*, 337–343.
(18) Aceto, M. D.; Martin, B. R.; May, E. L. In *CRC Handbook of Stereoisomers: Drugs in Psychopharmacology*; Smith, D. F., Ed.; CRC Press: Boca Raton, FL, 1983; pp 67–78.
(19) Gyrmek, L.; Unna, K. R. *J. Pharmacol. Exp. Ther.* **1960**, *128*, 30–36.
(20) Wasserman, N. H.; Bartels, E.; Erlanger, B. F. *Proc. Natl. Acad. Sci. U.S.A.* **1979**, *76*, 256–259.
(21) ChemGraf, created by E. K. Davies, Chemical Crystallography, Oxford University, developed and distributed by Chemical Design Ltd., Oxford.

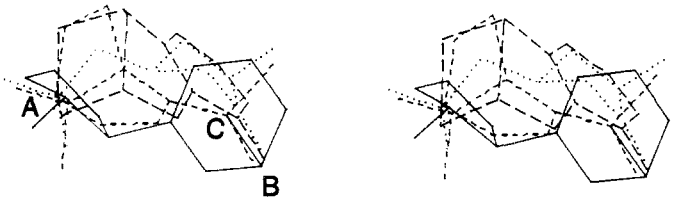


Figure 1. A stereo picture of one typical solution from the ensemble distance geometry treatment of the four agonists: (−)-nicotine (solid line), (−)-cytisine (long dash), (−)-ferruginine methiodide (short dash), (−)-muscarone (dotted). The conformation of each molecule is such that various distance and chirality constraints are best obeyed. The distance constraints are derived from the covalent structure of each molecule, from the requirement that there be no intramolecular bad-contacts, and from the requirement that equivalent pharmacophore atoms (labeled A, B, and C in Chart I) from each molecule be superimposed with a tolerance of 0.3 Å.

and lower distance bounds matrices as discussed above. We then followed the standard distance geometry algorithm to produce docked conformations of the molecules. We allowed each step in the refinement to proceed for 1000 iterations.

Once a possible pharmacophore has been derived from the ensemble approach, we can check molecules individually to see if they can reach the pharmacophore geometry. We treated each molecule in the bottom row of Chart I individually within the distance geometry formalism with the constraints that the A–B, A–C, and B–C distances be within the range of distances consistent with the pharmacophore. Again we chose the cationic center of each molecule as group A. There was only one set of candidates for group B and C for *trans*-3,3′-bisQ (the azo nitrogen and associated dummy atom) and for trimethaphan (the ureido oxygen and adjacent carbon). For strychnine there are two oxygens (and associated atoms) that are candidates for group B (and C). For dihydro-β-erythroidine there are three candidates. Each possibility was tried for these two molecules.

Results

Deriving the Pharmacophore. The set of equivalenced atoms for the four superimposed agonists did not give rise to violations of the triangle inequality. An alternate set, wherein the ether oxygen rather than the carbonyl oxygen is chosen as group B in muscarone, did lead to a violation; this impies that no superposition is possible for the alternate set.

Generation of 25 solutions takes ca. 3.5 CPU hours on our VAX 11/785 with floating point accelerator. Each ensemble solution represents a collective structure for all four superimposed agonists. EMBED records the root-mean-square violation (R) of the interatomic distances in each final refined structure from the smoothed upper and lower bounds. For the set of 25 structures the values of R cluster around 0.06 and 0.10. Inspection of the structures reveals that 15 solutions with $R = 0.10$ contain either distorted bond lengths and angles or bad van der Waals contacts. These represent cases where the initial three-dimensional coordinates could not be refined to accommodate the distance bounds and chirality contraints. In contrast, those 10 structures with $R = 0.06$ contain no such distortions. We therefore consider only these 10 "good" structures for further analysis. One of these structures is shown in Figure 1. We can expect each individual molecule in each of the structures to be in an energetically allowed conformation since any intramolecular hard-sphere contacts, disallowed in the lower bounds matrix, have been eliminated during refinement.

Table I. Range of Distances between Pharmacophore Atoms A, B, and C over 10 Ensemble Solutions

molecule	distance, Å		
	A–B	A–C	B–C
nicotine	4.5–4.9	3.8–3.9	1.2[a]
cytisine	4.9–5.0	4.2–4.3	1.2
ferruginine methiodide	4.4–4.6	3.7–3.8	1.2
muscarone	4.9–5.0	4.2–4.4	1.2

[a] This distance is fixed by the C=O covalent bond length.

