New Methods in Computer-Aided Drug Design

ROBERT P. SHERIDAN and RENGACHARI VENKATARAGHAVAN*

Medical Research Division, Lederle Laboratories, American Cyanamid, Pearl River, New York 10965

Received March 4, 1987 (Revised Manuscript Received May 21, 1987)

Introduction

Pharmaceutical research has been successful in identifying therapeutic agents by using conventional screening techniques. In this approach, large numbers of randomly selected compounds, either natural products or synthesized compounds, are tested in a battery of biological assays, or "screens". The combination of random selection and the ability of many screens to quickly evaluate the biological activity of a compound provides a practical means of identifying new "leads", structural classes with potential in a specific therapeutic area. Once a lead is found, chemists synthesize variations or "analogues" on the basic structure in an attempt to increase its activity and reduce its toxicity. From a small set of active compounds with low toxicity, usually one is then tested in clinical trials. In a few fortunate cases, compounds developed in this way will reach the market as drugs. To be marketable, a compound must be novel in addition to having good activity and low toxicity. The chief merit of the screening approach is that it periodically uncovers structural classes of compounds not previously known or never before used in a particular therapeutic area. However, the process, being largely based on trial and error, requires large amounts of time and money. For every 10 000 or more compounds synthesized each year in an average pharmaceutical company, less than one makes it to the market. Any method that allows the pharmaceutical chemist to increase the likelihood of synthesizing an active analogue or to increase his ability to find, or even design, novel leads is of enormous commercial interest.

For this reason many pharmaceutical companies explore structure–activity methods. The number of such methods has greatly expanded in recent years, along with the availability of commercial software and computer graphics systems. We like to divide structure–activity methods into two categories depending on how chemical structure is represented: topological/statistical methods and geometric modeling methods.

In a "topological" approach[1-6] only the "flat" chemical structure of a molecule is taken into account. Statistical, or "pattern recognition", techniques are commonly used to find structure–activity relationships for large numbers of chemical structures represented in this way.

In "modeling" methods,[6-11] the chemist considers the properties of molecules in three dimensions. Conformational analysis, quantum mechanics, and molecular mechanics techniques are important here. Interactive molecular graphics,[7,8] which allows the chemist to manipulate molecules in three dimensions and to perceive spatial information, is essential. Structure–activity relationships derived from modeling are often expressed in terms of pictures rather than in statistical rules. Modeling methods can be further subdivided into those most suitable for comparing small molecules and those most suitable for studying the interaction of small molecules with macromolecules (usually proteins). Recently, work has begun to reconcile the results from topological and modeling efforts.[12]

Our efforts at Lederle have focused in the past 10 years on developing methods of computer-aided drug design and applying them to practical problems. This review, divided into three parts, touches on some of the methods developed recently at Lederle. The first part describes how a topological approach to molecular structure is used to enhance conventional random screening and aid in the process of drug design. The second part addresses a technique of finding common geometric features in sets of small molecules, with the goal of generating a three-dimensional model of a common receptor. The third part describes an approach to studying the interaction of small molecules with macromolecules on the basis of molecular shape.

Topological Approaches

The pharmacological activity of a molecule at its site of action is due to the spatial arrangement and electronic nature of its atoms. However, most molecules are flexible and it is often impossible or impractically time-consuming to specify which conformation(s) of the molecule are important, especially when large numbers of molecules have to be considered. It is common,

Robert P. Sheridan was born in Passaic, NJ, in 1953. He received his B.S. from Upsala College in 1975 and his Ph.D. in Biochemistry from Princeton University in 1979. He next joined the Institute for Cancer Research as a postdoctoral fellow. He served as an Instructor in the Chemistry Department of Rutgers University before joining Lederle in 1983, where he is currently a Project Leader in Molecular Modeling. Dr. Sheridan is interested in the development and application of computer-based drug design methods.

Rengachari Venkataraghavan obtained his Ph.D. degree from the Indian Institue of Science in 1963. He was then a postdoctoral fellow at the National Research Council of Canada. He joined Cornell University as a Senior Research Associate and persued his research interests in computer-aided acquisition and analysis of spectroscopic data. He moved to Lederle Labs in 1977 where he is now the director of the Biomedical Research Computing Section.

* To whom correspondence should be addressed.

(1) Martin, Y. C. In *Quantitative Drug Design*; Grunewald, G. L., Ed.; Marcel Dekker: New York, 1978; Med. Res. Ser. Vol. 8.
(2) Hansch, C.; Leo, A. *Substituent Constant for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979.
(3) Mager, P. P. *Multidimensional Pharmacochemistry: Design of Safer Drugs*; Academic: New York, 1984; Vol. 20.
(4) Stuper, A. J.; Brugger, W. E.; Jurs, P. C. *Computer Assisted Studies of Chemical Structure and Biological Function*; Wiley: New York, 1979.
(5) Clementi, S. In *Drug Design: Fact or Fancy?*; Jolles, G., Wooldridge, K. R. H., Eds.; Academic: New York, 1984; pp 73–92.
(6) Franke, R. *Theoretical Drug Design Methods*; Elsevier: New York, 1984; Vol. 7.
(7) Langridge, R.; Ferrin, T. E.; Kuntz, I. D.; Connolly, M. L. *Science* 1981, *211*, 661–666.
(8) Meyer, E. F., Jr. In *Drug Design*; Academic: New York, 1980; Vol. IX, pp 268–299.
(9) Cohen, N. C. In *Advances in Drug Research*; Testa, B., Ed.; Academic: New York, 1985; Vol. 14, pp 42–146.
(10) Hopfinger, A. J. *J. Med. Chem.* 1985, *28*, 1133–1139.
(11) Marshall, G. R. *Annu. Rev. Pharmacol. Toxicol.* 1987, *27*, 193–213.
(12) Hansch, C.; Klein, T. E. *Acc. Chem. Res.* 1986, *19*, 392–400.

