

GenStar: A method for de novo drug design

Sergio H. Rotstein and Mark A. Murcko*

Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, MA 02139, U.S.A.

Received 5 August 1992

Accepted 14 October 1992

Key words: Drug design; Protein structure; Drug–ligand interactions; HIV protease inhibitors; Carbonic anhydrase inhibitors; Ligand design; Protein active site; Enzyme inhibitors; FKBP-12 inhibitors

SUMMARY

A novel method, which we call GenStar, has been developed to suggest chemically reasonable structures which fill the active sites of enzymes. The proposed molecules provide good steric contact with the enzyme and exist in low-energy conformations. These structures are composed entirely of sp^3 carbons which are grown sequentially, but which can also branch or form rings. User-selected enzyme seed atoms may be used to determine the area in which structure generation begins. Alternatively, GenStar may begin with a pre-docked ‘inhibitor core’ from which atoms are grown. For each new atom generated by the program, several hundred candidate positions representing a range of reasonable bond lengths, bond angles, and torsion angles are considered. Each of these candidates is scored, based on a simple enzyme contact model. The selected position is chosen at random from among the highest scoring cases. Duplicate structures may be removed using a variety of criteria. The compounds may be energy minimized and displayed using standard modeling programs. Also, it is possible to analyze the collection of all structures created by GenStar and locate binding motifs for common fragments such as benzene and naphthalene. Tests of the method using HIV protease, FK506 binding protein (FKBP-12) and human carbonic anhydrase (HCA-II) demonstrated that structures similar to known potent inhibitors may be generated with GenStar.

INTRODUCTION

Computer-assisted drug design is a rapidly growing field [1–5]. Although there is still no drug which can unambiguously be said to have been ‘computer designed’, there are clear examples in the recent literature of progress in that direction [6–8].

The past few years also have witnessed exciting advances in the techniques for the determination of macromolecular structures [9,10]. At the same time, homology based methods for the prediction of protein structure are improving [11–14]. In the near future it will be common for the investigator to have a high-resolution structure of the target enzyme, or at least a quite reasonable model derived from homology. Given this outlook, it is desirable that computational tools be

*To whom correspondence should be addressed.

developed which can analyze protein active sites and suggest compounds which may bind to these sites. Ideally, these methods would be graphical, interactive, fast, and produce diverse families of chemically and biologically reasonable structures. Such ‘rational de novo drug design’ methods do not yet exist.

It may be argued that a prerequisite for the development of de novo drug design methodology is a quantitative description of the interactions between a drug and its biological receptor. This is a continuing area of important research in theoretical chemistry [15,16]. The factors which contribute to binding include hydrogen bonding [17–21], the hydrophobic effect [22,23], van der Waals and dispersion interactions [24–26], solvation effects [27–29] and other electrostatic interactions [30–34].

Methods have been developed for analyzing specific types of interactions such as hydrogen bonding. Goodford et al. have developed GRID, which places small fragment probes at many regularly spaced grid points within the active site, determining the most favorable scores. It has been tested on a variety of enzyme–inhibitor complexes and found to reproduce the positions of important hydrogen bonding groups [35,36]. A related program is HSITE, which generates a map of the hydrogen-bonding regions of an enzyme active site, including the probability of hydrogen-bond formation at each point [37,38].

There are other ways to determine favorable locations of fragments. Miranker and Karplus have modified CHARMM [39] so that molecular dynamics simulations may be performed without nonbonding interactions between solvent molecules. This allows the solvent molecules to overlap each other in energetically favorable regions and greatly improves the efficiency with which such locations may be identified [40]. A variety of simulated annealing approaches have been described recently which are well suited for determining the location of molecular fragments [41–43]. Lee and Rose have described a method for determining the surface topology of a protein, looking for prominent features such as ‘bumps’ and ‘pits’. These can represent, for example, hydrophobic pockets which might be filled in order to maximize drug binding [44]. Lesk has described a technique called ‘molecular speleology’ to explore clefts in proteins and suggest candidate ligands using mathematical techniques borrowed from X-ray crystallography [45].

One of the difficulties with any fragment-based method is that it is a challenging problem to design a drug candidate that properly orients each of the fragments necessary for tight binding, and is synthetically tractable and pharmacologically reasonable. Several preliminary methods to accomplish this have been reported. Lewis and Dean have developed methods for using ‘spacer skeletons’ of appropriate size to match ligand atoms to the correct binding sites in enzyme active sites [46,47]. Since then, Lewis has proposed the use of a ‘diamond lattice’ to determine favorable ways of spanning the distance between distant regions of an active site [48]. Lewis and co-workers have developed an elegant, efficient method that treats the atoms in the structures selected by DOCK as an *irregular lattice* which can be used to connect distant atoms and/or fragments in chemically novel ways [49]. One of the most advanced methodologies in the area is CAVEAT, developed by Bartlett and co-workers [50]. CAVEAT uses databases of cyclic compounds which can act as ‘spacers’ to connect any number of fragments already positioned properly in the active site. Such tools allow the modeler or chemist to quickly generate hundreds of possible ways to connect the fragments already known or suspected to be necessary for tight binding [51]. Several other research groups have pioneered the use of 3D databases for drug design, especially searching for molecules which match a predicted pharmacophore. For instance, Van Drie et al. [52] and

Martin [53] have developed ALADDIN, while the group at Lederle has produced 3DSEARCH [54].

Several different methods have been developed for analyzing known active-site structures and suggesting ligands with complementary steric and electrostatic properties. The best known of these is DOCK which uses a fast sphere-matching algorithm to dock compounds from a user-supplied database in an enzyme active site [55]. In one application of the method, each compound was considered to be composed of several fragments which were docked separately, allowing some flexibility, although this capability does not appear to be generally available [56]. Another enhancement includes the consideration of electrostatics [57]. DOCK was used to select screening candidates from a large commercial database for possible HIV protease inhibitors. Enzymatic assays subsequently determined very weak HIV protease inhibitory properties in the well-known antipsychotic haloperidol [58], although there is as yet no published crystal structure to support the structural binding hypothesis put forth by the program.

Two recently reported programs, CLIX [59] and LUDI [60], represent significant advances in fragment-based drug design. These programs have the same basic approach to drug design as that found in CAVEAT [50,51] and also the work by Lewis and Dean [46,47] and Lewis et al. [49,49]. CLIX uses the output from GRID calculations [35,36], carried out with a variety of probes, to characterize the receptor site in terms of an ensemble of favorable binding positions for different groups or 'fragments'. This information is then used to query a chemical database for candidate molecules which have good coincidence of individual fragments with members of the ensemble. CLIX uses information from the GRID potential maps to suggest possible changes in the structures pulled out of the Cambridge Structural Database [61,62] to improve their binding. As a test case, sialic acid was found to bind well to a mutant influenza-virus hemagglutinin structure, in good agreement with available structural information. LUDI also accepts the output from GRID, but in addition the program can determine a list of *interaction sites* into which to place both hydrogen bonding and hydrophobic fragments. Both a rule-based approach and a statistical contact pattern derived from the Cambridge Structural Database may be used for this purpose. LUDI then uses a library of ~ 600 linkers to connect up to four different interaction sites into *fragments*. Then, smaller 'bridging' fragments such as CH₂ and -COO- are used to connect these fragments. For the enzyme DHFR, the placements of key functional groups in the well-known inhibitor methotrexate were reproduced by LUDI. For trypsin, the rule-based approach to fragment generation failed to reproduce the known conformation of benzamidine; however, the statistical contact pattern method did place this fragment in its proper orientation.

