Journal of Computer-Aided Molecular Design, 11 (1997) 243–255.

J-CAMD 397

A Monte Carlo method for finding important ligand fragments from receptor data

Stan Burt[a], Charles Hutchins[b] and Peter J. Zielinski[c,*]

[a]*National Cancer Institute, Frederick, MD 21702, U.S.A.*
[b]*Abbott Laboratories, D46Y AP10, and* [c]*D9P9 AP32, Abbott Park, IL 60064-3500, U.S.A.*

Received 1 May 1996
Accepted 26 February 1997

Keywords: Simulated annealing; Rational ligand design; Ligand design; Structure-based ligand design

Summary

A simulated annealing method for finding important ligand fragments is described. At a given temperature, ligand fragments are randomly selected and randomly placed within the given receptor cavity, often replacing or forming bonds with existing ligand fragments. For each new ligand fragment combination, the bonded, nonbonded, polarization and solvation energies of the new ligand–receptor system are compared to the previous configuration. Acceptance or rejection of the new system is decided using the Boltzmann distribution $e^{-E/kT}$, where E is the energy difference between the old and new systems, k is the Boltzmann constant and T is the temperature. Thus, energetically unfavorable fragment switches are sometimes accepted, sacrificing immediate energy gains in the interest of finding a system with minimum energy. By lowering the temperature, the rate of unfavorable switches decreases and energetically favorable combinations become more difficult to change. The process is terminated when the frequency of switches becomes too small. As a test, the method predicted positions and types of important ligand fragments for neuraminidase that were in accord with the known ligand, sialic acid.

Introduction

The extraordinary structural data produced by recent refinements in structure elucidation techniques, such as X-ray crystallography and NMR, have inspired the creation of innovative de novo ligand design methods to exploit these newly available structures. Software that allows the chemist to visualize, interpret and quantify stereochemical relationships [1] has improved the odds for successful ligand development. Computer software for structure-based ligand design has found wide use in the pharmaceutical industry and the predictive accuracy of this software is improving [2]. Von Itzstein et al. [3] have recently reported on the benefits of computing tools in designing potent inhibitors of influenza virus sialidase. Eventually, ligand design software may predict the structure of the best ligands without the need to synthesize and test less effective intermediates [2]. However, there are presently few examples of useful ligands that have been indisputably designed by computer. The science of ligand design is far from mature and remains a field open to new ideas.

Although the distinction is sometimes clouded, most computational schemes for ligand design may be loosely classified into two broad categories according to what initial data are essential to the scheme, as explained below. Each of these schemes, regardless of category, represents a point of view from which one hopes to design ligands for a specific receptor.

The first class consists of a diverse group of methods that require both known ligand and known receptor information. DOCK [4–7] is one of the oldest and most familiar techniques in this class. The DOCK approach is based upon the belief that steric complementarity is a major factor in molecular recognition [8]. A candidate ligand molecule is selected from a database of structures and oriented with respect to the chosen receptor pocket until the optimal steric fit is obtained. Early versions used only geometric fit for evaluating ligands, ignoring all other chemical information [9,10]. Electrostatic and van der Waals energy estimates have more recently been incorporated into these matching procedures [11]. Newer versions also support some ligand and receptor flexibility

*To whom correspondence should be addressed.

in order to accommodate the docking process [12–14]. By repeatedly testing different ligands taken from the database, ligands having the most favorable binding energies may be selected.

Monte Carlo, or Metropolis, algorithms have more recently been developed to improve docking operations. In the method of Goodsell and Olson [15], a random walk of a trial ligand in a rigid receptor is made with random displacements in the position and torsional angles of the ligand. The effect of these displacements on the interaction energy between the trial and target molecules is then computed. Displacements that cause a decrease in energy are accepted, while those that do not decrease energy are accepted with probability

$$P(E) = e^{-E/kT}$$

where E is the energy difference, k the Boltzmann constant and T the temperature. In the multiple-start Monte Carlo docking method [16], different configurations of a randomly positioned rigid molecular fragment are repeatedly docked with a rigid receptor. Each docking attempt is then scored using computationally inexpensive energy potentials to estimate the interaction energy. These scores suggest where fragments should be placed to enhance binding. The ultimate goal of this approach is to develop a library of independent rigid probes, or fragments, that can be assembled to form a complete ligand.

Statistical methods have been applied for the purpose of extracting common surface features from regions of dissimilar ligands having a similar arrangement of atoms but very different bonding [8]. The accessible surface of each ligand is fitted into a mold of the accessible surface structure of their common binding site and oriented until atoms on the molecular surfaces that are important to binding correspond. A correspondence between the positions of the ligand molecules can then be established, the ligands superimposed and a statistical comparison made between their common accessible surface and any energetic parameter mapped onto the surface.

The second class of methods requires only receptor information for generating ligands or ligand descriptors. The results of these methods range from energy contours to complete ligands. Contouring schemes generally require only a modest library of representative energy probes for generating contours, but do not construct ligands. The value of these schemes lies in their ability to suggest favorable locations for ligand fragments. The most popular contouring procedure is GRID [17–20], a program that produces nonbonded and hydrogen bonding energy contours using ligand probes consisting of a heavy atom united with any attached hydrogen atoms. The program places probes within the protein receptor pocket and computes the nonbonded interaction energy between the probe and receptor as the probe position is varied. The probe atoms are placed on a three-dimensional lattice, or grid, in the region of the receptor and the nonbonded interaction energies between the probe and receptor are calculated. A comparison of contours generated with different probes may suggest regions favorable for each probe. It is important to note that these methods only consider the interaction between each probe and the receptor, and not intraligand interactions.

HINT [21] is an interesting example of a contouring method that uses the idea of a hydropathic map, incorporating the belief that hydrophobicity is a significant thermodynamic factor in ligand binding. This program finds the hydrophobic interaction profile of a known molecule and uses it to predict the profiles of complementary molecules. Profiles of the receptor may be used to predict a ligand structure or profiles of known ligands may be used to do the opposite and predict a receptor structure.

The multiple copy simultaneous search (MCSS) method takes a somewhat different aim at the ligand problem [22]. Like GRID, MCSS determines energetically favorable positions for probes and produces an assembly of fragments rather than contours. Many copies (1000–5000) of acetonitrile, methanol, acetate, methyl ammonium, dimethyl ether, methane and acetaldehyde probes are distributed throughout the binding site in positions not occupied by the fixed receptor target. The interaction energies of these copies with the receptor are simultaneously minimized in the environment of the protein. Interactions between the probes are not considered. The forces experienced by the receptor are normalized according to the number of probes, as probes are eliminated during the course of the minimization. In the end, a number of specific probes are positioned to minimize receptor–ligand interaction energy, one characteristic that distinguishes MCSS from contouring strategies. In a recent paper [23], fragments suggested by the MCSS method are connected using a pseudo-energy function as a guide and the resulting structures are further minimized by a Monte Carlo scheme in order to eliminate any stretched bonds or bad contacts.