Similarities between molecular conformations in the 10 structures are measured by the conventional method of root-mean-square coordinate deviation (rms) after centroid superposition and least-squares rigid-body rotation.[22] Among the 10 good structures, we see only a single solution for the pharmacophore common to the four agonists. The rms deviation between the corresponding 12 (four sets of three) "pharmacophore" atoms is no more than 0.2 Å between any pair of structures; this is well within the superposition tolerance of 0.3 Å. The range in the distances between pharmacophore atoms in each molecule over the 10 structures is given in Table I. The A–B distance varies from 4.4 to 5.0 Å and the A–C distance varies from 3.7 to 4.3 Å. The B–C distance 1.2 Å is fixed by the length of the standard C=O bond or, in the case of nicotine, by the defined distance between the dummy atom and the pyridine nitrogen. The idealized pharmacophore triangle, then, has sides 4.8 ± 0.3, 4.0 ± 0.3, and 1.2 Å.

Although there is only one pharmacophore common to all the agonist molecules in the ensemble, there may be more than one conformation of each molecule that can reach the pharmacophore geometry. For convenience in comparing the conformations generated by distance geometry, we consider two conformations to be in the same "family" if the rms deviation between them is less than 0.3 Å. For cytisine, which is by itself partly flexible, there is only one family of conformations compatible with the pharmacophore geometry. For nicotine there are clearly two families of conformations (six examples of the first and four of the second in the 10 structures). The families differ by a rotation around the bond between the pyrolidine and pyridine rings as well as in the pucker of the pyrolidine ring. There are two families for ferruginine methiodide (six of the first and four of the second). The families differ by a rotation around the bond connecting the carbonyl group to the bicyclic nucleus. We treat this bond as rotatable because the equivalent bond in the related agonist anatoxin appears to be less conjugated than expected for an enone.[23] There are at least three families (seven of the first, two of the second, one of the third) for muscarone, which differ in the ring pucker and the dihedral angle around the first bond connecting the ring to the quaternary nitrogen. Where there are two or more families of conformations compatible with the pharmacophore geometry, there is no way to tell which is the active one for an individual agonist. However, as we discuss later, the superposition of volumes of several agonists is enlightening.

Testing the Pharmacophore. For each of the molecules used to test the pharmacophore geometry, we generated 10 solutions with the constraint that the A–B distance be 4.6–5.0 Å and the A–C distance be 3.7–4.3 Å (the B–C distance is already fixed). In each case, $R < 0.06$ for all 10 structures.

Strychnine is essentially rigid even without the imposition of the pharmacophore constraints. Even so, this

(22) Nyburg, S. C., *Acta Crystallogr., Sect B* **1974**, *30*, 251–253.
(23) Koskinen, A. M. P.; Rapoport, H. *J. Med. Chem.* **1985**, *28*, 1301–1309.

molecule fits the pharmacophore geometry fairly well. Only the oxygen so labeled in Chart I is a candidate for group B, the fixed A–B distance being 5.0 Å. The largest departure from the pharmacophore geometry is in the A–C distance: 4.6 Å vs. the ideal range of 3.7–4.3 Å.

The ring nucleus of trimethaphan has a slight freedom to pucker without the pharmacophore constraints. Only one particular pucker fits the pharmacophore geometry.

Two oxygen atoms in dihydro-β-erythroidine can reach the ideal A–B distance, the lactone oxygen (labeled B in Chart I) and the methoxy oxygen. On the basis of structural analogy to strychnine, we prefer the lactone oxygen. There are at least three families of conformations of dihydro-β-erythroidine that fit the pharmacophore geometry with the lactone oxygen as group B. The three families have different puckers in the outer rings.

In *trans*-3,3′-bisQ, the benzene rings and the azo groups must be in the same plane, but there are two possibilities for the N–N–C3–C2 dihedral angle, 0° and 180°. We tried both possibilities separately. The A–B distance is independent of the dihedral angle, but the proper A–C distance (about 4.0 Å) can be reached only if the dihedral angle is 0°. The only other degree of freedom is the position of the trimethylammonium group relative to the benzene ring. There are essentially only two possible values (symmetrically related on either side of the ring) for the dihedral angle around the appropriate bond. Because of steric interference of the trimethylammonium group with the ring, the A–B distance is slightly long: 5.2 Å vs. the ideal range of 4.5–5.0 Å.