Nishibata and Itai have developed LEGEND [63], which builds a structure sequentially from randomly selected atom types which are positioned with random torsion angles. A candidate atom is selected automatically if it is not bumping either the enzyme or any previous atoms in the growing drug molecule. After a structure is complete, charges are assigned to all atoms and the structure is energy minimized in the active site. From the many structures generated by LEGEND, a separate post-processing program, LORE, may be used to select the more interesting structures for graphical analysis.

Finally, two groups have recently published exciting methods for determining the correct orientation of peptide substrates in enzyme active sites. Unlike all the methods discussed previously, these methods avoid the difficult problem of connecting isolated fragments by using build-up procedures that linearly connect each fragment to the preceding one. The tradeoff is that a much more limited set of fragments may be considered. Singh et al. have developed GEMINI

[64], which uses information on the packing of amino acid side chains from a database of crystallographically determined structures to suggest conformations of peptide ligands to their receptors. It was demonstrated that GEMINI can reproduce the crystallographic orientation of various peptides bound to endothiapepsin, carboxypeptidase A and thermolysin. Moon and Howe have described GROW [65], which uses a build-up procedure to determine the best peptidal inhibitor or substrate for a given enzyme. A large predefined library of conformations of each amino acid is used in the construction process. Each conformation of each residue is tested according to a molecular mechanics forcefield, and the set of N-lowest energy possibilities is carried along to the next step. Significantly, both conformational (intramolecular) enthalpies and solvation free energies are included in the analysis done by GROW. Trial studies with the aspartyl protease rhizopuspepsin were quite successful at reproducing the conformation of a reduced peptide inhibitor whose structure has been determined crystallographically.

In this paper we describe the GenStar method. GenStar generates enzyme inhibitors by growing one sp^3 carbon atom at a time. We have tested our method against three well-characterized enzymes, FKBP-12, HIV protease and human carbonic anhydrase. As our test cases show, this approach suggests novel chemical structures that fill active site cavities, and is a useful tool in collaborative drug design. We also discuss the strengths and weaknesses of the technique and our current efforts to develop more advanced methods.

METHOD

The GenStar approach may be described in several steps. These are: (1) data preparation and program input; (2) representation of the active site region; (3) determination of candidate positions for inhibitor atoms, including the special case in which a ‘core fragment’ is used to start the process; (4) scoring candidates according to ‘good steric fit’; (5) selection of an inhibitor atom among the candidates; (6) handling of two special cases, chain branching and ring closure; (7) criteria for termination; (8) structure output; and (9) post-processing options.

(1) Data preparation and program input

The user first prepares a PDB file containing the non-hydrogen atoms of the enzyme of interest. Next the user determines the Cartesian coordinates of the upper and lower bounds of the active site space. The enzyme and the active site boundaries are used by a pre-processing program to generate a *closeness grid* (described below). In addition, the user must prepare a list of the enzyme active site atoms that can be used as ‘seed’ positions for initiating structure generation. In the case where structure generation is to begin from a ‘core’ inhibitor, the user must supply a PDB file containing the core structure (with the hydrogens removed). This file is fed into a pre-processing program which generates a *corefile* which can be interpreted by GenStar. When executing GenStar, the user also specifies the number of inhibitors to generate from each seed position as well as the maximum number of atoms in the structure. In addition, the user may request that inhibitor atoms not extend beyond a certain distance from the starting seed position or core inhibitor structure; this is called the *spatial extents check*. Finally, the user may also specify that only structures in which one or more rings are generated should be kept. All of these variables specified at run-time are listed and described in Table 1. There are a number of other constants

that control program execution that are stored in a separate user-editable file. These are listed in Table 2, and are described in the following sections.

(2) Representation of the active site region

The goal of the algorithm is to place atoms wherever they will have the best interaction with the enzyme. One way to place atoms in the active site would be to use a grid representation that subdivides the space into equidistant gridpoints. However, such a limitation is artificial and unnecessarily restrictive. For this reason we have decided to allow atoms to be generated at any Cartesian coordinates. However, the advantage of using a grid to divide space is the fact that each gridpoint can contain information about the space around it – for example, a list of the neighboring enzyme atoms. Such knowledge about the local space can greatly reduce the need for pointless computation. For example, when we are evaluating a candidate atom, we need to perform some basic checks to avoid errors such as intramolecular bumping. Local space knowledge allows us to limit our distance checking to the atoms that are close to the candidate. To this effect we have implemented what we call the *closeness grid*. This is a rectangular grid in which each gridpoint contains a list of all the enzyme atoms that are up to 5 Å away from it. Therefore, to find out if a candidate is bumping the enzyme we determine which gridpoints are surrounding that candidate, and measure the distance between the candidate and the enzyme atoms specified in those gridpoints. Typically, the use of a closeness grid increases performance more than tenfold. Currently the gridpoint spacing is set to 1 Å (see Table 2). Since atom placement is not restricted to gridpoints, there is no need for finer spacing.

(3) Determination of candidate positions for inhibitor atoms

The growth of the first three atoms of an inhibitor are handled by special cases, while the fourth and all subsequent atoms are handled in the general case.

First atom

The user requests that generation starts in the vicinity of a certain enzyme atom, which we refer to as the ‘seed’ atom. The user may specify any number of seed atoms. For this case we generate

TABLE 1
INPUT REQUIRED FROM USER AT PROGRAM RUNTIME

When starting from seed atoms	When starting from a core
Name of input PDB file	Name of input PDB file
Name of closeness grid file	Name of closeness grid file
Basename of output compound files	Name of file containing structure core
Number of enzyme atoms to use as seeds	Number of compounds to generate
Atom numbers of enzyme seed atoms	Option to discard structures without rings
Number of compounds for each seed atom	Max. number of atoms per compound
Option to discard structures without rings	Number of branch points from core
Max. distance to grow from starting point	Atom numbers of core branch points
Max. number of atoms per compound	Max. distance to grow from branch points