The results of the methods described above may be used as a starting point for structure-generating programs such as SPROUT [24], BUILDER [9], LEGEND [25,26], CLIX [27], LUDI [28], HOOK [29] and GROUPBUILD [7]. These algorithms construct entire ligand molecules, often around key known fragments. Some of these algorithms, based on a scheme suggested by Lewis and Dean [30,31], begin by building a skeleton that spans the receptor pocket and satisfies steric requirements. In other cases, fragments are placed first and an interconnecting skeleton is then built, while in a few cases such as LEGEND [25], GEMINI [32], GROW [33] and GenStar [34] fragments are sequentially added without using a skeleton. To fill in a skeleton, ligand fragments often containing more than one heavy atom are drawn from a library of candidates

placed at a vertex of the skeleton and then scored according to the interaction energy with the receptor. A fragment is then randomly selected from among those with the most favorable energies. Empty skeleton sites are usually filled sequentially beginning from some initial point, although some algorithms are able to backtrack if a dead end is reached [25,26]. This process of selecting and placing ligand fragments is by no means exhaustive for even moderately sized receptors. Several hundred structures may be generated and subjected to post run processing in order to determine the best candidates, but this is a tiny fraction of the number of possible ligand fragment combinations.

With so many methods to aid in ligand design readily available, one might wonder what worthwhile improvements could be added. Methods that use known ligands, those in the first category, are sometimes unsuccessful because extensive databases are needed in order to have a reasonable chance of finding good ligand matches for a given receptor. Ultimately, the ability to generate novel ligand structures or descriptors using small fragment building blocks is highly desirable. Methods in the second category have this ability, but can be improved.

The methods in the second category do not give full consideration to relationships between ligand atoms that clearly affect binding. Bonding between ligand atoms largely determines their partial charges, thus defining the charge compatibility of nonbonded fragments and ultimately influencing binding energy. This can be seen by observing that powerful nonbonded ligand–ligand forces, such as van der Waals and Coulombic forces, are among the prominent factors governing ligand architecture and are known to be key in establishing acceptable spatial and energetic relationships among nonbonded atoms. The van der Waals forces effectively prevent nonbonded atoms from approaching one another as closely as bonded ones. Coulombic forces, which depend upon the partial charges and therefore the bonding of the atoms involved, may favor or discourage the presence of nearby nonbonded ligand atoms [32]. Thus, ligand atom positions, their partial charges and bonding are intimately involved with binding energy.

Likewise, it is well known that intramolecular hydrogen bonding and the hydrophobic effect are important to protein folding [1,32,35]. Both forces depend upon the presence of nearby ligand atoms and influence the energy differences between conformations of the free and bound forms of the ligand.

One way to improve upon existing schemes is to more fully develop intraligand energy calculations, but several difficult problems must be addressed in order to do this. As stated above, ligand atom bonding largely determines binding energy, so an accurate procedure to assign bonds and maintain correct valence states during ligand building

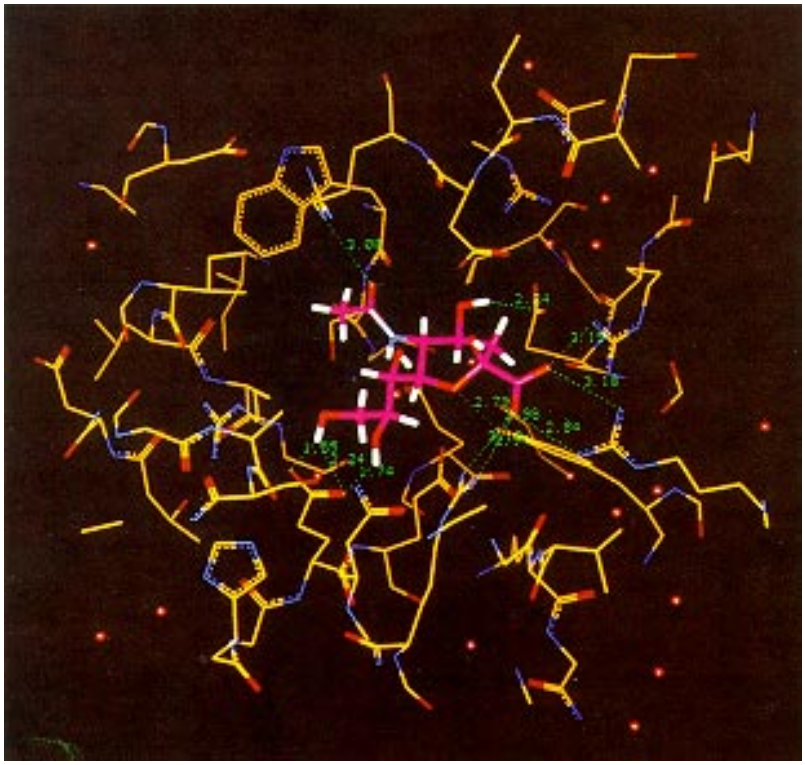

Fig. 1. Sialic acid bound to the neuraminidase enzyme. Oxygen atoms are shown in red, carbon atoms in magenta and nitrogen atoms in blue. Only hydrogen bonds between the ligand and receptor are shown in these figures. All ligand hydrogen bond distances are measured between heavy atoms. All the figures were produced using the INSIGHTII software.

is crucial. Partial charges and bonded energies must be calculated once bonding has been established and updated as ligand building proceeds. Both intramolecular and intermolecular hydrogen bonding energies must be computed during molecule building so that this process can be monitored and controlled. An algorithm that guides this process of sifting through the enormous number of ligand fragment combinations and associated interactions with the purpose of minimizing binding energy has a formidable task. The remainder of this paper describes an algorithm for performing this task.

The algorithm presented here fits into a broader strategy that addresses important issues only partially dealt with by existing strategies. A central point of this plan is to maximize information relevant to binding energy and use it during the placement of ligand fragments. For example, as fragments are bound together to form a ligand, their partial charges are greatly affected. Calculated interaction energies should reflect this fact. However, methods such as MCSS view ligand fragments as independent entities that energetically interact with the receptor but not among themselves. Ligand fragment bonding is only performed, for example using the HOOK routine, after MCSS has finished placing the ligand fragments. This strategy does not use bonding information to adjust partial charges. Instead, fixed partial charges are used during fragment placement so that charge changes that may significantly affect binding are not accounted for.

We illustrate a second central point of our plan by considering a scenario involving a ligand being built using a sequential routine, one that systematically builds a ligand by bonding one fragment to the next. If the routine only considers ligand–receptor Coulombic interactions, it may place two positively charged ligand fragments only a few bonds apart and create an energetically unfavorable situation. However, if the routine does check for interactions with other already placed ligand fragments, it may place a fragment without a strong positive charge in the second position because of the strong, preexisting positively charged first fragment. But it may have been better to put the less positive fragment in the first position and the more positive fragment in the second position. The problem is that sequential routines do not have an effective mechanism for dealing with situations where older fragments become outdated. In our plan, fragment placement is iterative. Previously existing fragments are often replaced as the ligand evolves.

A third central point of our plan has to do with how the actual energy minimization is conducted during the building process. While the decision to add, replace or remove fragments is based on energy, it is carried out using an annealing algorithm that helps avoid local energy minima traps during the actual building process. Randomly selected fragments are sometimes added to or removed from the ligand even when there is an unfavorable energy consequence. In principle, the algorithm decreases the chance that the building process will paint itself into a corner by freezing in older fragments and overly constraining the placement of new fragments.

Overview of the proposed method

An annealing algorithm to search for the collection of ligand fragments that binds best to the chosen receptor site is proposed here [36]. The algorithm effectively searches for the best ligand fragment arrangements in a manner that avoids evaluating the vast majority of energetically unacceptable ligand fragment combinations. The shape and atomic composition of the ligand are determined by the dimensions of the receptor site and the interaction energy of ligand fragments with the ligand–receptor complex. The current method falls between GRID and structure-generating algorithms in terms of the structures it produces. Like GRID, the method places ligand fragments on a 3D lattice and evaluates the interaction energy, but rather than energy contours the final results are collections of fragments. Unlike sequential structure-generating methods, the current scheme does not produce complete molecules because a constraint to produce a single molecule has not yet been added. However, the method is better suited than sequential methods for identifying key fragment combinations because atoms are positioned randomly rather than sequentially. Random positioning is an effective means for replacing ligand fragments that become obsolete as changes in the ligand being built alter intraligand interaction energies.