Steric Requirements for Agonists. Little is known about the stereoselectivity of nicotinic agonists, but there is evidence that at least one agonist is strongly stereoselective.[17] (–)-Cytisine is the only active chiral agonist for which there is a single conformation that can achieve the pharmacophore geometry. Imagine that the pharmacophore atoms of (–)-cytisine in that conformation are oriented in the plane of the page so that the letters A, B, C are arranged counterclockwise. The bulk of the volume of this molecule is then in front of the plane. This orientation of the bulk of the volume relative to the pharmacophore triangle defines an absolute "handedness". It is reasonable to assume that the receptor cavity has a shape roughly complementary to (–)-cytisine when agonists are bound and that a molecule with the opposite handedness would not activate the receptor.

When a molecule has two or more families of conformations generated by distance geometry that can attain the pharmacophore geometry, those families with the correct handedness, i.e., those conformations that most resemble (–)-cytisine in shape, are most likely to represent the agonist-like or "activating" conformations. For (–)-muscarone, (–)-ferruginine methiodide, (–)-nicotine, and *trans*-3,3′-bisQ, only one of the two or three families has the correct handedness. The slightly less active enantiomers (+)-nicotine and (+)-muscarone can also attain the correct handedness. For all the agonists, the correct handedness can also be defined by noting that groups B and C are contained in a planar array of atoms and that group A is behind the plane of those atoms. Out-of-plane arrangement of the cationic center has been suggested previously[17,20] as being important for agonist activity. Our best guess for the activating conformation of each agonist is shown in Figure 2. Also shown is the conformation of each antagonist that can attain the pharmacophore geometry.

Having selected the likely activating conformation for each agonist, we can dock the agonists together by a simple least-squares rigid-body superposition of the pharmacophore atoms onto an ideal pharmacophore triangle. When this is done, the volumes of all the agonists overlap to a great extent. The union of the volumes for all the agonists defines the space that an agonist may occupy. This volume is shown in Figure 3.

Discussion

Previous Models of the Nicotinic Pharmacophore. Previous authors have proposed models for the nicotinic pharmacophore. Pauling and Petcher[11] noted in the crystal structures of large rigid antagonists like tubocurarine an 11-Å separation between quaternary nitrogens and, in some cases, an oxygen 5 Å from each nitrogen. Palmer et al.[13] examined a set of steroidal blockers and suggested a pharmacophore consisting of a quaternary nitrogen and an electronegative site (either a carbonyl oxygen or the π cloud of a double bond) 5–6 Å away. Wasserman et al.[20] proposed a pharmacophore based on the structure of *trans*-3,3′-bisQ: an electronegative atom in a hydrophobic plane with a cationic center 1.5 Å above the plane and 5.2 Å from the electronegative atom.

Since these previous authors explicitly considered only two groups, a cationic center and an electronegative atom, we can say only that the A–B distance in our pharmacophore is consistent with their proposals. Beers and Reich,[12] however, dealt with three implicit groups. They considered the agonists cytisine and nicotine and the antagonists strychnine, trimethaphan, and β-erythroidine. The first group is the cationic center of a molecule, equivalent to our point A. The second group is a "receptor point", which we will call D, defined relative to a selected electronegative atom, equivalent to our group B. D is along the C=O bond (for carbonyl O as the electronegative atom) or angle bisector (for –O– or –N– as the electronegative atom) and at the van der Waals radius of the electronegative atom. The A–D distance of 5.9 Å proposed by Beers and Reich is geometrically compatible with our pharmacophore triangle A–B–C. Our equivalent of point D would be along the C → B direction 1.4 Å (the van der Waals radius of O) from B. Given the triangle distances 4.8 Å (A–B), 4.0 Å (A–C), and 1.2 Å (B–C), the A–D distance is then determined to be almost exactly 5.9 Å. Not surprisingly, where we consider the same molecules, our choice of essential groups and active conformations resembles that of Beers and Reich.