0001-4842/87/0120-0322$01.50/0

therefore, for structure–activity methods to consider only the "topology" of a molecule, that is, those aspects that are contained in a two-dimensional structural diagram. A great deal of the three-dimensional information for a molecule is presumed to be implicit in the topological description despite the simplifications involved. With a topological approach we can take advantage of the computerized "connection table" databases of many thousands of compounds commonly maintained by chemical and pharmaceutical companies.

There are two aspects of these methods that are important: the descriptors and the technique to relate these descriptors to activity. Molecular descriptors are numerical values which represent selected features of a molecule. There have been many descriptors proposed in the literature: physical properties (hydrophobicity, electron-donating ability) of substituents, molecular shape, the presence of substructures, indexes of molecular connectivity, etc. We refer the reader to reviews in this area.[1-6]

Each molecule can be represented as a location in a high-dimensional space, each dimension corresponding to a descriptor. The problem of relating activity to structure becomes the problem of relating activity to location in the space. Again, a variety of methods have been described for dealing with the problem: nearest-neighbor analysis, partial-least-squares analysis, principle component analysis, linear regression, discriminant-plane analysis, etc. These have also been reviewed.[1-6]

Atom Pairs and Topological Torsions. We have presented[13,14] two new descriptors, the "atom pair" (AP) and "topological torsion" (TT), and some methods of relating the descriptors to biological activity which we find extremely useful in the industrial environment. Both AP's and TT's are related to substructure descriptors proposed by others, and we refer the reader to the citations in ref 13 and 14.

An AP is a substructure composed of two non-hydrogen atoms i and j and an interatomic separation

$$\langle\text{atom type } i\rangle\text{–}\langle\text{separation}\rangle\text{–}\langle\text{atom type } j\rangle$$

where ⟨atom type i⟩ contains information about the element type, the number of non-hydrogen neighbors, and the number of π electrons; ⟨separation⟩ is the number of bonds in the shortest bond-by-bond path that connects atoms i and j. A structure with N non-hydrogen atoms is represented by the aggregate of $N(N-1)/2$ AP's.

A TT is a substructure of four atoms i, j, k, and l which are directly bonded:

$$\langle\text{atom type } i\rangle\text{–}\langle\text{atom type } j\rangle\text{–}\langle\text{atom type } k\rangle\text{–}\langle\text{atom type } l\rangle$$

A structure with M sets of four consecutively bonded atoms will be described by M TT's.

AP's and TT's capture complementary aspects of chemical structures, AP's taking into account long-range properties and TT's taking into account short-range substructures. The AP and TT descriptors were developed to be general enough that we can generate structure–activity relationships for sets of compounds

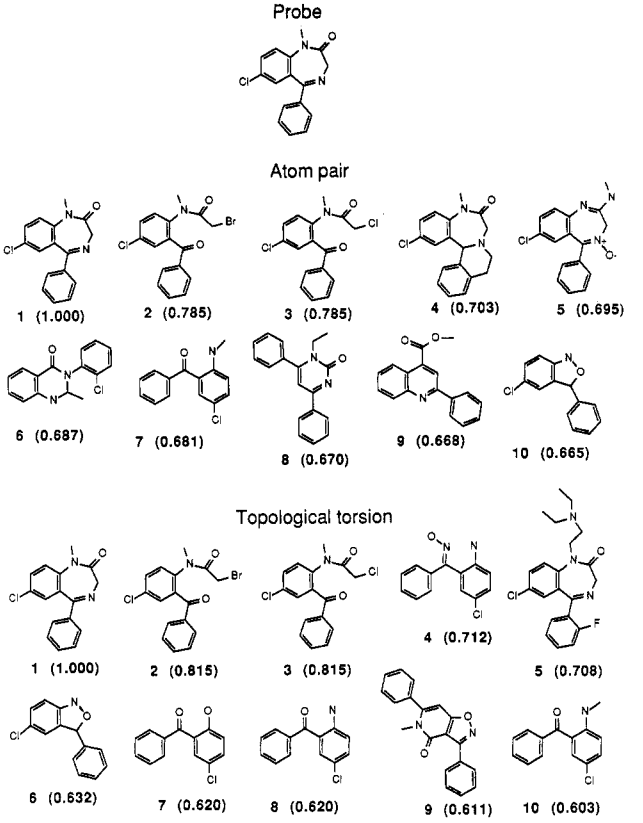


Figure 1. Chemical structures in the *Fine Chemicals Directory*[16] which are the most similar to the probe diazepam on the basis of atom pair and topological torsion descriptors. The similarity values are shown in parentheses.

with diverse chemical structures but specific enough in the aggregate to discriminate between closely related isomers. It is noteworthy that the descriptors are unbiased by any prejudice about what part of a molecule is the "core" and which the "substituent" or about the type of substructures that "should" be important.

Similarity Probe. Given these descriptors, it is possible to define the "similarity" between two structures as the number of descriptors they have in common. Similarity ranges from 1.0 (complete identity) to 0.0 (nothing in common). We have found this similarity criterion to be an extremely useful tool in enhancing the productivity of screening. Often, for instance, one would like to screen as diverse a set of compounds as possible. For this application the similarity criterion can be used in a converse way: Given a set of compounds, one can choose a subset such that each compound is *dissimilar* to every other.

A more commonly used application is the "similarity probe". This is used to select molecules from a large database which are the most similar to a "probe" structure and thus most likely to show similar biological activities. We commonly use as probes leads from the literature or active compounds found by previous screening effort.

As an example of a probe molecule consider diazepam. Diazepam belongs to the class of benzodiazepines,[15] compounds which bind to specific receptors on the GABA-R chloride channel complex and which show a variety of central nervous system effects, acting as

(13) Carhart, R. E.; Smith, D. H.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1985**, *25*, 64–73.