The goal of this annealing algorithm is to minimize an objective function, E, which represents the energy of the ligand–receptor system. The objective function used in this study is discussed below. The energy of the system, as estimated by E, is altered by adding or removing ligand fragments. Receptor atoms are held fixed. To begin minimizing E, a random change is made to an initial arbitrary arrangement of ligand fragments located within the chosen site. A randomly chosen individual fragment is placed at a random position within the receptor region.

The energy of the ligand–receptor system is estimated using the objective function after each fragment placement. The energy change between the new and old systems, E, is then calculated by subtracting the energy of the old system, E_{old}, from that of the new, E_{new}:

$$E = E_{new} - E_{old} \quad (1)$$

If the energy difference is negative ($E < 0$), the new arrangement is accepted and replaces the old. On the other hand, if the difference is not favorable, for which $E \geq 0$, the switch is accepted only if Eq. 2 holds.

$$X < e^{-E/kT} \quad (2)$$

In this equation, k is the Boltzmann constant, T is the temperature of the system and X is a randomly picked number between 0 and 1.

Simulated annealing begins with a relatively high temperature. After some number of attempted placements, the temperature is either reduced according to an annealing schedule and again held, or the annealing process is terminated because too few changes took place. According to Eq. 2, energetically unfavorable modifications become less likely as the temperature decreases. Favorable ligand fragments are eventually frozen into place as the temperature decreases further.

To implement the above scheme, coordinate data are read into a uniformly spaced 3D lattice that surrounds the space occupied by the ligand–receptor system. Each receptor and ligand fragment center is associated with a unique grid node that tracks fragment information. For a given fragment, information may be retrieved using the three grid coordinate indices that identify the grid node. In this treatment, ligand fragments may be placed only on grid nodes. The number of lattice nodes, specified beforehand, thus determines the resolution to which ligand fragments are placed.

The initial state of the receptor–ligand system is assumed to be given, with known atom types and coordinates. Partial charges on the receptor atoms must be given, but the partial charges are not required for ligand fragments since these are generated by the program as explained below. All calculations use original receptor coordinates with the exception of solvation energy calculations, as will be explained. The complete prototype library, given in Table 1, includes the NULL fragment that has no size, charge or surface area and effectively deletes any fragment it replaces.

Objective function

The objective function in this study is of the form

$$\begin{aligned} E = {} & \Re(\text{bonded}) + \Re(\text{nonbonded}) + \Re(\text{solvation}) \\ & + \Re(\text{solvent–solute polarization}) \\ & + \Re(\text{hydrogen bond}) \end{aligned} \quad (3)$$

The above summations are taken over all ligand–ligand and ligand–receptor pairs in order to minimize the overall system energy. Receptor–receptor interactions are not included because the receptor, and hence the interaction energy between receptor fragments, is assumed to be fixed.

Bonded and nonbonded energies

Expressions for van der Waals, electrostatic, bonded and nonbonded energies are taken from the DREIDING force field along with related parameter values [37]. Bonded terms include stretching and 1,3 angle bending terms. The nonbonded energies of Eq. 3 are not included between bonded atoms. The manner in which bonds and partial charges are assigned to ligand fragments is discussed separately below.

TABLE 1
COMPLETE FRAGMENT LIBRARY WITH FRAGMENTS CHOSEN IN TRIALS INDICATED

Fragment	Description	Used in trials 1–4?
NULL	Has no atoms, size, charge, etc.	Yes
CH_3	$-CH_3$	Yes
CH_2	$-CH_2-$	Yes
CH	CH with three open bonds	Yes
C	C is one double, two single bonds	Yes
NH_2P	$=NH_2^+$	No
NHP	NH^+ quaternary	No
NH_2	$-NH_2$	No
NH_3	$-NH_3^+$	No
NH	-NH-	Yes
N	=N-	No
OH_2	Water	No
OH	-OH	Yes
O	-C=O	Yes
O^-	$-O^-$	Yes
OX	Ether oxygen -O-	No
F	Fluorine -F	No
Cl	Chlorine -Cl	No
Br	Bromine -Br	No
P	Phosphorus -P	No
SH	Thio -SH	No
S	Sulfur -S-	No

Solvation and solvent–solute polarization energies

The 3D grid influences solvation energy calculations since it is used in computing the standard state free energy that includes cavity formation and solvent dispersion [38]. The solvation energy is calculated by moving ligand atoms to the nearest lattice node and marking each lattice point that falls within the van der Waals surface of the atoms. Surface points not overlapping other points are tallied and the accessible surface area of each fragment is calculated as suggested by Wang and Levinthal [39]. The standard state free energy of cavity formation and solvent dispersion is calculated from the accessible surface by the method of Cramer and Truhlar [38,40–45], who also derived an expression for the standard state free energy associated with solute–solvent polarization. The polarization free energy is a function of distance and partial charges and is computed after ligand fragment charges are established.

Hydrogen bond energy

A novel competitive scheme estimates hydrogen bonding between fragments. Fragments with hydrogens to donate compete with other donors for fragments able to accept some or all of their hydrogens. Fragments like OH

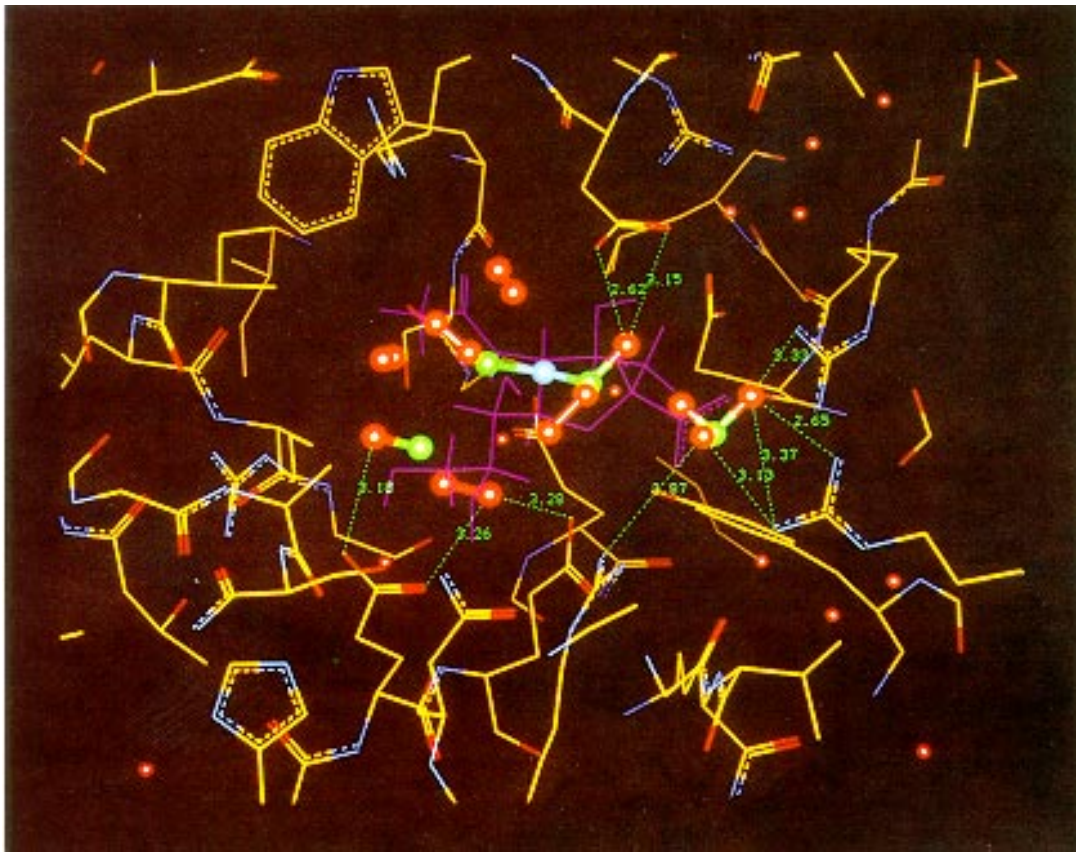

Fig. 2. Results of trial 1. (In Figs. 2–5, all ligand hydrogen bond distances are measured between heavy atoms.)