TABLE 2
KEY PARAMETERS

<i>Closeness grid parameters</i>				
GRIDPOINT SPACING	1.0 Å			
MAX SIZE ON X	30 Å			
MAX SIZE ON Y	30 Å			
MAX SIZE ON Z	30 Å			
<i>Distance and scoring constraints</i>				
INH_INH_DIST	2.25 Å	Minimum intra-molecular distance		
MIN_HEAVY_ATOM_DIST	3.0 Å	Minimum for heavy atoms (except N and O)		
MIN_O_OR_N_DIST	2.8 Å	Minimum for O and N atoms		
MAX_H_BOND_DIST	3.5 Å	Maximum distance for a hydrogen bond		
MAX_CONTACT_DIST	5.0 Å	Maximum distance for enzyme contact		
HBOND_SCORE_INCREMENT	1.0	Additional score for good hydrogen bonding contact		
MIN_RING_SCORE	5.0	Minimum acceptable score during ring closure		
<i>Default constraints for bond lengths, bond angles and torsion angles</i>				
MINBL	1.48 Å	Minimum bond length		
MAXBL	1.60 Å	Maximum bond length		
DELTABL	0.06 Å	Bond length increment		
MINDEG	106°	Minimum bond angle		
MAXDEG	118°	Maximum bond angle		
DELTADEG	4°	Bond angle increment		
MINTOR1	55°	Minimum torsion for low-energy conformation 1		
MAXTOR1	80°	Maximum torsion for low-energy conformation 1		
MINTOR2	160°	Minimum torsion for low-energy conformation 2		
MAXTOR2	200°	Maximum torsion for low-energy conformation 2		
MINTOR3	280°	Minimum torsion for low-energy conformation 3		
MAXTOR3	305°	Maximum torsion for low-energy conformation 3		
DELTATOR	5°	Torsion increment		
MIN_SHELL_ATOM1	3.0 Å	Minimum radius of the shell used to position the first atom		
MAX_SHELL_ATOM1	5.0 Å	Maximum radius of the shell used to position the first atom		
ANGLE_DELTA	10.0°	Rotational increment used to build the spheres of candidate points for the first two atoms, and also used for building the torus used for the third atom and the general case		
<i>Definition of ring closure constraints</i>				
MAX-RING	7	Maximum number of atoms allowed in a ring		
<hr/>				
Flexibility variable	Tight	Medium	Loose	Explanation
LFLEX	0.00 Å	0.04 Å	0.20 Å	Additional bond length flexibility
AFLEX	0.00°	10.0°	25.0°	Additional bond angle flexibility
TFLEX	0.00°	10.0°	60.0°	Additional torsion angle flexibility

a spherical shell around the seed atom. The minimum and maximum radii of the shell are currently 3 Å and 5 Å, respectively, but may be changed by the user (Table 2). This shell consists

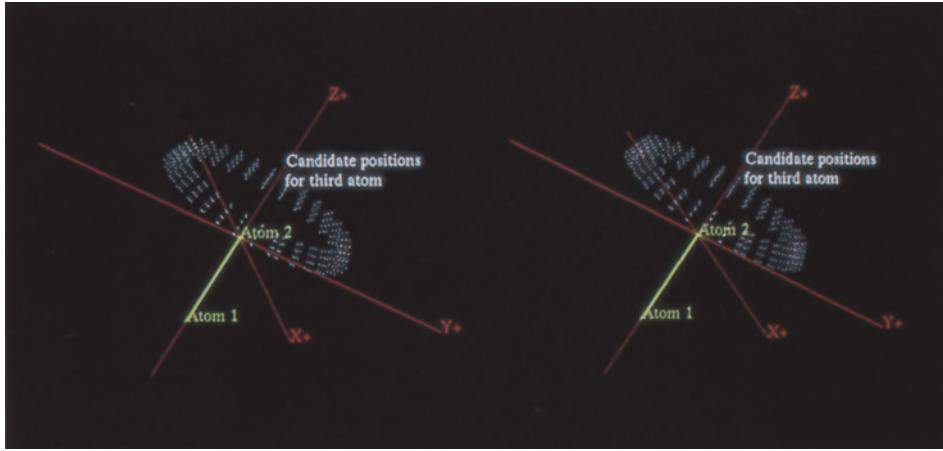


Fig. 1. Torus for positioning the third atom.

of ~ 400 points which make up the set of candidate positions for the location of the first inhibitor atom. These positions are scored according to the algorithm described below, and one is chosen randomly from among the set of candidates with the highest scores. If no acceptable candidate is found, generation is aborted. Usually this means that the seed atom is in a location such that all the candidates either bump into the enzyme or lie outside of the active site as previously defined.

Second atom

To generate the second inhibitor atom the only constraint is bond length since for a two-atom system both and torsion angles are irrelevant. A shell is constructed around the first atom, and the minimum and maximum radii of the shell are defined as the standard lower and upper bounds for bond lengths, listed in Table 2. Each of ~ 450 candidates is scored and the selection is made as before.

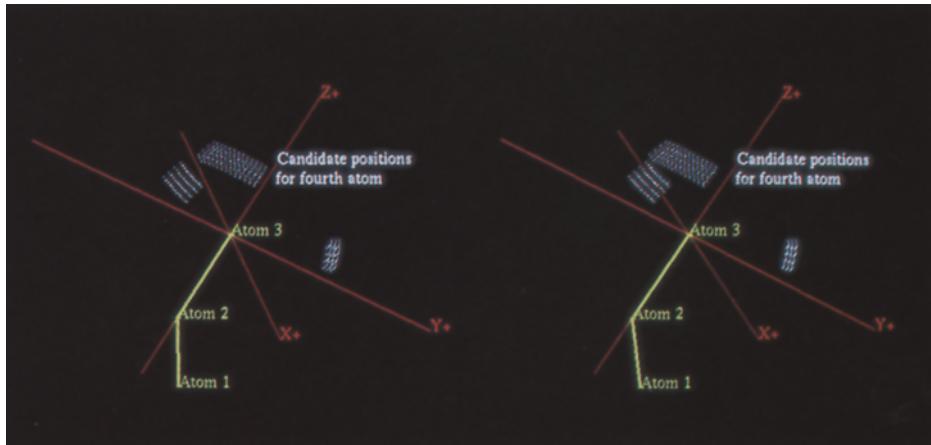


Fig. 2. Interrupted torus for positioning the fourth and subsequent atoms.

Third atom

When choosing the third atom we must ensure that the candidates satisfy the constraints of bond length and bond angle. A polar coordinate system is used to generate the candidate atom positions that obey the specified constraints. This operation is performed during the initialization phase of GenStar, and the candidate positions which form a torus-shaped template are stored (Fig. 1). After that, when generating the third atom of an inhibitor, we only need to move the torus to the coordinate frame of the first two atoms. This avoids the repeated generation of candidate positions, which is computationally demanding due to the use of innately slow trigonometric functions. With the default parameters ~ 400 candidates are considered.

Additional atoms

At this point there are at least three atoms in the inhibitor structure and all candidates must therefore respect the constraints of torsion angle as well as bond length and bond angle. Some values of torsion angles result in very-high-energy conformations and are therefore considered undesirable. During the initialization phase of the program we generate a toroidal volume of possible locations for the new atom. This torus is similar to the one created for the special case of the third atom, as described above, but this time we rule out torsions that would result in high-energy conformations. The result is a template which has three interruptions in its circumference (Fig. 2). The ranges of acceptable torsion values are listed in Table 2. The ranges that we currently use correspond to low-energy rotational states for normal hydrocarbons [66] and produce ~ 350 candidates. Whenever a new atom must be generated, we simply move the template to the correct coordinate frame.