(14) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 82–85.

(15) Haefely, W.; Kyburz, E.; Gerecke, M.; Mohler, H. In *Advances in Drug Research*; Testa, B., Ed.; Academic: New York, 1985; Vol. 14, pp 166–322.

anxiolytics, anticonvulsants, and hypnotics. Ten structures from the *Fine Chemicals Directory*[16] which are most similar to the probe diazepam are shown in Figure 1. It is obvious that the similarity of these structures to the probe is at a more general level than can be captured by a substructure search. Also, it should be noted that AP's and TT's capture different aspects of structures; while some structures are the same in the AP and TT list, expecially toward the highest similarity values, there are significant differences between the lists.

We often find that structures selected by the similarity probe have biological activities related to that of the probe several times more often than expected by chance. This is true even for compounds which are not simply analogues of the probe. Mecloqualone, for instance, which appears as **6** in the AP list in Figure 1, is a non-benzodiazepine with sedative and hypnotic activity.

Trend Vector Analysis. When large numbers of molecules are available for which the biological response has been measured, we can begin to ask about which structural features distinguish the more active from the less active molecules. Imagine that we have a "training set" of structures, each with an activity value associated with it. Each molecule can be represented as a location in a high-dimensional space, each dimension associated with distinct *type* of descriptor (AP or TT) in the set. A "trend vector" **T**, pointing from the inactive molecules toward the active molecules, can be calculated by a formula analogous to that for calculating a dipole moment, with "activity" replacing charge. **T** summarizes the activity data in a chemically meaningful way: The vector component T_k is more positive as descriptor k is more closely associated with active molecules and more negative as it is more closely associated with inactive molecules. Calculating a trend vector is rapid, and it involves no adjustable parameters, thus avoiding the mathematical difficulties that arise from an excess of descriptors.

The length of **T** can be used to decide whether the structure–activity relationship described by the vector is statistically significant. The vector calculated from the "real" activity data is thought to contain meaningful information if it is significantly longer than vectors calculated from spurious data created by randomly reassigning the original activities to the wrong structures.

The vector representation is also useful in that the angle between two vectors can be taken as a measure of their "parallelism". Often by comparing trend vectors in this way, we detect a previously unsuspected similarity in the structure–activity relationships of two distinct biological activities.

Given a significant trend vector, we can calculate the predicted activity (or rank) of any arbitrary structure by calculating the projection of the structure on **T**. We insist that at least 95% of the descriptor types in the structure be "recognized" by the trend vector (that is, be present in the training set from which the vector was calculated) before rank can be assigned, since the contribution of the unrecognized descriptors to the activity is undefined. Although the correlation between the observed activity and rank for molecules within a typical training set appears quite modest when compared to those correlations generated by fitting methods which use adjustable parameters, the correlation for structures not already "fit" is sometimes comparable.[13] Whether AP or TT descriptors are better in capturing a structure–activity relationship appears to depend on the training set.

We have developed two applications of trend vector based predictions. The first is a graphics-driven program with which a chemist can interactively draw a chemical structure and ask for the predicted activity. Given immediate feedback, the chemist can quickly design a molecule with a higher rank than the molecule he started with. The program also can make suggestions about atom replacements and additions that would tend to increase the rank. In the second application, one can scan a large database of compounds, calculate the rank for each, and save the structures with the highest ranks. These are the structures most likely to show activity.

The clearest way to demonstrate the utility of the second application is by a retrospective experiment. A diverse set of about 4500 Cyanamid proprietary compounds were tested in the laboratory[17] for binding to the insect nicotinic acetylcholine receptor. This receptor, present in both vertebrates and invertebrates, is a transmembrane glycoprotein located in the postsynaptic membrane of cholinergic systems. The purpose of this screen was to identify compounds which could, like many plant products such as nicotine, act as insecticides by disrupting the function of acetylcholine as a neurotransmitter. The compounds were called "active" or "inactive" depending on their ability to displace a radiolabeled receptor–specific protein neurotoxin from a receptor preparation.[18] An active compound could be an agonist (which induces the same effect as acetylcholine) or antagonist (which blocks the effect). This large set was randomly divided into two smaller sets of approximately equal size. A trend vector (in this case the topological torsion descriptor proved better) was constructed by using one set as the training set. This vector was used to calculate the relative rank of compounds in the remaining "test" set.[14] In a screening program, one is interested in finding a sufficiently large set of chemically diverse actives by screening the smallest number of compounds. By random screening, in order to find, say, 50% of the actives in the test set, we would expect to have to screen 50% of the set. However, if we screen the test set in the order of decreasing theoretical rank, we find about 50% of the actives in the first 10% of the compounds screened. This represents a 5-fold increase in "hit rate" (in this case from about 1% to about 5%) and a corresponding 5-fold saving in screening effort.

By itself, a 5-fold increase in hit rate is not necessarily impressive; chemists and biologists often achieve higher hit rates when selecting analogues of known actives. But typically the active structures found from trend vector ranking are more diverse, and thus the probability is increased for discovering an unexpected new

(16) *Fine Chemicals Directory Handbook*; Fraser Williams (Scientific Systems): London, 1983–1985. Connection tables distributed by Molecular Design Ltd., Hayward, CA.

(17) Kukel, C.; Chalmers, A.; Jennings, K., personal communication.

(18) Jones, S. W.; Sudershan, P.; O'Brien, R. D. *J. Neurochem.* **1981**, *36*, 447–453.