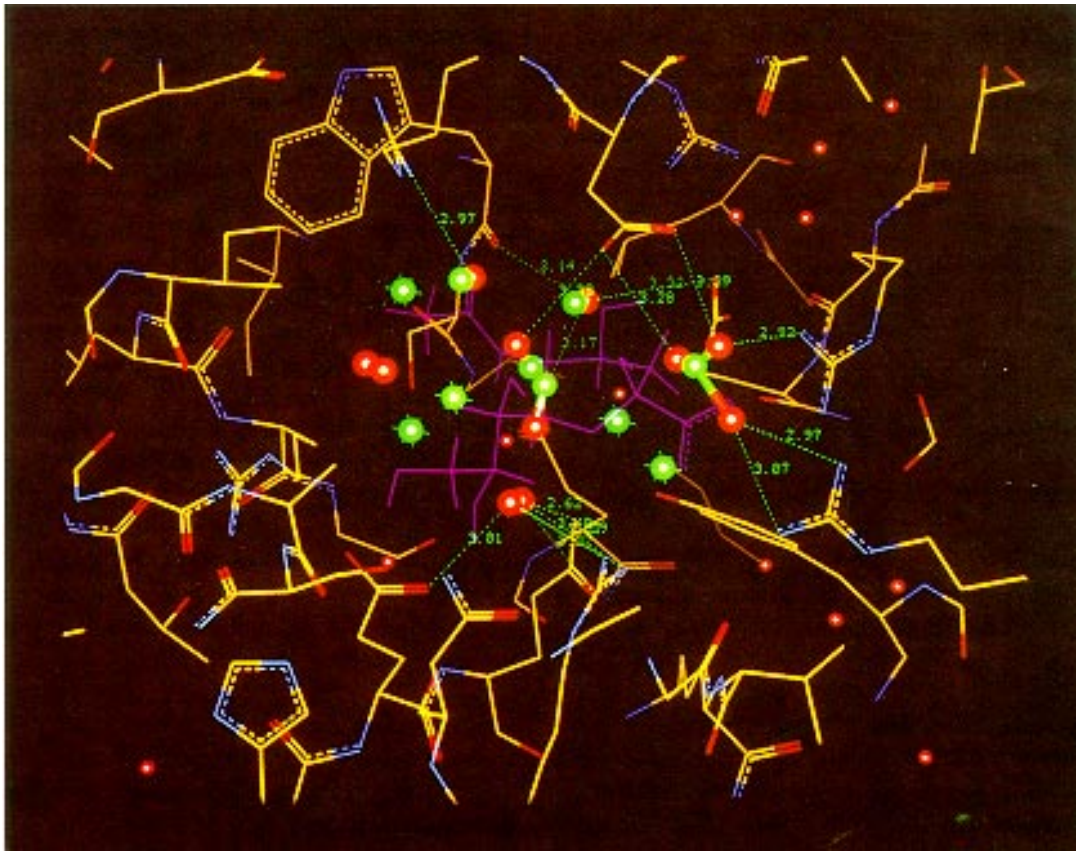

Fig. 3. Results of trial 2.

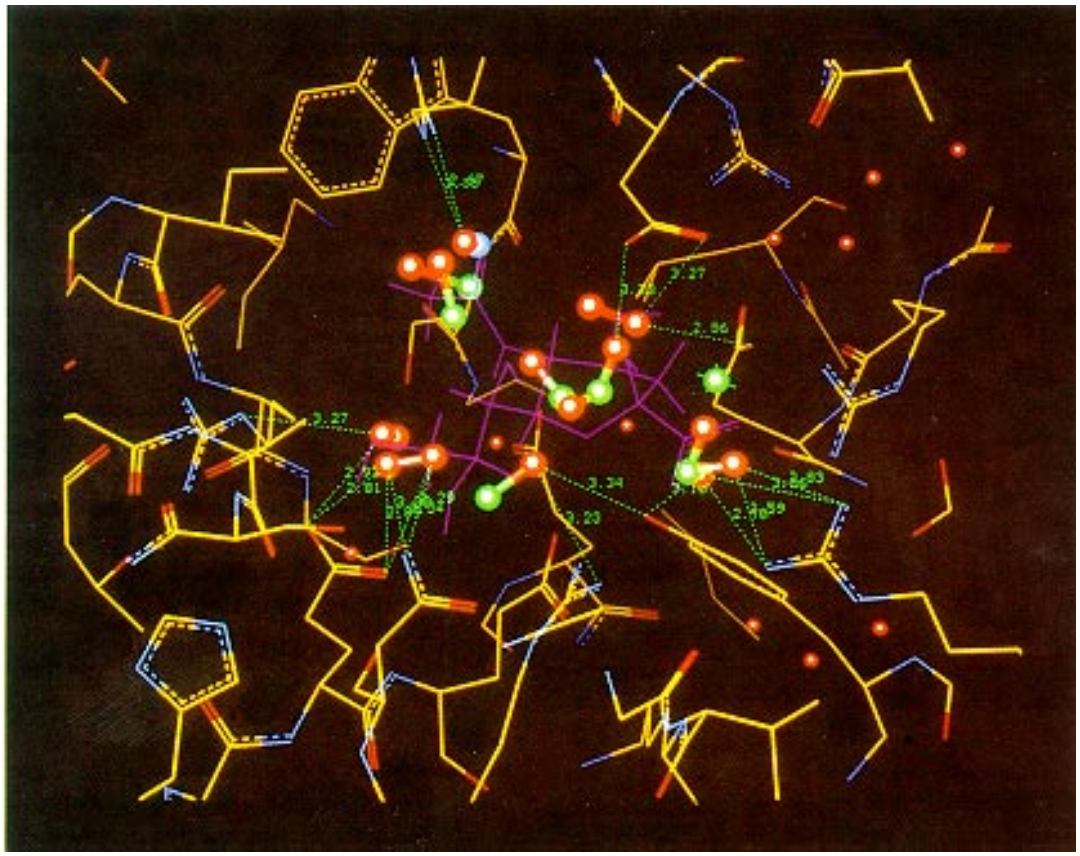

Fig. 4. Results of trial 3.