Generating structures from a 'core' inhibitor

Alternatively, the user may supply a core inhibitor that has previously been docked into the active site. This option allows the user to begin with a structure or fragment which is known (or hypothesized) to bind to the active site in a specific orientation. The user loads a core structure and then selects the atom(s) from which inhibitor generation should begin. Generation immediately starts from the specified atoms in the usual manner. The user can place a limit on the maximum distance from the starting atom(s).

(4) Determination of enzyme contacts and scoring

After ~ 350 candidate atoms are selected using the methods described above, each is scored. Our scoring algorithm is straightforward. First we ensure that the candidate atom lies within the boundaries of the active site. Next, if the spatial extent option was selected by the user, we ensure that each candidate is within the user-specified distance of either the seed enzyme atom or the core branch point, whichever is relevant. Then we check to see whether the candidate is bumping into any other atom already in the currently growing inhibitor structure. Such intramolecular contacts disqualify a candidate except in the special case of ring closure discussed below. At this point we begin the scoring process. We determine the eight gridpoints from the closeness grid that form a cube in space surrounding the candidate. From these gridpoints we can extract the list of enzyme atoms which are close to the candidate inhibitor atom. Using this list we ensure that the candidate is not bumping into any enzyme atom. Next, we simply add 1.0 for each enzyme atom which lies

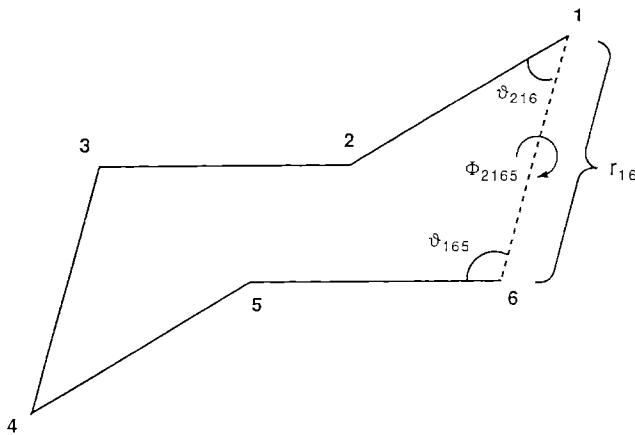


Fig. 3. The internal coordinates measured when determining whether or not to close a ring.

between $X \text{ \AA}$ and 5 \AA from the candidate atom. When the enzyme atom in question is a carbon or sulfur, $X = 3.0 \text{ \AA}$; when the enzyme atom is nitrogen or oxygen, we set $X = 2.8 \text{ \AA}$ to account for the smaller radius of N and O atoms. A similar scheme has been used by Yamashita et al. to determine the binding sites of metal ions on protein surfaces [67]. To this score we add an additional 1.0 if the enzyme atom is N or O and the separation is less than 3.5 \AA . This provides a slight preference for inhibitor atoms that are well situated to form hydrogen bonds to the enzyme. To increase the importance of H-bonding this value may be adjusted by the user.

(5) Selection of an inhibitor atom

Of all the hundreds of candidate atoms, we determine the one with the highest score. Then, we

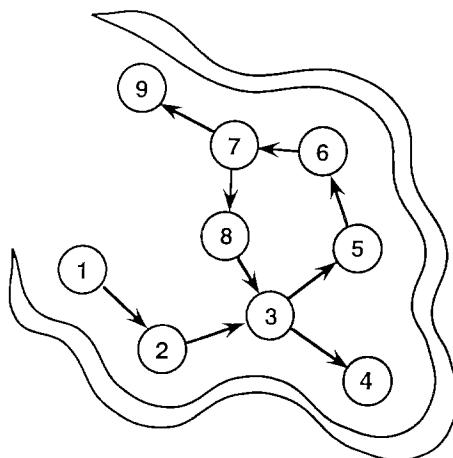


Fig. 4. The direction of the atom-by-atom build-up process used by GenStar. Atoms 3 and 7 are branch points, while atom 8 closes the 5-membered ring.

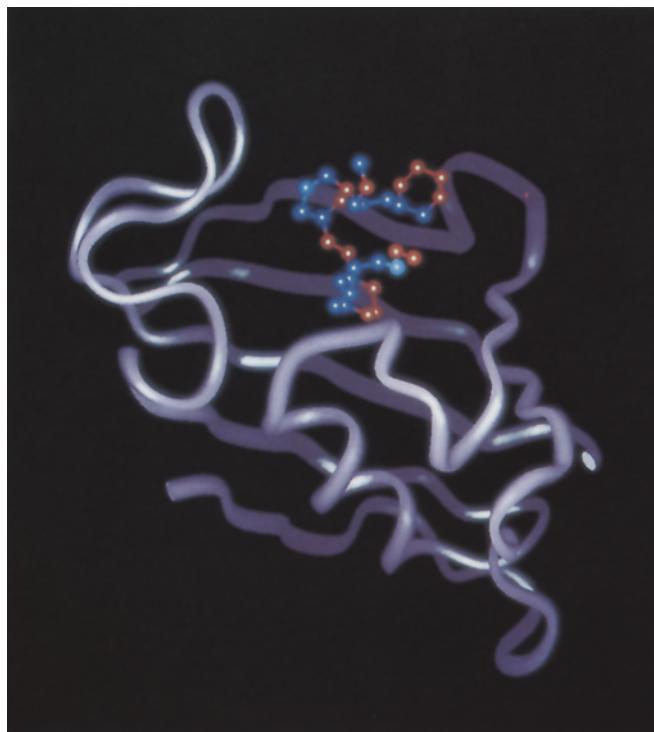


Fig. 5. Color coding the surface of an inhibitor using the electrostatic potential (ESP) of the enzyme. Red represents regions of positive ESP, indicating that electronegative atoms should be added, while blue represents regions of negative ESP, indicating that electropositive atoms should be added.

find all other scores within 20% of the highest score. Typically, out of ~ 350 candidates, there are 10–20 that fall into this category. Then, we randomly select one atom from among these high-scoring candidates. The reasons for introducing randomness to the process are to compensate for the approximations in the scoring algorithm and to diversify the shapes of the structures generated.

(6) Special cases: branching and ring closure

Sometimes, the scoring algorithm will determine that there are candidates from different conformational families (e.g., trans and gauche $+$) that both have very good scores, suggesting that it may be worth exploring space in more than one direction. In this case, the program adds the current atom to a list of ‘branch points’ before a candidate is selected. The existence of these branch points allows for a much more realistic generation of structures, since real chemical compounds are seldom linear. Every time the structure being generated reaches a dead end, meaning that adding another atom would either bump into the enzyme or extent beyond the active site grid, the program selects the next unexplored branch point remaining in the list.