An interesting point of contrast between this and previous work is in the selection of molecules from which to derive the pharmacophore. Most previous authors derived their pharmacophores from the more rigid molecules, usually antagonists (there are several completely rigid antagonists that are suitable), and then considered the more flexible agonists. For this type of approach a conformational search method is not necessary. In contrast, we derive our pharmacophore from flexible agonists using a general conformational search method and then show that an additional agonist and three antagonists can reach the pharmacophore. It is encouraging to note that, although we consider a different set of molecules and use a different method, we obtain a similar pharmacophore to that obtained by previous authors.

Our current idea about the nicotinic binding site incorporates previous pharmacophore models and is consistent with the structures of protein toxins,[24] nearly

(24) Walkinshaw, M. D.; Saenger, W.; Maelicke, A. In *Structural Aspects of Recognition and Assembly in Biological Macromolecules*; Balaban, M., Sussman, J. L., Traub, W., Yonath, A., Eds.; Balaban ISS: Philadelphia, 1981; pp 313–324.

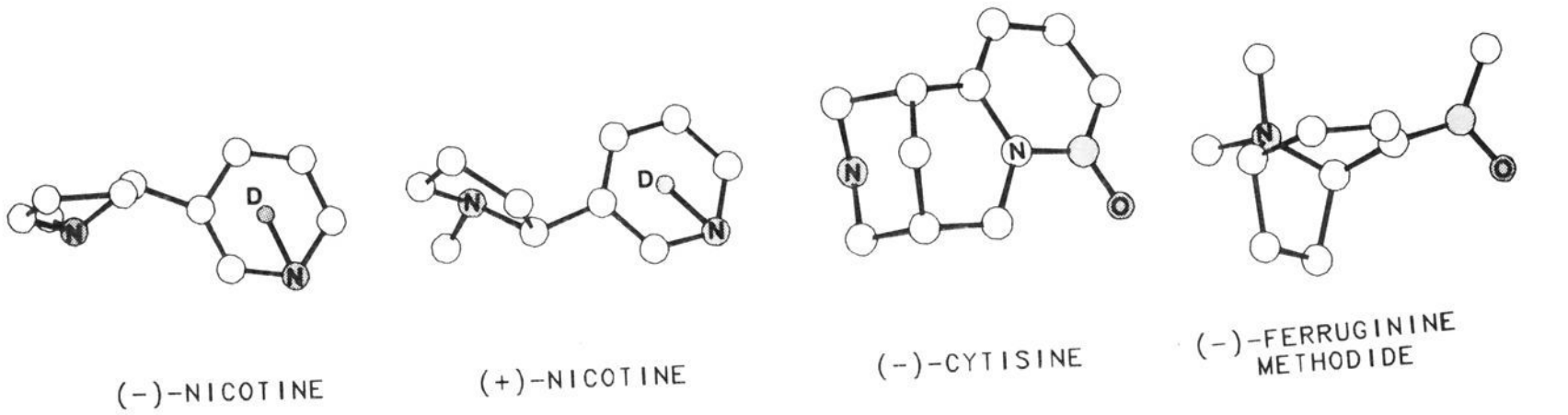


Figure 2. Likely active conformations of nicotinic agonists and antagonists derived from distance geometry. Each conformation is one member of a "family" of closely related conformations that can achieve the pharmacophore geometry. The agonists also have the same handedness of the molecular volume relative to the pharmacophore triangle as (−)-cytisine. All agonists (top two rows) are oriented so that the pharmacophore atoms A, B, and C are in the plane of the page. The idealized pharmacophore triangle (box) is shown in the same orientation. For clarity, the antagonists (bottom row) are oriented with the A–C side of the pharmacophore triangle tilted toward the reader. The orientation of the phenyl rings in trimethaphan cannot be derived from the pharmacophore triangle and the rings are positioned arbitrarily. There are at least three families of conformations of dihydro-β-erythroidine (which differ in the pucker of the methoxy-substituted ring); one is shown.