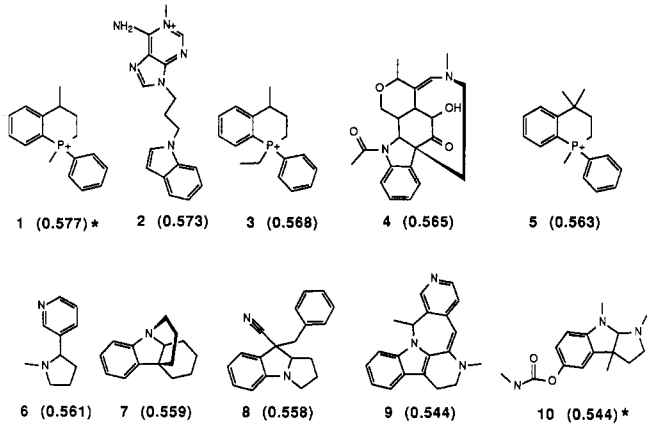


Figure 2. Structures in the Cambridge Crystallographic Database[19] most likely to bind to the insect nicotinic receptor. The ranks (relative predicted activities) of these structures as calculated from the topological torsion trend vector are shown in parentheses. Structures marked with an asterisk happened to be present in the training set of proprietary compounds from which the trend vector was calculated.

class of active compounds. To illustate this, we have applied the trend vector described above to the connection tables of the Cambridge Crystallographic Database.[19] The structures most likely to bind to the nicotinic receptor are shown in Figure 2. Structures **1** and **10** (the alkaloid eserine) happen to be also included in the training set of proprietary compounds; they are active in the screen. Structures **3** and **5** resemble structure **1**. The known agonist nicotine appears as structure **6**. The remaining structures are not analogues of any structure in the training set. Structure **9** is the plant alkaloid decussine, which shows muscle-relaxing activity[20] similar to that of strychnine, a nicotinic antagonist. The remaining structures, especially **4** (the alkaloid strychnobrasiline), are also likely to show related activity.

Active Analogue Approach: Ensemble Distance Geometry

The "lock and key" analogy has been a useful idea in pharmacology. In that analogy ligands, e.g., drug molecules (keys), exert their effects by binding to receptors (locks). Whereas real keys have only physical shape as the important property, drug molecules have a number of physicochemical properties. The "pharmacophore hypothesis" is the simplifying assumption that, to be activated, the receptor must recognize a "pharmacophore", a set of essential chemical groups common to all active molecules. Since the structure of most interesting receptors is not known, one must deduce the corresponding "pharmacophore model" from sets of active and closely related inactive molecules. Such a model has two important features: the pharmacophore geometry, i.e., the unique three-dimensional arrangement of the essential groups (analogous to the pattern shared by all keys), and the binding site volume available for occupancy by ligands (analogous to the keyhole).

Deducing the pharmacophore geometry is complicated by the fact that most interesting ligand molecules are conformationally flexible. The first step in deducing the geometry is to choose the essential groups in each of a set of active molecules such that the equivalent of each group (which can be an atom, the center of a ring, etc.) appears in each molecule. The problem of finding the pharmacophore geometry (discussed in ref 21 and 22) can then be stated: Find low-energy conformation(s) for each molecule such that there is a arrangement of the groups common to all molecules. If there is no such arrangement, the initial choice of the groups must be revised. If there is more than one arrangement, more constraints must be applied.

Once a unique pharmacophore geometry has been deduced, one has the further task of deciding for each molecule which of the low-energy conformations that can attain the pharmacophore geometry is the "receptor-bound" conformation. The union of the volumes of all receptor-bound conformations, docked together such that the equivalent groups are superimposed, provides a minimum volume for the binding site. A molecule that can attain the pharmacophore geometry may still be inactive if it extends outside the binding site volume.

Two approaches have been previously explored for finding pharmacophore geometries from sets of flexible molecules. The first approach was developed by Marshall and co-workers.[23] Other workers (for example Schulman et al.[24]) have used similar concepts with slight variations. A second approach came from Crippen and co-workers.[25] Recently, we described[26] an "ensemble" extension of distance geometry that provides a very different approach. Distance geometry is a technique that addresses the problem: given a set of N points (usually atoms in chemical applications) and a matrix of upper and lower bounds for the distances between them, generate three-dimensional coordinates such that the distance bounds are satisfied. An algorithm for accomplishing this was developed by Crippen and coworkers and has been described in detail.[27,28] Structures generated by this algorithm represent Monte Carlo samplings of conformation space within the constraints of the distance bounds. The key to our extension is to include the atoms of all the molecules to be considered in one large distance bounds matrix. The distance bounds are set from the covalent structure of the molecule, from requirement that there be no intramolecular hard-sphere band contacts, and from the requirement that equivalent groups from different molecules be superimposed. The final structures generated by the entire algorithm consist of several molecules, each in a low-energy conformation, superimposed

(19) Allen, F. H.; Bellard, S.; Brice, M. D.; Cartwright, B. A.; Doubleday, A.; Higgs, H.; Hummelink, T.; Hummelink-Peters, B. G.; Kennard, O.; Motherwell, W. D. S.; Rodgers, J. R.; Watson, D. G. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1979**, *B35*, 2331–2339.
(20) Rolfsen, W. N. A.; Olaniyi, A. A.; Sandberg, F.; Kvick, A. H. *Acta Pharm. Suec.* **1980**, *17*, 105–111.
(21) Motoc, I.; Dammkoehler, R. A.; Mayer, D.; Labanowksi, J. *Quant. Struct.-Act. Relat. Pharmacol., Chem. Biol.* **1986**, *5*, 99–105.
(22) Marshall, G. R. *Ann. N.Y. Acad. Sci.* **1985**, *439*, 162–169.
(23) 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.; American Chemical Society: Washington, DC, 1979; ACS Symp. Ser. No. 112, pp 205–226.
(24) Schulman, J. M.; Sabio, M. L.; Disch, R. L. *J. Med. Chem.* **1983**, *26*, 817–823.
(25) Crippen, G M. *J. Med. Chem.* **1979**, *22*, 988–997.
(26) Sheridan, R. P.; Nilakantan, R.; Dixon, J. S.; Venkataraghavan, R. *J. Med. Chem.* **1986**, *29*, 899–906.
(27) Crippen, G. M. *Distance Geometry and Conformational Calculations*; Research Studies Press: New York, 1981.
(28) Havel, T. F.; Kuntz, I. D.; Crippen, G M. *Bull. Math. Biol.* **1983**, *45*, 665–720.