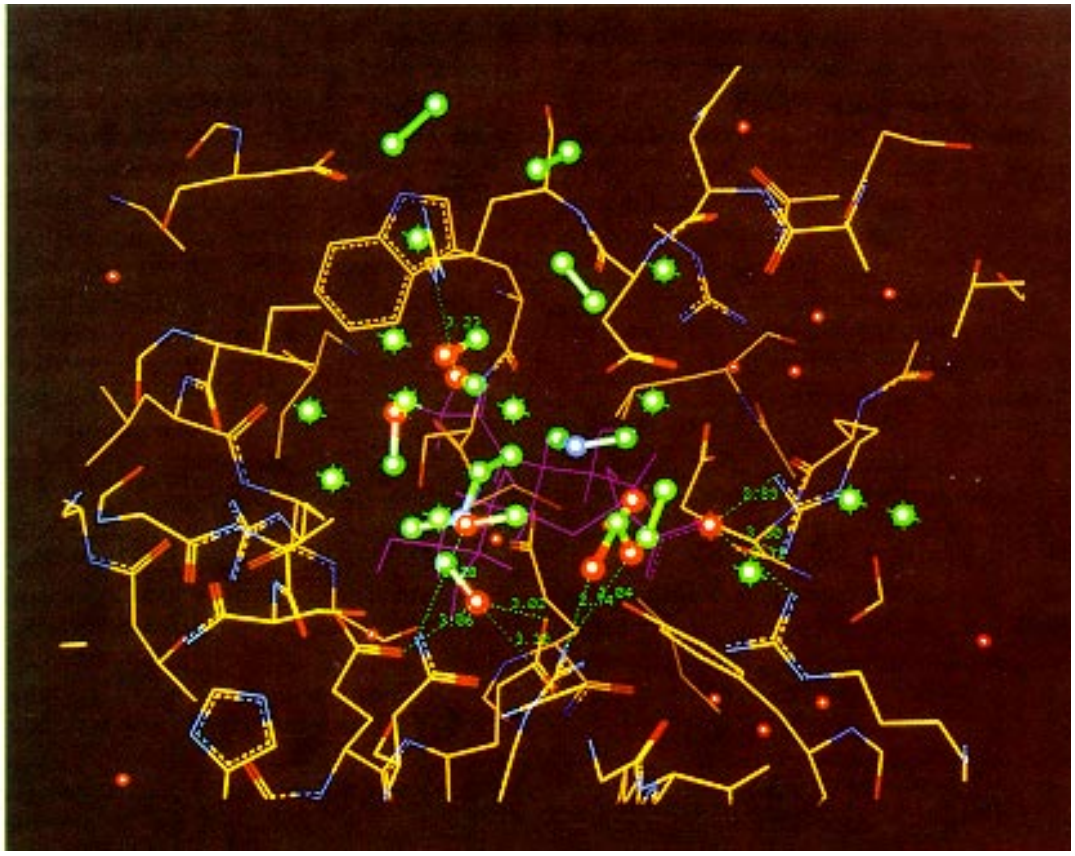

Fig. 5. Results of trial 4.

may be both donors and acceptors. The scheme ensures that the number of hydrogens any fragment can donate or accept is never exceeded. Donor–acceptor pairs with the strongest hydrogen bonding energies are kept and the others are discarded. The scheme is executed whenever a switch is attempted involving fragments able to participate in hydrogen bonding.

For ligand fragments that donate hydrogens, the scheme orients these hydrogens in order to maximize the DREIDING hydrogen bonding energy, E_{hb}, given by Eq. 4. This is not so in the case of donating receptor fragments where original hydrogen coordinates are always used. Ligand fragments may hydrogen bond with both receptor and other ligand fragments. In the case of ligand–ligand hydrogen bonding, the hydrogens of both fragments are oriented.

$$E_{hb} = D_{hb}[5(R_{hb}/R_{DA})^{12} - 6(R_{hb}/R_{DA})^{10}] \cos^4(\angle DHA) \quad (4)$$

where $\angle$DHA is the bond angle formed between the center of the hydrogen donor, the hydrogen and the center of the accepting fragment. R_{DA} is the distance between the donor and acceptor in Å. D_{hb} and R_{hb} are 4.0 kcal/mol and 2.75 Å, respectively.

The routine begins by searching the entire ligand–receptor space for hydrogen bond donors. If a donor is found, the search is temporarily halted. If the donor is a receptor fragment, the coordinates of its hydrogens are retrieved from a file and used for estimating hydrogen bond energy. For ligand donors, the positions of hydrogens are oriented as explained below. In either case, a sweep to locate all possible hydrogen bond acceptors within a 6 Å distance of the donor is initialized. If an acceptor fragment is found, another 6 Å region about it is searched for other possible hydrogen bond donors. Thus, three searches may occur simultaneously. The first is a search for donors, the second for acceptors, and the third for competing donors to a given acceptor.

During the third sweep, the hydrogen coordinates of receptor hydrogen donor fragments are used to compute E_{hb} for each hydrogen in the donor fragment. If the donor is a ligand fragment, its first hydrogen is oriented directly toward the acceptor so that E_{hb} is computed using Eq. 4 with $\angle$DHA set to zero. The coordinates of these donors are placed on a list along with the estimated E_{hb}. The third sweep then continues searching for other donors in this manner until the entire region around the acceptor has been explored. At this point, the list of competing donors is sorted according to E_{hb} and trimmed so that it at most contains the number of donor hydrogens that the acceptor can accommodate. If the coordinates of the original donor fragment appear on this list, then the coordinates of the acceptor and E_{hb} are placed on a second list. The second search then resumes in this manner, building a list of possible acceptors for the original donor. When the search is completed, the second list is sorted by E_{hb}. At this point, the second list contains acceptors able to accommodate at least one hydrogen bond from the original donor ranked by E_{hb}. The list of acceptors serves as a starting point for computing E_{hb} for the donor's second and third hydrogens, if they exist.

We now describe how hydrogen bonding is determined when the donor is a receptor fragment. Here, stored hydrogen coordinates are used in the case of a receptor donor to determine the angle $\angle$DHA in Eq. 4. If the donor has second and third hydrogens, the coordinates of each acceptor that can potentially accommodate one or both hydrogens are reentered on the second list along with their new E_{hb}. Thus, an acceptor may appear more than once on the list. To establish the final donor list, the donor list is pruned so that it contains no more acceptor entries than the donor has hydrogen atoms to donate. The routine then resumes the first search for donors and the process repeats until all donors have been examined.

A similar situation takes place if the donor is a ligand fragment, except that hydrogen coordinates are calculated rather than read from a file. Mathematical details of the calculations are somewhat long and are left to the Appendix. This calculation begins by pointing the first hydrogen being donated directly at the acceptor having the best E_{hb}, so that the donor, acceptor and hydrogen lie on a straight line. The E_{hb} between any second hydrogen and acceptor on the list requires that $\angle$DHA in Eq. 4 is found. This is done by rotating the second hydrogen about the line of the first hydrogen bond so that the new acceptor lies in the plane of the first two hydrogens and their central atom. During rotation, proper angles and distances between donor hydrogens and the central atom are maintained, these depending on the elemental type of the central atom. The plane may be formed with the second hydrogen in either of two positions, so we choose the position that places it closest to the target acceptor. $\angle$DHA is thus determined so that E_{hb} may be estimated. By repeating this process for every acceptor on the list, the second bond may be assigned and the position of the second hydrogen finalized. The position of any third donor hydrogen is determined once the second is fixed.

Bonding

The bond routine begins by searching for ligand fragment centers. When a fragment with unmatched valence electrons is found, a temporary bond is assigned between the fragment and all ligand atoms within a specified distance. However, temporary bonds are not assigned between fragment pairs that do not have matching bond orders, such as between the carbonyl oxygen and hydroxyl fragments. Bond stretching, nonbonded electrostatic and van der Waals energies are computed using DREIDING parameters [37].