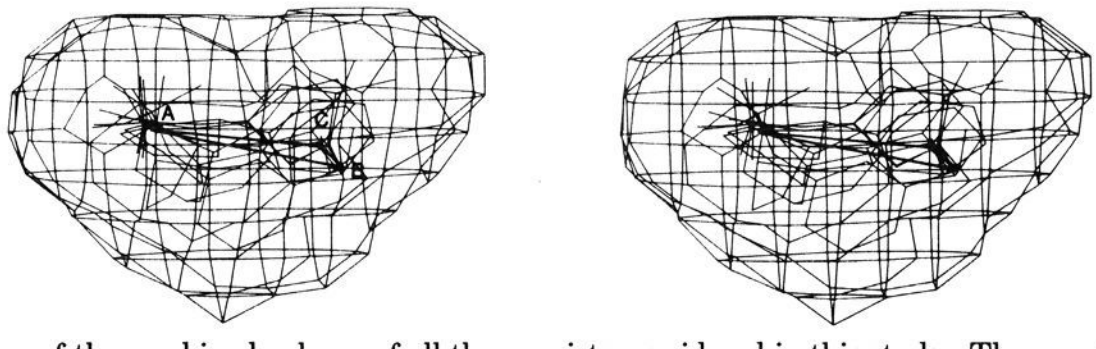


Figure 3. A stereo picture of the combined volume of all the agonists considered in this study. The agonists, in the conformations shown in Figure 2, were docked so that the pharmacophore atoms were best superimposed with an ideal pharmacophore triangle (bold lines). The symmetrical structure of *trans*-3,3′-bisQ suggests that the binding site for agonists is symmetrical and able to accommodate two pharmacophore triangles. We show only one half of such a binding site and truncate *trans*-3,3′-bisQ accordingly.

symmetrical antagonists like tubocurarine,[11] and symmetrical agonists like *trans*-3,3′-bisQ.[20] The binding site is thought to contain two anionic or electronegative groups, 11 Å apart, each capable of interacting with a cationic center on the agonist or antagonist. About 5 Å away from each anionic group is a donor group that can hydrogen

bond to an electronegative atom. (Our pharmacophore, which contains only one cationic center, would interact with half of such a binding site.) Nothing in this model for the binding site distinguishes agonists from antagonists. However, there is experimental evidence that, although agonists and antagonists compete for the same binding sites (two per multimeric receptor protein), they might not bind in exactly the same way. Also, agonists and antagonists may differentially bind to different quaternary states of the receptor protein. (See the review by Conti-Tronconi and Raftery.[25]) Since the pharmacophore geometry we and previous authors derive is attainable by agonists and diverse types of antagonists, it is probable that both agonists and antagonists are recognized by the anionic and donor groups in the binding site. Some property other than the pharmacophore geometry must be important for agonism. One possible suggestion is that to be an agonist a molecule must not only be able to achieve the pharmacophore geometry but also be confined to a specific volume. Antagonists are almost always large molecules, which if docked onto the agonists, would certainly extend outside the agonist volume shown in Figure 3. It may be that the extra bulk of antagonists prevents the receptor protein from reaching the open-channel quaternary form.

Comparison of Ensemble Approach with Previous Approaches. Two approaches have been previously explored for finding conformations of flexible molecules for which superposition of essential groups is possible. All methods, including ours, depend on the assumptions of the pharmacophore hypothesis and are limited by our ability to pick out the right combination of groups to constitute the pharmacophore for each molecule. The first approach was developed by Marshall and co-workers.[1] For each active molecule, each rotatable bond is systematically rotated by a fixed increment so that all possible combinations of dihedral angles are generated. For each allowed conformation (a conformation for which there are no intramolecular bad contacts), the distances between pharmacophore groups are saved. Those sets of intergroup distances that can be achieved by all active molecules represent possible pharmacophore geometries. Several other workers have used similar concepts while changing the strategy slightly (for example Schulman et al.[2]). A second approach comes from Crippen and co-workers.[3-7] For each molecule, rotatable bonds are systematically rotated and the upper and lower bounds over all allowed conformations are recorded for the distances between pharmacophore groups. For each intergroup distance the greatest lower bound and least upper bound among all the molecules are found. These define upper and lower bounds for the intergroup distances common to all molecules. Distance geometry is used to generate a three-dimensional arrangement of "site points" from the common upper and lower bounds. This binding-site model can be used to dock additional molecules and rationalize binding data, but only the initial steps that generate the binding-site model, equivalent to a pharmacophore model, are of interest to us here.