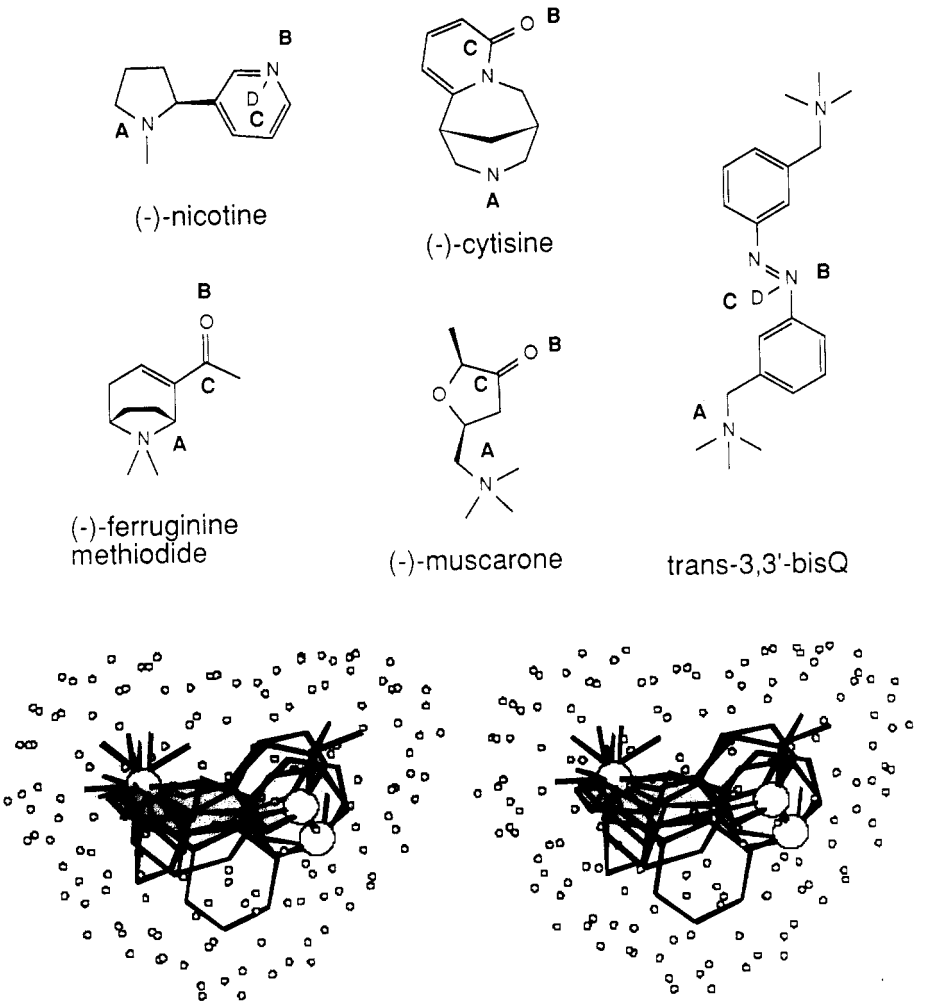


Figure 3. (Top) A set of nicotinic agonists. (−)-Nicotine, (−)-cytisine, (−)-ferruginine methiodide, and (−)-muscarone were used to generate the pharmacophore geometry by using the ensemble approach. Trans-3,3′-bisQ was checked for consistency with the pharmacophore model. Atoms labeled D are "dummy" atoms along the angle bisector and 1.2 Å from the atom to which they are attached. In each molecule the bold letters A, B, and C indicate the atoms identified with essential groups in the pharmacophore. (Bottom) A stereopicture of the combined volume of the agonists: (+)- and (−)-nicotine, (+)- and (−)-muscarone, (−)-cytisine, (−)-ferruginine methiodide, and *trans*-3,3′-bisQ. The agonists, in their likely receptor-bound conformations, were docked so that the pharmacophore atoms were best superimposed. The symmetrical structure of *trans*-3,3′-bisQ suggests that the receptor can accommodate two pharmacophores. Most agonists occupy only half of such a site. We consider only that half-site and truncate *trans*-3,3′-bisQ accordingly. The mean position of each type of pharmacophore atom is shown by a large circle: A, B, and C are arranged counterclockwise with A at the left. Small dots indicate the solvent-accessible surface of the combined volume of the agonists.

at the essential groups. The pharmacophore geometry is found by inspecting these final structures. The ensemble approach has some advantages over previous methods in that it is possible to include steric interactions and common chiral constraints between molecules, in that flexible rings are easily handled, and in that the computation time is independent of the number of rotatable bonds.

One receptor where we applied the ensemble approach with great success is the vertebrate nicotinic acetylcholine receptor (reviewed by Changeux et al.[29]). Nicotinic agonists induce an open-channel form of the receptor. Antagonists bind to the receptor but do not open the channel. To elucidate the pharmacophore for this receptor, we studied an ensemble of four semirigid agonists: (−)-nicotine, (−)-cytisine, (−)-ferruginine methiodide, and (−)-muscarone. Figure 3, top, shows the assignment of specific atoms as essential groups A, B, and C to be superimposed. A is a cationic center, B is an electronegative atom capable of accepting a hydrogen bond, and C helps define the direction of hydrogen bonding around B. We generated several sets of superimposed structures and found that there was only one possible pharmacophore geometry within the error of superposition: a triangle with sides 4.8 (A–B), 4.0 (A–C), and 1.2 Å (B–C). This geometry is consistent with the earlier model of Beers and Reich.[30] We found that the pharmacophore geometry could be also attained by the antagonists strychnine, trimethaphan, dihydro-β-erythroidine, and the agonist *trans*-3,3′-bis-[(trimethylammonio)methyl]azobenzene (*trans*-3,3′-bisQ).