Occasionally, a candidate atom is appropriately situated to close a ring. Although the method implicitly allows for rings of any size to be formed (from cyclopentane to any macrocycle) we have purposely limited the size of rings to seven atoms (Table 2). Removing the limitations on ring sizes

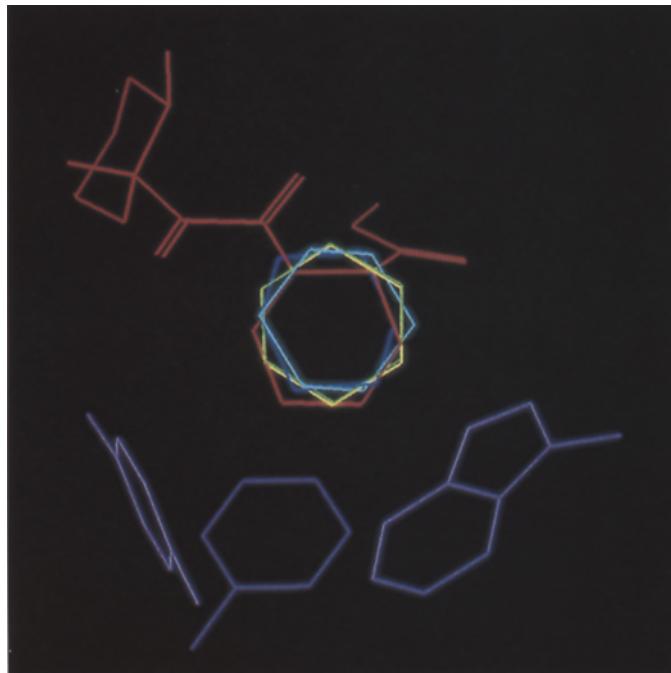


Fig. 6. Cyclohexane rings which closely match the orientation of the pipecolic ring found in the FKBP-12 inhibitors, FK506 and rapacycin. Rings have been energy minimized in the active site.

tended to produce over-complicated structures. The program uses a set of ring closure tolerances which determine whether the proposed ring has acceptable internal coordinates (Fig. 3). These tolerances may be different from the normal ranges of bond length, angle, and torsion. For convenience, three sets of tolerances called *loose*, *medium*, and *tight* are predefined (Table 2). *Loose* criteria allow more variation in the internal coordinates and lead to the identification of more rings by the program. *Tight* criteria use more stringent rules and produce fewer rings. Our tests have shown that tighter criteria will tend to produce larger rings (such as cycloheptane), whereas looser criteria allow for the creation of smaller ones such as cyclopentane. If a proposed ring meets the user-specified tolerances, then the candidate atom is scored in the usual way. If the score of the atom is above a threshold value, the program will choose that atom, and form the ring, rather than any other candidates that may have higher scores. This is allowed since rings tend to add beneficial structural rigidity and it therefore seems reasonable to give them some degree of priority.

To make the concepts of branching and ring closure more clear, Fig. 4 demonstrates a sample build-up procedure. Starting with atom 1 the process continues as shown in the directions of the arrows. Atoms 4 and 5 both make good enzyme contacts and have high scores, so atom 3 is marked as a branch point. Build-up continues with atom 4, but it is not possible to go any further, so growth then transfers to atom 5. Atoms 6 and 7 again make good enzyme contacts. Atom 8 can close the 5-membered ring, so it is selected, and atom 7 is automatically declared a branch point. Nothing can be added to atom 8, so growth then continues with atom 9.

(7) Termination criteria

An inhibitor is ‘finished’ when any of several conditions is met. Either: (a) a user-specified number of atoms has been reached or (b) the inhibitor has reached a dead end, i.e., no more atoms can be added and all branch points have been explored. If there are at least four atoms in the structure, the inhibitor is saved. The user can specify the number of structures to generate. On an SGI Indigo workstation the generation of a hundred structures, each containing approximately 50 atoms, typically takes about ten minutes.

(8) Structure output

Each structure generated by GenStar is written out in PDB format, one file per structure. The file names all have a unique beginning specified by the user. The B-factor column is used to record the score of the atom. HEADER and REMARK records are used to store information about the run such as the user name, directory path, date and time, enzyme name, active site grid dimensions, seed atoms, core structure used, and other parameters.

(9) Post-processing

Since the program can generate many structures quickly, it is desirable to have ways of reducing the number of structures that the user must examine and to minimize the effort required. We have created a series of post-processing tools, including options for removing unwanted structures, structure clean-up, visualization, structure modification, and several kinds of ‘pattern recognition’. These options are essential for allowing the synthetic chemists to efficiently review the structures.

Removing duplicate structures

Approximately duplicate structures may be eliminated based on differences in atomic positions. Also, the user can specify enzyme atoms or residues which must be in contact with the inhibitor structure. For example, a user may only be interested in inhibitors which are in contact with a particular ‘greasy pocket’ binding site in the enzyme. Structures which lack such contacts may be eliminated. The user can control the required proximity.

Structure cleanup

As it generates inhibitors, GenStar also builds an InsightII [68] script for (a) reading in each inhibitor; (b) growing hydrogens; (c) fixing potentials and charges; and (d) energy minimizing with the enzyme fixed. The minimization removes poor inter- and intramolecular contacts and ensures that the docked conformation of the proposed structure is fully relaxed.

Visualization

Using a pre-defined grid representation of the electrostatic potential (ESP) of the active site we can color a Connolly surface [69] of each inhibitor to show the electrostatic properties of the enzyme in the vicinity of each inhibitor atom. We can use the same information to color code each atom in a stick figure. Examples are shown in Fig. 5.

Structure modification

Also using the ESP grid for guidance, we have written InsightII script files which automatically modify the hydrocarbon structures that GenStar produces to place heteroatoms in regions where the opportunity exists for strong, favorable electrostatic interactions with the enzyme (such as hydrogen bonds).

Pattern recognition

When taken together, all the structures generated during a GenStar run fill the majority of the active site. Lewis et al. have described such a collection of atoms as an *irregular lattice* which can be used to find paths to connect distant fragments in an active site [49]. We carry out two kinds of searching on the irregular lattice generated from a collection of GenStar structures:

(a) *High-density pathways.* The active site is filled with a diamond lattice of 1.54 Å spacing and superimposes on the irregular lattice of GenStar-derived structures. At each grid point in the diamond lattice we count the number of atoms from the irregular lattice that are within 0.8 Å. We then write out a PDB file containing the coordinates of all points in the diamond lattice which are ‘visited’ by at least 10% of the structures generated. Next, we identify the point on the diamond lattice with the highest score, and trace the path of highest density through the diamond lattice, and write those coordinates to a separate PDB file. These two PDB files are then displayed in the context of the enzyme active site.

(b) *Fragment recognition.* Using straightforward graph theory we locate benzene, naphthylene, cyclohexane, cyclopentane, and acetone ‘fragments’ which may be formed by combining atoms from the irregular lattice [70]. Because *every* atom in our lattice makes good contact with the enzyme, the fragments generated in this manner should, in principle, provide significant binding energy. These fragments may then be displayed to give the chemist another set of suggestions for synthetic starting points.