The ensemble method differs greatly from previous methods in the order in which structural information is processed. Both previous methods could be called "incremental" methods. They start with three-dimensional representations of individual molecules, generate intergroup distance information by systematic rotation, then intersect the information from each molecule to find intergroup distances common to all molecules (i.e., to find the pharmacophore geometry). In contrast, the ensemble method *starts* with a distance bounds representation of a set of molecules, wherein the superposition of groups between the molecules is assigned directly. Three-dimensional coordinates of all the molecules are generated and refined together as the last step. The pharmacophore geometry is found by analyzing the final structures. For a set of points in three dimensions, there are more interpoint distances ($N(N-1)/2$) than degrees of freedom ($3N-6$) and the interpoint distances are necessarily correlated. Any method that represents conformations as distances must take the correlation into account at some stage. In the previous methods, which start with three-dimensional molecules, the correlation information is implicitly included when the distance information is generated. In the ensemble method, any correlation information that is not included in the initial distance bounds is accounted for during the refinement of the three-dimensional structures.

An important advantage of treating a number of molecules as an ensemble is that certain critical intermolecular information, which is important for certain applications, can be included. For example, the possibility of representing steric interactions between molecules, impossible in the incremental approach, allows us to use "excluded volume" information. Also, in the incremental approach only intramolecular distance information is shared between molecules. Finding common intergroup distances among a set of molecules is a necessary condition for the groups in all the molecules to be superimposable but not a sufficient one for more than three groups, since distance information alone does not take absolute handedness into account. In the refinement step of the ensemble approach, in which the superpositions between molecules are treated directly, handedness information is taken into account for all molecules simultaneously. The ensemble approach is, of course, limited by the fact that finding distance geometry solutions becomes more computationally intensive as the number of atoms in the problem (i.e., the number of molecules in the ensemble) increases. The practical limit for EMBED is about 200 atoms. However, more efficient algorithms, which can handle up to several hundred atoms,[16] have recently been developed.

Representing molecular structures as a set of distance bounds has great advantages in treating those aspects of structure that are more easily expressed in local distances than in dihedral angles. An important example where this is true is that of flexible rings. All possible ring puckers are consistent with a single set of 1–2 bond lengths and 1–3 distances and are allowed as distance geometry solutions (within the constraints of steric interactions between ring substituents). In contrast, the dihedral angles in a flexible ring are strongly correlated and it is difficult to produce systematic rotations around the bonds such that the ring maintains its connectivity. This is especially true for fused-ring systems. The ability of the ensemble approach to represent complex flexible ring systems in a simple way is, we believe, the most important improvement over previous methods. Most of the molecules studied in this paper could not be properly treated earlier. There are aspects of molecular structure, on the other hand, that cannot be simply handled by a distance geometry method. In a systematic search method it is easy to specify that a bond may have only two possible dihedral angles (e.g., cis or trans), but this is impossible to specify in a single set of distance bounds.

Any method that includes a systematic search is always limited by the fact that the computational time to com-

(25) Conti-Tronconi, B. M.; Raftery, M. A. *Ann. Rev. Biochem.* 1982, *51*, 491–530.

plete the search goes up very rapidly with the number of rotatable bonds. Despite clever schemes to make the searches more efficient, the method described by Marshall et al.[1] is limited to a fairly small number of rotatable bonds (less than eight) or a fairly coarse (30° increment) search "grid". The computational time for the ensemble approach, which does not require a search, is independent of the number of rotatable bonds.

One important feature of distance geometry methods is that they are Monte Carlo methods. That is, conformational space within the distance constraints of the ensemble is randomly sampled. This has the drawback that we may have to take a large number of samples to be sure not to miss an interesting solution. Also we have to arrange the solutions into families for analysis. In contrast, a systematic search can in principle generate a complete set of distinct conformations (although, in practice, some interesting conformations may be missed if the grid is too coarse). For problems for which a complete set of solutions must be found, systematic search methods may be preferred. For problems in which a representative sample of solutions (or the fact that a solution does not exist) is sufficient, Monte Carlo methods may be applied to advantage.

Conclusion

We describe a new application of distance geometry in which two or more molecules are treated as an ensemble. With this approach we can find a common pharmacophore from a small set of biologically active molecules and generate coordinates for the set of molecules in their receptor-bound conformations such that their essential groups are superimposed. This approach has several advantages over previous methods of finding pharmacophore geometries, especially for molecules containing flexible rings.

We find only one pharmacophore geometry compatible with all four nicotinic agonists, assuming that the cationic center and a pair of atoms that form a dipole are important for activity. The pharmacophore triangle formed by the three atoms has sides 4.8, 4.0, and 1.2 Å. The pharmacophore geometry, compatible with previous models in the literature, can be reached by agonists and antagonists not in the original set used to derive the pharmacophore. We suggest that a specific arrangement of the pharmacophore triangle and the bulk of the volume of agonist molecules defines a "handedness" essential for agonist activity. By docking together the conformations of various agonists that achieve the pharmacophore geometry and that have the correct handedness, we can derive a volume occupied by agonists on the receptor.