Fragments then compete for bonding partners in order to determine the best all-around choices. A given ligand fragment with n available single or multiple bonds is considered bonded to another ligand fragment if two conditions hold: (i) The difference between the bonded and nonbonded energies of the fragments must be one of the n smallest of all bond pairings possible with the given fragment. (ii) The same condition must hold for the second fragment – that is if the second fragment has m possible bond pairings, then the energy difference associated with the fragments must be among the m smallest. These bonding preferences are established using the following recursive routine.

The bonded–nonbonded energy differences between a bond donor and a possible partner are calculated using the equation

$$E_{diff} = E_{bstretch} - E_{nb} \quad (5)$$

$E_{bstretch}$ is the energy due to bond stretching and E_{nb} is the nonbonded Coulombic and van der Waals energy between the two fragments. E_{diff} is placed on a list along with other relevant information and sorted. A similar list is constructed for each bond acceptor, then all donors to that acceptor, all acceptors to those donors, and so on, until a predetermined number of recursions are reached. At the final level of recursion, the list of partners is sorted and trimmed so that it does not exceed the number of bonds that the current fragment is capable of forming. This list is returned to the requesting fragment where each requested list is cross-matched with that of the caller, rejecting any mismatches. The list is then trimmed so that the total number of valence electrons used in bonding does not exceed the number available to the calling fragment. A final check is performed once all bonds are assigned by comparing each fragment to that of its partners to again verify that each is on the other's list. If the verification fails, no bond is assigned. This novel bonding scheme yields reasonable assignments using only two levels of recursion.

Ligand bonding and partial charge calculations

This implementation uses a modified version of the Lewis–Langmuir method [46] to estimate fragment charges. In the modified version, fragment charges are calculated by simply adding the hydrogen charges to that of the heavy atom. After each fragment switch, fragments with unmatched bonds remaining are assumed to be connected to carbon for the purpose of calculating partial charges.

Fragment placement

A new ligand fragment may be placed so that it overlaps and thus requires deletion of an existing ligand fragment. However, receptor fragments may not be deleted, so the new configuration is rejected if the new fragment overlaps a receptor atom. Fragments are considered to overlap if their distance from each other is less than 0.1 times the combined van der Waals radii of the fragments. After preliminary checks, the new ligand fragment is positioned and E_{new} is calculated from Eq. 1. The acceptance or rejection of the fragment is based on the Monte Carlo algorithm discussed earlier.

Results

Many factors enter into choosing a test case for the algorithm. It is desirable to use available crystal structure data for a ligand bound to a protein having well-characterized binding interactions. If possible, the binding site should include both polar and nonpolar regions in order to assess how well the algorithm discriminates polarity when positioning ligand fragments. It is also desirable to have knowledge of any ligand groups that are thought to be important to binding so that these can be compared in the test results.

The neuraminidase active site was chosen as the test receptor. Sialic acid bound to neuraminidase in the crystal structure obtained from the Protein Data Base is shown in Fig. 1. This site is a well-studied [28] cavity with polar and nonpolar regions. The site was chosen in order to compare the newly built ligand with the sialic acid observed in the crystal structure. The interaction of a carboxylic acid group with arginine residues of the enzyme is a common feature of sialic acid-based inhibitors and substrates that bind at this site. This site provides a good test of the algorithm's ability to place polar carboxylic and hydrogen bonding fragments along with nonpolar fragments.

To test the algorithm, protein atoms within 6 Å of the sialic acid in the neuraminidase active site were selected to be used as the test receptor. A total of 367 heavy atoms, including several important waters of hydration, were thus selected. A subset of the complete fragment library was chosen for these trials, as indicated in Table 1. Notably, the subset did not contain an ether oxygen even though one is present in the known ligand.

Four separate trials of the algorithm were carried out; each required approximately 2 weeks to complete on a Silicon Graphics IRIS 4D/340 with the program running on one processor. Each trial began with an empty receptor pocket. The results of these trials are shown in Figs. 2–5. Bound sialic acid, whose purple outline appears in each figure, serves as a reference to compare against the fragments placed in each trial. The first three trials used 20 000 grid nodes at a spacing of approximately 0.77 Å. Fragment switching was restricted to a subregion slightly larger than the region occupied by sialic acid when bound to the enzyme. The subregion contained less than 1000 nodes. The last trial used 80 000 nodes at a spacing of about 0.49 Å and a switching subregion having dimen-

sions about 6 Å larger than the subregion used for the first three trials.

The initial value of kT was set to 10 in all the trials. In the first trial, this initial value was held for approximately 4000 iterations before changing kT to 5 for the same number of iterations. kT was then dropped to 1 and held for about 40 000 iterations. The results of trial 1 are shown in Fig. 2. In the second trial, approximately 10 000 iterations were performed for kT values of 10, 5, 2.5 and 0.01. These results are shown in Fig. 3. The results of the next trial are listed in Fig. 4. Here, kT was dropped from 10 after 10 000 iterations directly to 0.01, for which 20 000 iterations were completed. The results of the last trial are shown in Fig. 5. Although fewer iterations were performed at each kT value during trial 4, the computational time was similar to that of the previous three trials because of the larger switching region and number of grid nodes used. Fifteen hundred iterations were performed at kT = 10, 2000 iterations at kT = 2.5 and 1.0 and 4000 iterations at kT = 0.01.

All four trials produced a ligand with a carbon connected to three acid oxygen atoms. The location is near the three receptor arginine residues in the active site and close to the position of the acid group of the bound sialic acid. No acid oxygen atoms were positioned elsewhere in the binding site although several other types of oxygen atoms are present. Nonpolar carbon atoms are also positioned at hydrophobic locations, such as the region near the methyl group of the acetyl in sialic acid, in these examples.

Since the present energy equation does not include terms or constraints to promote the formation of a single, connected ligand molecule, the ligands appear as groups of fragment clusters. As a result, clusters form that have no open bonds available for linking to other clusters. For example, several O=O clusters appear in the results. Each oxygen is uncharged and has no Coulombic interaction. However, an examination of these clusters using molecular graphics suggests that favorable hydrogen bonding with the receptor may account for their survival.

About six hydrogen bonds are formed between sialic acid and the receptor besides the hydrogen bonds formed with the carboxyl group. The ligand of trial 1 (Fig. 2) shows about five hydrogen bonds, trial 2 (Fig. 3) shows at least 10, trial 3 (Fig. 4) has at least 14 and about eight hydrogen bonds are evident in trial 4 (Fig. 5).

The extra hydrogen bonds in trials 2 and 3 are largely due to several well-positioned oxygen atoms, each having multiple hydrogen bonds with the protein. In trial 2 (Fig. 3), two well-positioned oxygen atoms form six hydrogen bonds. Similarly, in trial 3 (Fig. 4) two oxygen atoms make five hydrogen bonds.

These results demonstrate the ability of the algorithm to position ligand fragments that are capable of forming hydrogen bonding despite the constraint that these fragments are placed only on grid nodes. Many of the ligand oxygen atoms participating in hydrogen bonds are themselves bonded to other ligand oxygen atoms. As a result, these oxygen atoms do not have valence electrons available for bonding to other atoms, resulting in clusters of bonded fragments that prevent the formation of a single, complete and connected ligand molecule. This cause of fragment clustering can be eliminated by forbidding pairings that consume all available bonds of both ligand fragments.