For (−)-nicotine, (−)-ferruginine methiodide, (−)-muscarone, and *trans*-3,3′-bisQ there is more than one low-energy conformation that can attain the pharmacophore geometry. However, if we assume that these three molecules should fit into approximately the same volume as (−)-cytisine, for which there is only one such conformation, we are able to decide on a unique re-

(29) Changeux, J.-P.; Devillers-Thiery, A.; Chemouilli, P. *Science* **1984**, *225*, 1335–1345.

(30) Beers, W. H.; Reich, E. *Nature (London)* **1970**, *228*, 917–922.

ceptor-bound conformation for each. This is also true for (+)-nicotine and (+)-muscarone, which are also agonists. Once all the agonists are docked together, we get an idea of the volume that agonists can occupy on the receptor. This volume is shown in Figure 3 (below). Given the pharmacophore and the allowed volume, one may start to design new agonists to fit the model.

Applications of Shape Matching

One important aspect of ligand–receptor interactions is thought to be the complementarity in their shapes. Methods have been suggested[31,32] for docking two macromolecules by shape. Kuntz et al.[33] proposed a shape-matching method for docking rigid small molecules onto a receptor of known structure. In collaboration with his group[34,35] we have made several extensions of this method.

The shape-matching method requires shape representations of the receptor binding site and of the ligand. To characterize the shape of the receptor, we generate[33] a set of spheres outside the solvent-accessible surface[36] which fill all the pockets and grooves on the receptor. (In the lock and key analogy, the atoms of the receptor represent the lock and the receptor spheres approximate the shape of the keyhole.) Over a typical macromolecule there are several distinct sets of overlapping spheres, representing surface invaginations of various sizes, but the largest set usually corresponds to the observed binding site. The shape of a ligand can be most simply defined by the location of its non-hydrogen atoms. Both the receptor and ligand are treated as rigid.

The geometrically possible ways to orient a ligand in the binding site are found by a method of systematic distance matching[33] that maps a subset of ligand atoms into a subset of pockets, each pocket defined by a sphere center, that can receive them. Typically, there may be a few hundred orientations for a given ligand. These orientations can then be processed further depending on the application.

In one application,[34] we automate the search for geometrically plausible ways to fit a known ligand onto a receptor of known three-dimensional structure. Our approach assumes that shape complementarity is useful criterion for finding the right binding mode.

Most ligands are flexible molecules. The shape-matching algorithm can be used if the flexible ligand is approximated as a small number of large rigid fragments. Orientations for each fragment in the binding site are generated as described above. For each fragment, we eliminate those orientations that result in a significant overlap between the fragment and the receptor and eliminate those orientations that are nearly coincident with another orientation. To recreate the ligand, we systematically pair the orientations of the two fragments and save the orientations in which specified atoms from each fragment can be rejoined as they were joined in the intact ligand. A set of orientations chosen in this way constitutes a "binding mode" for the ligand. We then divide the binding modes into "families" and energy minimize each family in the presence of the receptor using AMBER[37] (assisted model building and energy refinement). We show as an example the case of methotrexate docked to dihydrofolate reductase (DHFR). DHFR, a key enzyme in the metabolism of all living organisms, catalyzes the reduction of dihydrofolate to tetrahydrofolate. DHFR inhibitors act as antitumor agents (e.g., methotrexate) and antibacterial agents (e.g., trimethoprim). The crystal structures of several DHFR–inhibitor complexes are known.[38-40] For the purposes of this application, the most interesting inhibitor is methotrexate, an analogue of the substrate folate. The bound conformation of methotrexate shows strong shape complementarity to the active site of DHFR.

Spheres were generated to define the shape of the methotrexate binding site on DHFR. Methotrexate was approximated as two large fragments called fragment 1 and fragment 2. The receptor and fragments are shown in Figure 4, top and middle, respectively. Orientations were generated for each fragment and were processed as described. We found four families of binding modes. Two of these, in which the "pteridine" portion of the ligand pointed outward, were of high energy. The two low-energy families are shown in Figure 4 (below). One looked very much like the crystallographically observed binding mode of methotrexate, and one was in an "inverted" mode very much like that thought to be assumed by folate.[38]

We are currently developing shape matching as a tool for designing novel ligands to bind to a given receptor.[35,42] In these applications, one scans a databse of small molecules to find those molecules which have the best shape complementarity to the receptor binding site and then modifies the molecules to complement the receptor in chemical properties. The shape of the receptor may be known from X-ray crystallography or be derived from pharmacophore models of the type described in the previous section.