Acknowledgment. This work was supported in part by the U.S. Army Medical Research and Development Command, Contract No. DAMD17-84-C-4111.

Registry No. (-)-Nicotine, 54-11-5; (-)-cytisine, 485-35-8; (-)-ferruginine methiodide, 85514-41-6; (-)-muscarone, 16980-76-0; *trans*-3,3′-bis O, 83800-31-1; strychnine, 57-24-9; trimethaphan, 7187-66-8; dihydro-β-erythroidine, 23255-54-1; (+)-muscarone, 4780-69-2; (+)-nicotine, 25162-00-9.

Structure–Taste Correlation of L-Aspartyl Dipeptides Using SIMCA Method

Yoshikatsu Miyashita, Yoshimasa Takahashi,† Chiyozo Takayama, Kazuo Sumi, Kazuya Nakatsuka, Takehiko Ohkubo, Hidetsugu Abe,† and Shin-ichi Sasaki*

School of Materials Science and Research Center for Chemometrics, Toyohashi University of Technology, Tempaku-cho Toyohashi 440, Japan. *Received November 26, 1984*

One of the pattern recognition techniques, the SIMCA method, has been applied to structure–taste studies on L-aspartyl dipeptides (L-Asp-NH-R). The sweet and bitter taste class models of the peptides were obtained by using five structural descriptors, such as molar refractivity, and four kinds of STERIMOL parameters. The classification rates were calculated to be 87% and 81% for sweet and bitter peptides, respectively. The SIMCA method has also suggested that two factors, shape and size, of the C-terminal amino acid moiety R in the dipeptides are extremely important to model their taste qualities.

Many attempts[1] have been carried out to develop new sweeteners. It is important to investigate the structure–taste relationships to obtain information of sweetener designing.

Recently, van der Heijden[2] investigated the quantitative structure–sweetness relationships of L-aspartyl dipeptide analogues. Iwamura has performed studies on the correlation between structure and taste potency of perillartine analogues[3] and the structure–sweetness relationships of L-aspartyl dipeptide analogues.[4]

Kier[5] examined a series of perillartines for their sweet or bitter taste by using molecular connectivity indices and discriminant analysis. We have been interested in applying the pattern recognition methods to several structure–activity problems.[6] With regard to the structure–taste relations, for example, the classification of perillartines into sweet and bitter classes was studied by the pattern recognition method.[7]

Ariyoshi[8] proposed that the structure–taste correlation of L-aspartyl dipeptide analogues depends on the common molecular features relating to the sweet peptide through the Fischer projection formula of dipeptides.

In this study an attempt has been made to classify the sweet and bitter class dipeptides by the SIMCA pattern

† Research Center for Chemometrics.

(1) Crosby, G. A.; Dubois, G. E.; Wingard, R. E. *Drug Design*; Ariens, E. J., Ed.; Academic: New York, 1979; Vol. 8, p 215.
(2) van der Heijden, A.; Brussel, L. B. P.; Peer, H. G. *Food Chem.* **1978**, *3*, 207.
(3) Iwamura, H. *J. Med. Chem.* **1980**, *23*, 308.
(4) Iwamura, H. *J. Med. Chem.* **1981**, *24*, 572.
(5) Kier, L. B. *J. Pharm. Sci.* **1980**, *69*, 416.
(6) Miyashita, Y.; Takahashi, Y.; Daiba, S.; Abe, H.; Sasaki, S. *Anal. Chim. Acta* **1982**, *143*, 35.
(7) Takahashi, Y.; Miyashita, Y.; Tanaka, Y.; Abe, H.; Sasaki, S. *J. Med. Chem.* **1982**, *25*, 1245. Takahashi, Y.; Miyashita, Y.; Tanaka, Y.; Hayasaka, H.; Abe, H.; Sasaki, S. *J. Pharm. Sci.* **1984**, *73*, 737.
(8) Ariyoshi, Y.; Yasuda, N.; Yamatani, T. *Bull. Chem. Soc. Jpn.* **1974**, *47*, 326.

0022-2623/86/1829-0906$01.50/0