Discussion and Conclusions

The correct placement of the acid group in all four trials is most encouraging, especially considering the simplicity of the method. Nonpolar carbon atoms were also favorably placed and considerable hydrogen bonding is evident in the results. In two of the experimental trials, the algorithm formed significantly more hydrogen bonds than are found between the natural ligand and the protein.

Ligand–ligand interactions were included based on the belief that ligand–receptor interactions alone are not sufficient for constructing reasonable ligands. Ligand fragments that do not directly affect the ligand–receptor energy often have an indirect influence due to their interactions with other ligand fragments. An example of such an indirect influence occurs when an oxygen fragment bonds to a carbon fragment and the partial charges change in a manner that causes a favorable Coulombic interaction with the receptor.

The time necessary to complete the large number of calculations required by this method is a matter of some concern. Speed can be increased by mapping regions where fragments may be successfully placed within the target site before beginning the annealing process. The exclusion of nonproductive regions should improve the speed by preventing useless attempts at switching. Faster hardware, especially multiple processors, would obviously improve calculation speed.

This project achieved several important goals. Ligand fragments were placed in a manner that took into account the energy of the entire complex, rather than just ligand–receptor interactions. To do this, energy calculations addressed hydrogen bond orientations and interligand bonding. The results indicated that the method is useful for ligand design. Ligand atoms with widely differing polarities were favorably placed knowing only the structure of the receptor. The placement of important acidic and hydrogen bonding groups clearly demonstrates that the method is able to suggest the nature and location of important ligand fragments.

The trials described above have provided insights that will be implemented in future versions of the algorithm. For example, constraints to form a single ligand molecule

rather than multiple clusters will be added. We optimistically look forward to future explorations in rational ligand design with this approach.

Acknowledgements

This work is from a thesis submitted to the graduate school of the Illinois Institute of Technology in partial fulfillment of the requirements for the Ph.D. degree. One of us, P.J.Z., is deeply grateful to several individuals whose generous sharing of their knowledge, advice and time made this project possible. These are Peter Lykos, Stan Kantor and Ross Overbeek. Sincere thanks are also expressed to the Argonne National Laboratory for the use of their SP1 parallel computer in obtaining many of the results documented here.

References

1 Andrews, P., In Foye, W. (Ed.) Principles of Medicinal Chemistry, 3rd ed., Lea & Febiger, Philadelphia, PA, U.S.A., 1990, pp. 855–860.
2 Bugg, C., Carson, W. and Montgomery, J., Sci. Am., 269 (1993) 92.
3 Von Itzstein, M., Wu, W., Kok, G., Pegg, M., Dyason, J., Jin, B., Pahn, T., Smythe, M., White, H., Oliver, S., Colman, P., Varghese, J., Ryan, M., Woods, J., Bethell, R., Hotham, V., Cameron, J. and Penn, C., Nature, 363 (1993) 418.
4 Kuntz, I., Blaney, J., Oatley, S., Langridge, R. and Ferrin, T., J. Mol. Biol., 161 (1982) 269.
5 DesJarlais, R., Sheridan, R., Seibel, J., Dixon, S., Kuntz, I. and Venkataraghavan, R., J. Med. Chem., 31 (1988) 722.
6 Martin, Y., J. Med. Chem., 35 (1992) 2145.
7 Rotstein, S. and Murcko, M., J. Med. Chem., 36 (1993) 1700.
8 Namasivayam, S. and Dean, P., J. Mol. Graph., 4 (1986) 46.
9 Lewis, R., Roe, D., Huang, C., Ferrin, T., Langridge, R. and Kuntz, I., J. Mol. Graph., 10 (1992) 66.
10 Kato, Y., Akiko, I. and Iitaka, Y., Tetrahedron, 43 (1987) 5229.
11 Meng, E., Shoichet, B. and Kuntz, I., J. Comput. Chem., 13 (1992) 505.
12 Karfunkel, H., J. Comput. Chem., 7 (1986) 113.
13 DesJarlais, R., Sheridan, R., Dixon, S., Kuntz, I. and Venkataraghavan, R., J. Med. Chem., 29 (1986) 2149.
14 Billeter, M., Havel, T. and Kuntz, I., Biopolymers, 26 (1987) 777.
15 Goodsell, D. and Olson, A., Proteins, 8 (1990) 195.
16 Hart, T. and Read, R., Proteins, 13 (1992) 206.
17 Goodford, P., J. Med. Chem., 28 (1985) 849.
18 Boobbyer, D., Goodford, P., McWhinni, P. and Wade, R., J. Med. Chem., 32 (1989) 1083.
19 Wade, R., Clark, K. and Goodford, P., J. Med. Chem., 36 (1993) 140.
20 Wade, R. and Goodford, P., J. Med. Chem., 36 (1993) 148.
21 Kellogg, G. and Abraham, D., J. Mol. Graph., 10 (1992) 212.
22 Miranker, A. and Karplus, M., Proteins, 1 (1991) 29.
23 Caflisch, A., Miranker, A. and Karplus, M., J. Med. Chem., 35 (1993) 2142.
24 Gillet, V., Johnson, P., Mata, P., Sike, S. and Williams, P., J. Comput.-Aided Mol. Design, 7 (1993) 127.
25 Nishibata, Y. and Akiko, I., Tetrahedron, 47 (1991) 8985.
26 Nishibata, Y. and Akiko, I., J. Med. Chem., 36 (1993) 2921.
27 Lawrence, M. and Davis, P., Proteins Struct. Funct. Genet., 12 (1992) 31.
28 Böhm, H.-J., J. Comput.-Aided Mol. Design, 6 (1992) 61.
29 Eisen, M., Wiley, D., Karplus, M. and Hubbard, R., Proteins Struct. Funct. Genet., 19 (1994) 199.
30 Lewis, R., J. Comput.-Aided Mol. Design, 4 (1990) 205.
31 Lewis, R.A. and Dean, P.M., Proc. R. Soc. London, B236 (1989) 125.
32 Singh, J., Saldanha, J. and Thornton, J., Protein Eng., 4 (1991) 251.
33 Moon, J. and Howe, W., Proteins Struct. Funct. Genet., 11 (1991) 314.
34 Rotstein, S. and Murcko, M., J. Comput.-Aided Mol. Design, 7 (1993) 23.
35 Pickett, S.D. and Sternberg, M.J.E., J. Mol. Biol., 231 (1993) 825.
36 Zielinski, P.J., An Annealing Algorithm for Designing Ligands from Receptor Structures, UMI, Ann Arbor, MI, U.S.A., 1994, p. 1.
37 Mayo, S., Olafson, B. and Goddard, W., J. Phys. Chem., 94 (1990) 8897.
38 Cramer, C. and Truhlar, D., J. Am. Chem. Soc., 113 (1991) 8305.
39 Wang, H. and Levinthal, C., J. Comput. Chem., 12 (1991) 868.
40 Cramer, C. and Truhlar, D., J. Am. Chem. Soc., 113 (1991) 8552.
41 Cramer, C. and Truhlar, D., Chem. Phys. Lett., 198 (1992) 74.
42 Cramer, C. and Truhlar, D., J. Comput. Chem., 13 (1992) 1089.
43 Cramer, C. and Truhlar, D., J. Comput.-Aided Mol. Design, 6 (1992) 629.
44 Cramer, C. and Truhlar, D., Science, 256 (1992) 213.
45 Cramer, C. and Truhlar, D., J. Am. Chem. Soc., 114 (1992) 8226.
46 Allen, L., J. Am. Chem. Soc., 111 (1989) 9115.