RESULTS

We envision three principal uses for GenStar: (1) completely de novo fragment building and drug design, (2) modification of known drugs using a core structure, and (3) an iterative build-and-choose process, possibly starting from a small core. We have tested GenStar in each kind of situation. As test systems for each type of situation, we have considered FK506 Binding Protein (FKBP-12), HIV-protease and Human Carbonic Anhydrase type II (HCA-II). Several crystal structures of enzyme–inhibitor complexes are now available for each of these systems, and we wished to see whether GenStar would suggest candidates reminiscent of known drugs, bound in approximately the same way.

(1) *De novo design: FKBP-12*

FK506 and rapamycin are both immunosuppressive, macrocyclic natural products which inhibit the enzyme FKBP-12. Both FK506 and rapamycin contain a 6-membered piperidine ring in essentially the same location, at the bottom of a lipophilic active site pocket [71–74]. Using all the indole ring atoms from Trp⁵⁹ as seeds, and medium ring closure constraints (Table 2), a number of suggested positions for 6-membered rings were generated. After simple energy minimi-

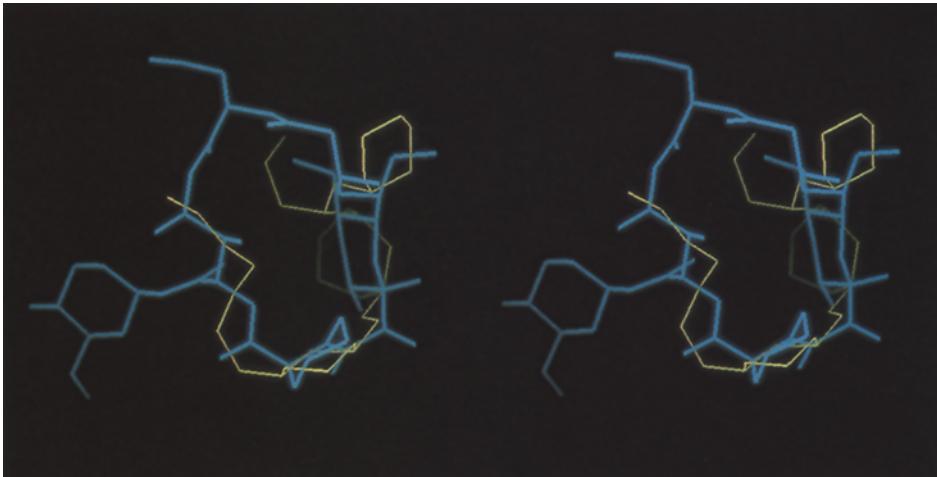


Fig. 7. Energy-minimized structure (in yellow) produced by GenStar which mimics the structure of FK506 (light blue) bound into FKBP-12.

zation with CHARMM [75] many of these ring systems migrated into a nearly exact match ($\text{RMSD} \sim 0.4 \text{ \AA}$) to the position seen in FK506 [72] and rapamycin [74] (Fig. 6). Also, many of the de novo structures generated from these seed atoms have approximately the same topology as

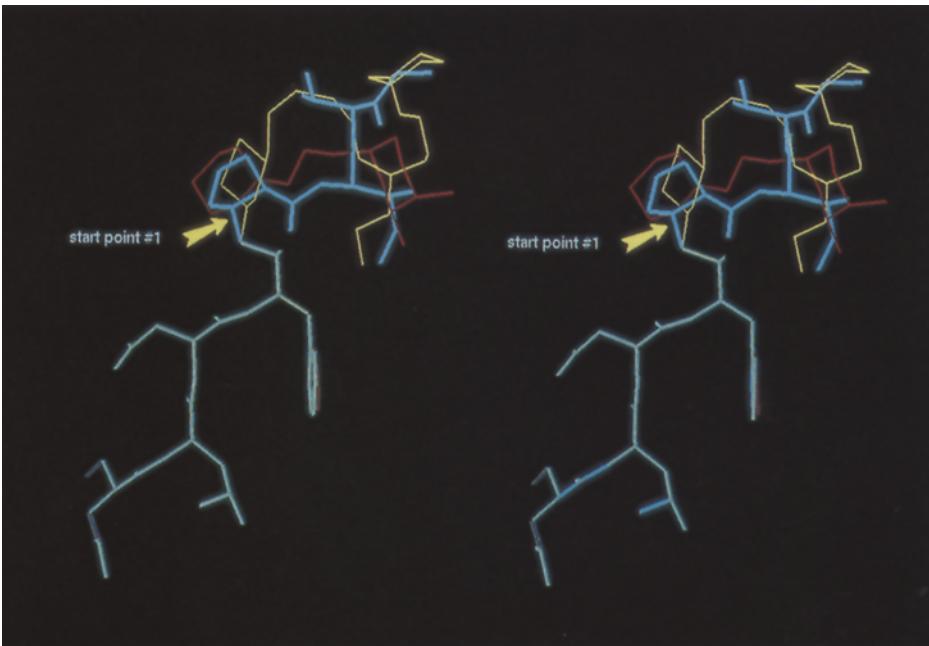


Fig. 8. Structures produced by GenStar (in yellow and red) which closely mimic the HIV-protease inhibitor JG-365 (light blue). An arrow points at the start points ('grow point') used to begin structure generation from the core. The structures have been energy minimized in the active site, with the 'core' portions constrained.

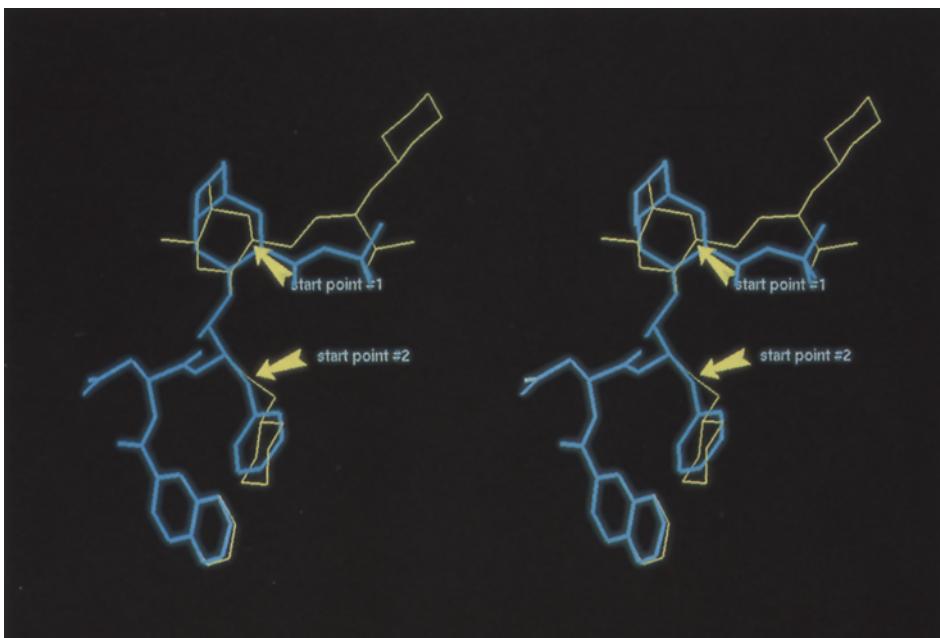


Fig. 9. GenStar structure (in yellow) which closely mimics the HIV-protease inhibitor Ro 31-8959 (light blue). Arrows point at the start points ('grow points') used to begin structure generation from the core. The structure has been energy minimized in the active site, with the 'core' portion constrained.