Conclusion

The goal of "designing" drugs, in the sense that one designs an airplane or a bridge, has not yet been realized since the pharmacological equivalent of the principles of aerodynamics or structural engineering remains largely unknown. The purpose of current methods of computer-aided drug design, then, is to generate statistical or graphical models consistent with experiment and then suggest, on the basis of the models, which new experiments will be the most fruitful. Until recently, practitioners of drug design methods were limited by the availability of appropriate software and computer graphics systems. Now that these barriers have been at least partly removed, the limit to generating a useful model is the lack of experimental data. We anticipate

(31) Wodak, S. J.; Janin, J. *J. Mol. Biol.* **1978**, *124*, 323–342.
(32) Santavy, M.; Kypr, J. *J. Mol. Graphics* **1984**, *2*, 47–49.
(33) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. *J. Mol. Biol.* **1982**, *161*, 269–288.
(34) DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. *J. Med. Chem.* **1986**, *29*, 2149–2153.
(35) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R., submitted to *J. Med. Chem.*
(36) Connolly, M. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.
(37) Weiner, P. K.; Kollman, P. A. *J. Comput. Chem.* **1981**, *2*, 287–303.
(38) Bolin, J. T.; Filman, D. J.; Matthews, D. A.; Hamlin, R. C.; Kraut, J. *J. Biol. Chem.* **1982**, *257*, 13650–13662.
(39) Blakley, R. L. In *Folates and Pterins I*; Blakley, R. L., Benkovic, S. J., Eds.; Wiley: New York, 1984; pp 191–253.
(40) Freisheim, J. H.; Matthews, D. A. In *Folate Antagonists as Therapeutic Agents I*; Sirotnak, F. M., Burchall, J. J., Ensminger, W. D., Montgomery, J. A., Eds.; Academic: New York, 1984; pp 69–131.
(41) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
(42) Sheridan, R. P.; Venkataraghavan, R., unpublished data.

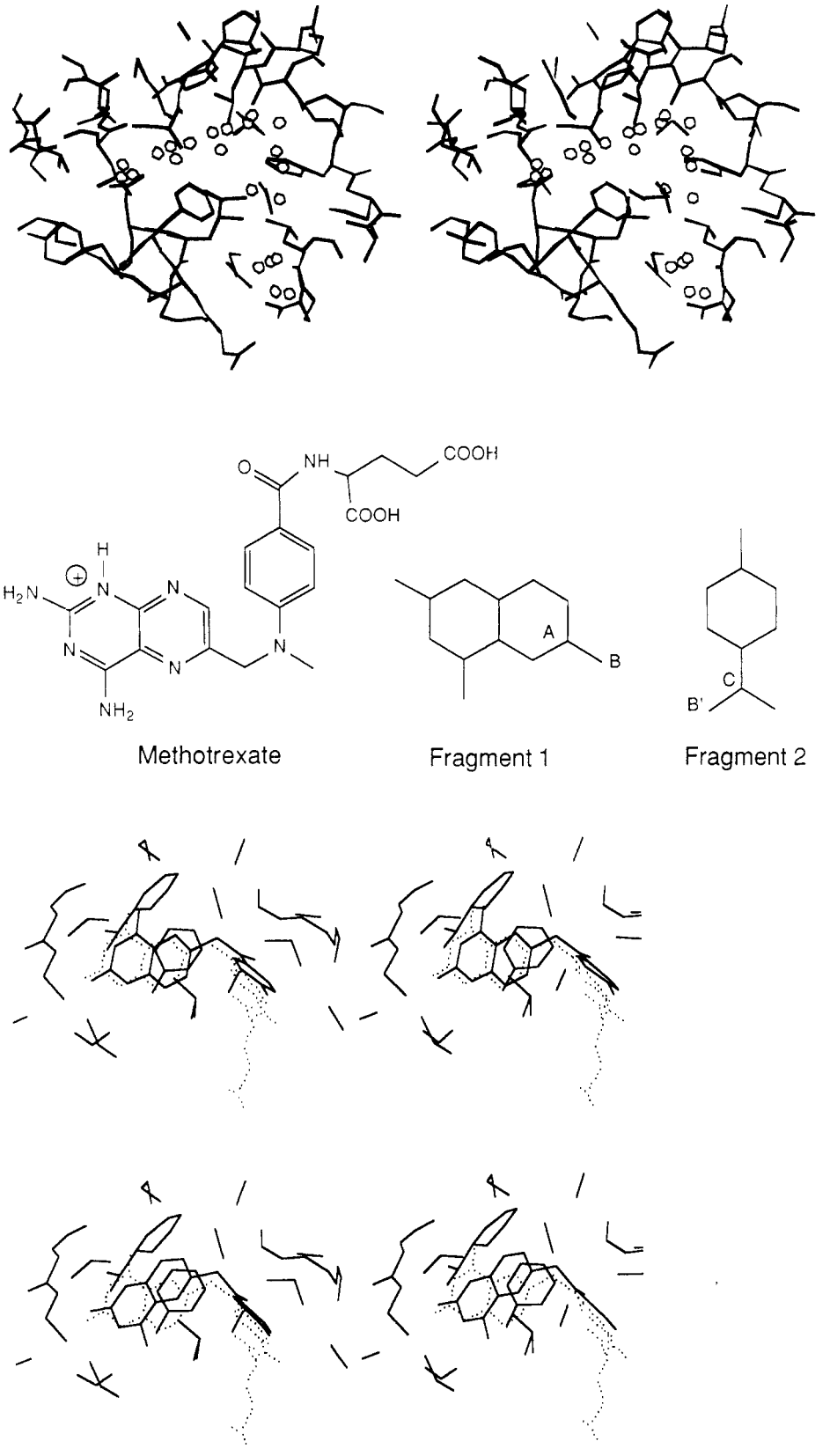


Figure 4. (Top) Representation of the shape of the active site from *L. casei* dihydrofolate reductase (Brookhaven Protein Data Bank[41] coordinate set 3DFR). The 22 spheres representing the site are shown as large circles. In the crystallographically observed binding mode of the inhibitor methotrexate, the pteridine ring of methotrexate fits toward the left, the benzoate portion fits in the corner, and the amino acid portion points down. (Middle) The shape of methotrexate is approximated as two large rigid framents. In selecting orientations of each fragment to be joined, we require that B and B′ be nearby (they are the same atom in the intact ligand) and that A and C be approximately the same distance as they would be in the angle A–B–C. (Bottom) Stereopictures of two low-energy binding modes for methotrexate (only those atoms in the fragments are included) on dihydrofolate reductase. The crystallographically observed binding mode for the entire methotrexate molecule (dotted) is shown for comparison. Explicit hydrogens were included on heteroatoms for the energy minimization, but these are omitted here for clarity. The upper mode closely resembles the observed binding mode, while the lower mode resembles the mode thought to be adopted by the substrate folate.

that two current trends will correct this situation. The first trend is the increasing availability of crystal structures of macromolcules which are of pharmacological interest and the explosion of molecular biology and protein engineering techniques which help us understand drug–receptor interactions at the molecular level. The second is the increasing availability of modeling tools to experimentalists, a trend which will no doubt result in a more timely testing of computer-generated models.