Appendix

Hydrogen bonding method for locating the second and third hydrogens

Begin by translating the origin to the center of the donor fragment so that the donor is located at the point (0,0,0). Let H_0 denote a vector from the donor to a hydrogen attached to the donor and let H_1 and H_2 likewise denote similarly defined vectors for additional hydrogens. Consider the case of a donor fragment with two hydrogens to donate. Let ψ be the fixed angle $\angle H_0DH_1$ formed between the two hydrogens with the donor center, D, as its vertex. Also let θ be the angle $\angle H_1DT$, where T is the vector from the donor to the targeted acceptor with which the second hydrogen is interacting.

Assume that H_0, D, H_1 and T lie in the same plane. Note that H_1 can always be rigidly rotated about the bond from D to H_0 to achieve this. Then the angles ψ and θ also lie in that plane and the angle $\sigma = \angle H_0DT$ is given by

$$\sigma = \theta + \psi$$

Let $\beta = \angle TH_1D$. By applying the law of cosines,

$$d(D,T)^2 = d(H_1,T)^2 + d(H_1,D)^2 - 2d(H_1,T)\,d(H_1,D)\cos\beta$$

or, upon rearranging,

$$\cos\beta = \frac{\left[d(H_1,T)^2 + d(H_1,D)^2 - d(D,T)^2\right]}{\left[2d(H_1,D)^2\,d(D,T)^2\right]}$$

To find $d(H_1,T)$, the law of cosines is applied again.

$$d(H_1,T)^2 = d(H_1,D)^2 + d(D,T)^2 - 2d(H_1,D)^2\,d(D,T)^2\cos\theta$$

At this point, $\cos\beta$ may be computed by substituting this last expression into the previous one. After simplification, one obtains

$$\cos\beta = \frac{\left[d(H_1,D) - 2d(D,T)\cos\theta\right]}{\sqrt{\left[d(H_1,D)^2 + d(D,T)^2 - 2d(H_1,D)\,d(D,T)\cos\theta\right]}}$$

where $\cos\theta$ may be easily calculated using the relationship

$$\theta = \sigma - \psi$$

The coordinates (u,v,w) of the second hydrogen, H_1, may be found using the known coordinates (a,b,c) of the first hydrogen, H_0, and the known coordinates (T_1,T_2,T_3) of the target acceptor, T. Since both the hydrogens and the target are in the same plane, it follows that the determinant of the matrix formed from these coordinates is zero,

$$\det\begin{vmatrix} u & v & w \\ a & b & c \\ T_1 & T_2 & T_3 \end{vmatrix} = 0$$

Furthermore, since the two hydrogens are separated by the fixed angle ψ, we have from elementary vector analysis that the dot product between the hydrogens is given by

$$H_1 \bullet H_2 = \rho^2\cos\psi$$

where ρ is the distance from each hydrogen to the donor's center. Likewise, H_1 and T are separated by an angle θ from which is obtained

$$H_1 \bullet T = \rho\, T\cos\theta$$

where T is the length of the vector from the center of the donor fragment to the center of the target acceptor fragment.

Combining these last three equations in the matrix equation form $\mathbf{M}H_1 = Y$ gives

$$\begin{vmatrix} a & b & c \\ T_1 & T_2 & T_3 \\ bT_3 - cT_2 & cT_1 - aT_3 & aT_2 - bT_1 \end{vmatrix} \begin{vmatrix} u \\ v \\ w \end{vmatrix} = \begin{vmatrix} \rho^2\cos\psi \\ T\rho\cos\theta \\ 0 \end{vmatrix}$$

Let $\det(\mathbf{M})$ be the determinant of the matrix on the left side of the last equation. Then the vector of H_1 coordinates, (u,v,w), is the solution to this matrix equation and is given by

$$u = \{\rho/\det(\mathbf{M})\}\{\rho\cos\psi[T_2(aT_2 - bT_1) - T_3(cT_1 - aT_3)] - T\cos\theta[b(aT_2 - bT_1) - c(cT_1 - aT_3)]\}$$

$$v = \{-\rho/\det(\mathbf{M})\}\{\rho\cos\psi[T_1(aT_2 - bT_1) - T_3(bT_3 - cT_2)] - T\cos\theta[a(aT_2 - bT_1) - c(bT_3 - cT_2)]\}$$

$$w = \{\rho/\det(\mathbf{M})\}\{\rho\cos\psi[T_1(cT_1 - aT_3) - T_2(bT_3 - cT_2)] - T\cos\theta[a(cT_1 - aT_3) - b(bT_3 - cT_2)]\}$$

At this point, the coordinates of all entities are known except for a possible third hydrogen, H_2. Assume that H_2 is located at coordinates (x,y,z). Observe that (x,y,z) is not uniquely determined by the known angles and distances that have been given or calculated thus far, but the value of the triplet (x,y,z) is constrained by the fact that the angle between these coordinates and the first two hydrogens is fixed, which leads to the relationship

$$H_1 \bullet H_2 = H_0 \bullet H_2 = \rho^2\cos\psi$$

from which it follows that this constraint may be written as

$$(H_1 - H_2) \bullet H_2 = (u-a)x + (v-b)y + (w-c)z = 0$$

Let T once again represent a vector from the donor to a target acceptor that is interacting with the third hydrogen. The coordinates (x,y,z) of H_2 desired are those that minimize the distance $d(T,H_2)$ or, equivalently, the square of that distance, $d(T,H_2)^2$.

The method of Lagrange undetermined multipliers may be used to find the minimal distance subject to the given constraint. To use this method, a new variable λ is introduced and the function $F_\lambda = F(H_0,H_1,H_2,T,\lambda)$ is defined as

$$F_\lambda = (T_1 - x)^2 + (T_2 - y)^2 + (T_3 - z)^2 + \lambda[x(u-a) + y(v-b) + z(w-c)]$$

Next, compute the derivatives F_x, F_y, F_z and F_λ of the function F with respect to each of the four variables contained in the last equation and set each derivation equal to zero. This step gives

$$F_x = -2(T_1 - x) + \lambda(u - a) = 0$$

$$F_y = -2(T_2 - y) + \lambda(v - b) = 0$$

$$F_z = -2(T_3 - z) + \lambda(w - c) = 0$$

$$F_\lambda = x(u - a) + y(v - b) + z(w - c) = 0$$

Upon rearranging and putting this system of linear equations into matrix form, we have

$$\begin{vmatrix} 2 & 0 & 0 & u-a \\ 0 & 2 & 0 & v-b \\ 0 & 0 & 2 & w-c \\ u-a & v-b & w-c & 0 \end{vmatrix} \begin{vmatrix} x \\ y \\ z \\ \lambda \end{vmatrix} = \begin{vmatrix} 2T_1 \\ 2T_2 \\ 2T_3 \\ 0 \end{vmatrix}$$

The (x,y,z) triplet that satisfies this equation is

$$x = K\{2T_1[(w-c)^2 - (v-b)^2] + 2(u-a)[T_2(v-b) + T_3(w-c)]\}$$

$$y = K\{-T_2[(w-c)^2 + 2(u-a)^2] + 2T_1(v-b)(u-a) + T_3(w-c)(v-b)\}$$

$$z = K\{(v-b)[T_2(w-c) - T_3(v-b)] + 2(u-a)[T_1(w-c) + T_3(u-a)]\}$$

where

$$K = -1/[2d(H_1,H_2)]$$

With this calculation of the position of H_2, the positions of all hydrogens have been determined.