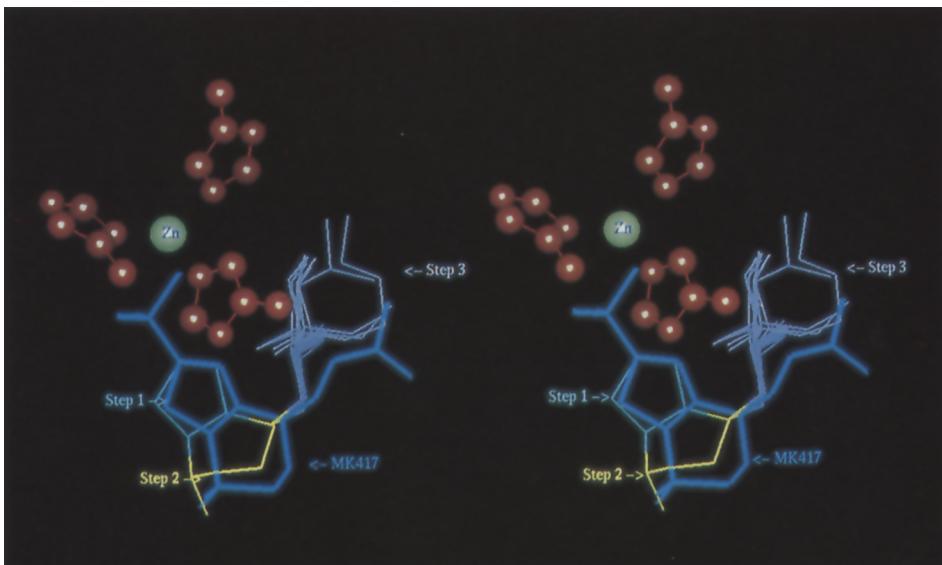


Fig. 10. GenStar structure developed in a two-step iterative process which closely mimics the known carbonic anhydrase inhibitor MK-417 (heavy blue lines). The Zn^{2+} ion and the three histidine residues that bind to it are shown as colored spheres. The coordinates of MK-417 were modeled using the information in Ref. 7. See text for details.

FK506 and rapamycin in the active site. A representative example is shown in Fig. 7. Clearly, GenStar is providing a reasonable suggestion for a novel class of compounds which closely mimics the known, sub-nanomolar macrocyclic inhibitors.

(2) Modification of known inhibitors: HIV protease

As a test with the HIV protease, we docked a ‘core’ inhibitor (structure **1**) composed of the P4 to P1’ nitrogen of the inhibitor JG-365 developed by Swain et al. and subsequently studied crystallographically [76]. Several interesting ‘hits’ from a single 50-structure run are shown in Fig. 8. Medium ring closure constraints were used (Table 2). For comparison, the actual X-ray coordinates of JG-365 are shown in light blue. Clearly a ring in the P1’ position has been generated, and the general shape of the known inhibitor has been preserved. Along the same lines, we docked the P3–P1’ core of Ro 31-8959, the Roche lead compound [77,78] (structure **2**). Again, compounds which closely mimic the known inhibitor are discovered, including the unusual P1’ 6-membered ring, as demonstrated in Fig. 9.

(3) Iterative design: carbonic anhydrase

It is well known that aromatic sulfonamides are excellent inhibitors of human carbonic anhydrase (HCA-II) [79]. The sulfonamide group has been shown to bind to the catalytic Zn²⁺ ion [80,81]. We began with methane sulfonamide docked in the crystallographically determined orientation [7,80,81] as our core (Fig. 10). Then, as a first step, we grew 10 additional atoms from the methyl to determine whether GenStar would identify a large percentage of cyclic structures. Throughout this process, medium ring closure constraints were used (Table 2). This run generated several inhibitors containing such rings, in all cases 5-membered rings which included the methyl carbon. After minimization, all of these cyclic compounds essentially converged to a single structure, which was then used as the core for the second round of structure generation. Here, we allowed new atoms to grow from any atom in the newly-formed ring, but again limited the number of new atoms to 10. This run resulted in several structures which formed a second ring, shown in yellow in Fig. 10. After minimizing this [5,7]-bicyclic structure, we began a third round, growing a maximum of ten atoms off the second ring. A high percentage of the structures generated placed acyclic alkyl side chains in a particular region of the active site. Several representative compounds are shown in Fig. 10, with the new atoms highlighted in purple. For comparison, the structure of MK-417, a potent HCA-II inhibitor that has been shown to lower intraocular pressure in man, is shown in blue. The structure of MK-417 is very closely related to another HCA-II inhibitor, MK-507, currently in Phase III clinical trials [82] for the treatment for glaucoma. Even though GenStar is currently incapable of directly producing an aromatic thienothiopyran ring system, the overall shape of the well known inhibitor is clearly suggested. Other, more novel structural classes of HCA-II inhibitors were suggested; these will be discussed elsewhere [83].

DISCUSSION

The results of the test runs with our three model systems demonstrate that the GenStar

approach provides interesting, useful suggestions for inhibitor design. Completely de novo structures may be created, or specific subsites (such as the P1' pocket in HIV protease) may be filled. Pattern matching, carried out on the ‘irregular lattice’ of atom positions derived from the union of all generated structures, provides additional useful suggestions for novel compounds.

We intended with this method to begin to explore a build-up approach to de novo drug design. Usually the structures that are generated by the program are synthetically difficult. For example, some contain many stereocenters. However, this has been less of an issue than we anticipated. The method has been effective in fighting against a kind of ‘writer’s block’ of molecular modelers, by suggesting new and unexpected ways to fill the active site, rather than attempting to generate compounds that can be readily synthesized. The usefulness of the method lies in providing ‘raw material’ for further collaborative modeling studies between synthetic and computational chemists.

We still consider it a de novo design method, since this raw material is generated from no more data than the crystal structure of a protein. It is worth remembering that successful collaboration depends on a mixture of visualization, synthetic analysis, chemical intuition and novel methodology such as GenStar, and we expect that this will continue to be essential no matter how advanced the techniques may become. In part, this is because tight binding to an enzyme is a necessary but not sufficient condition for a successful drug, and GenStar (like all de novo drug design programs) is currently limited solely to considerations of drug binding.