The authors gratefully acknowledge the contribution of their present and former colleagues Dr. R. Carhart, Dr. D. Smith, Dr. J. S. Dixon, Dr. N. Bauman, Dr. R. Nilakantan, Dr. V. N. Balaji, Dr. C. Peishoff, and Dr. D. Dunn. The enthusiasm of Dr. G. Sutherland and Dr. A. Oronsky for novel techniques in drug design is greatly appreciated.

Laser Photolysis of Silylene Precursors

PETER P. GASPAR,* DEWEY HOLTEN, and STANISLAW KONIECZNY

Department of Chemistry, Washington University, St. Louis, Missouri 63130

JOYCE Y. COREY

Department of Chemistry, University of Missouri, St. Louis, Missouri 63121

Received November 14, 1986 (Revised Manuscript Received April 14, 1987)

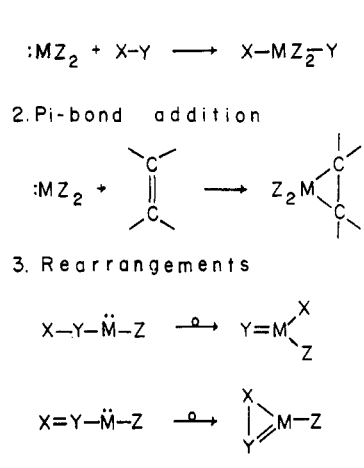
Introduction

The reactions of compounds containing divalent silicon atoms, generally called silylenes,[1] were of particular interest to us from the beginning of our efforts to employ the mechanistic ideas of organic chemistry beyond the first row of the periodic table. The reactions of divalent carbon compounds—carbenes—are so distinctive[2] that it seemed possible to address the differences between reactions of silylenes and carbenes even while the mechanisms of carbene reactions were being elucidated.

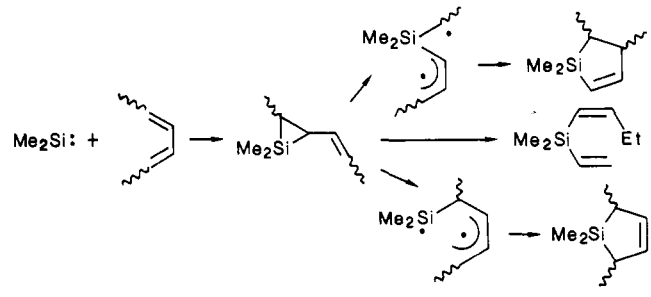
It has been found that the most important types of transformations of carbenes are shared by silylenes,[3] as shown in Scheme I. Since these reactions are both unusual and useful, we were curious about exactly how they occur, and we wanted detailed mechanistic knowledge in order to exercise greater control over them.

Our knowledge of silylene reaction mechanisms is rather rudimentary.[3] Almost every known silylene has a singlet electronic ground state, so that only singlet silylene chemistry is thus far known. Silylenes have been found to readily insert into Si–H, Ge–H, and O–H bonds, and C–H insertion is known as an intramolecular

Scheme I.
Comparable Reactions of Carbenes and Silylenes



Scheme II



Peter P. Gaspar was born in Brussels, Belgium. His undergraduate education was received at Caltech. His Ph.D. (1961) in organic chemistry at Yale, under William Doering, was followed by a NATO Fellowship at Heidelberg with Georg Wittig and further postdoctoral work at Caltech under George Hammond. A member of the faculty at Washington University, St. Louis, since 1963 he has been a visiting professor at Princeton (twice) and in Lisbon where he was also a Fulbright Lecturer. With interests in reaction mechanisms of a variety of short-lived electron-deficient species, his work in organosilicon chemistry led to the Frederic Stanley Kipping Award in 1986.

Dewey Holten is Associate Professor of Chemistry at Washington University, St. Louis. He received his B.A. from Washington University (1973), his Ph.D. in physical chemistry under the direction of Martin Gouterman from University of Washington, Seattle (1976), and did postdoctoral work at Washington State University. His research is in picosecond and slower scale laser spectroscopy, primary events in photosynthesis and model systems, and transition metal–porphyrin photophysics and photochemistry.

Stanislaw Konieczny is an Associate in Teaching and Research in the Institute of Inorganic Chemistry and Technology of the Technical University, Gdansk, Poland, where he completed his Ph.D. in 1978. His research interests cover a wide range of silicon chemistry.

Joyce Y. Corey is Professor of Chemistry at the University of Missouri—St. Louis. An undergraduate at the University of North Dakota, she received her Ph.D. in inorganic chemistry from the University of Wisconsin in 1964, having worked with Robert West. Her research is in the synthesis and characterization of organometallic compounds containing elements from groups III and IV.

(1) Gaspar, P. P.; Pate, B. D.; Eckelman, W. *J. Am. Chem. Soc.* **1966**, *88*, 3878–9.
(2) Kirmse, W. *Carbene Chemistry*, 2nd ed.; Academic: New York, 1971.
(3) Gaspar, P. P. *React. Intermed. (Wiley)* **1978**, *1*, 229–277; **1981**, *2*, 335–385; **1985**, *3*, 333–427.

0001-4842/87/0120-0329$01.50/0