We believe that the success of the program results from a combination of factors. First is the requirement that each inhibitor atom must make good contact with the enzyme. Second is the allowance for branching and ring closure which helps to provide more realistic structures. Third is the allowance of a range of bond lengths, bond angles, and torsion angles; as more atoms are grown in each structure, the cumulative effect is to allow a more thorough exploration of the active site than would be possible with a fixed set of bond lengths, angles and torsions. Finally, and perhaps most important, is the inclusion of randomness to the atom selection process. This helps to satisfy the desire that novel structures be generated, and prevents us from depending too strongly on the scoring algorithm.

For de novo drug design, this tool is quite different than either DOCK [55] or GRID [35], two other programs widely used to suggest novel compounds. GenStar only uses sp^3 carbons, making it much less ‘thorough’ than GRID, which has many different types of atomic probes. On the other hand, the atoms are positioned without the restriction of fixed grid spacing, and the suggested atoms are liked together in chemically interesting ways, including rings and branching, as *molecules*. This offers more immediate feedback to the medicinal chemist. We consider the techniques complementary in this sense. DOCK, like GenStar, suggests ‘whole structures’ to the chemist, but the structures are limited to those stored in an available database, and the method is still rather CPU intensive. DOCK also cannot easily be used to modify an existing drug. Again, we consider our method to be complementary. This is not an ‘either-or’ situation, as we believe that a variety of drug design tools are necessary to fully meet the needs of the practicing drug discovery team.

The interesting method of Nishibata and Itai, LEGEND [63], is similar in some ways to GenStar. Individual atoms are sequentially added, and randomness is used to pick the type of atom and the dihedral relationship of each new atom to the previous atoms in the structure. However, there are important differences as well. GenStar allows a range of chemically reasona-

ble bond lengths and bond angles as well as torsion angles which allows for more variability during structure build-up. For each new atom, hundreds of candidates are considered. The selection of atoms in GenStar is governed by enzyme contacts and can be influenced by hydrogen bonding and ring formation; in LEGEND, intermolecular interactions are not taken into account at all during the atom generation process. It would seem that a very high sampling rate would be necessary to obtain even a small number of interesting compounds for further study. The literature description of LEGEND is somewhat brief, however, making more detailed comparisons difficult.

Ideally, a method such as GenStar should include an estimation of hydrophobic effects, which are believed to play a significant role in the binding of drugs to protein [22,23,84,85]. We are experimenting with simple functions which will approximate this contribution. There is no consensus yet in this field, but most recent literature in this field suggests that each \AA^2 of non-polar enzyme surface area buried during the formation of the drug–enzyme complex contributes approximately 20–25 cal/mol [86–96]. We use fast, approximate methods to estimate the amount of ‘hydrophobic enzyme surface’ which would be buried by each candidate atom; the larger this surface area, the greater the contribution the hydrophobic effect makes to the overall score of the candidate. Including this term in GenStar’s scoring function increases the probability that the method will fill hydrophobic cavities.

Our scoring algorithm, which is quite similar to the one used in DOCK [55], uses a hard-sphere potential. In other words, when any candidate atom is too close to any enzyme atom, GenStar rejects the candidate outright. Both GenStar and DOCK set the minimum distance criterion loosely, i.e. we allow closer non-bonded contacts than are normally found in real molecules. For instance, we allow a carbon–carbon separation of 3 \AA , despite the fact that such a separation is expected to be somewhat repulsive. This softening of the non-bonded repulsion term compensates for the fact that the enzyme is stationary during the drug design process. Recently, we have been considering some new potential functions, one of which is shown in Fig. 11. The basic idea is that

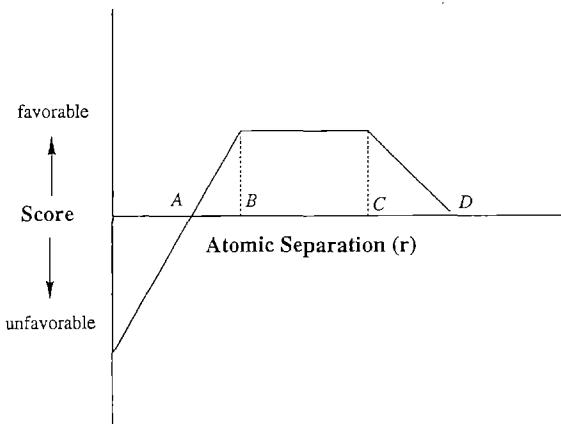


Fig. 11. Improved scoring function which more closely resembles an inverted Leonard-Jones potential. The atom–atom separation distance, r , is shown on the x-axis, while the score for the interaction is given on the y-axis. Positive score indicates an attraction between the two atoms. At small atom–atom separations ($r < A \text{ \AA}$) the score is repulsive (negative). At intermediate distances ($A \text{ \AA} < r < D \text{ \AA}$) favorable interactions exist and attractive scores result. At large atom–atom separations ($r > D \text{ \AA}$) there is essentially no dispersion interaction and the score is 0. See text for details.

rather than *eliminate* candidate atoms which have close non-bonded contacts with the enzyme, we *penalize* these candidate atoms. This helps us to avoid discarding candidate atoms which are overlapping the enzyme but which, after energy minimization, would have acceptable enzyme contacts. If the atomic separation, r , between the candidate atom and an enzyme atom is less than a Å, a negative (penalizing) score results. In effect, Fig. 11 may be viewed as a simplified, inverted Leonard-Jones potential.

Even when issues such as hydrophobic and solvation effects have been addressed, we are left with the fundamental limitation that GenStar builds molecules atom by atom, carbon by carbon. The molecules suggested by an atom-based approach tend to be more flexible than is desirable (although Genstar's ability to close rings is helpful in this regard). Attempting to make the structures more realistic by allowing placement of heteroatoms and different bond types would require a great deal of planning ahead when generating a structure. This, added to the combinatorial explosion of possibilities (because of additional bond and atom types), would make the method impractical. Another objection is that it is more difficult to incorporate information from other techniques (such as GRID [35]) about how several different fragments may be connected in the active site (as was recently demonstrated in the CLIX program [69]). To overcome these limitations, one would like to explore methodology in which molecules are constructed from larger fragments. Our current research is aimed at addressing these issues.

ACKNOWLEDGEMENTS

We thank Joshua Boger, David Deininger, and Roger Tung for careful review of this manuscript, Pat Connelly for reviewing the manuscript and for discussions on the hydrophobic effect, and Mason Yamashita for his review of the manuscript and for stimulating discussions about scoring functions. We also would like to thank Dr. Richard Lewis for kindly providing us with a preprint of his work on irregular lattices (Ref. 49).

REFERENCES

- 1 Cohen, N.C., Blaney, J.M., Humblet, C., Gund, P. and Barry, D.C., *J. Med. Chem.*, 33 (1990) 883.
- 2 Gund, P., Halgren, T.A. and Smith, G.M., *Annu. Rep. Med. Chem.*, 22 (1987) 269.
- 3 Marshall, G.R., *Annu. Rev. Pharmacol. Toxicol.*, 27 (1987) 193.
- 4 Fröhbeis, H., Klein, R. and Wallmeier, H., *Angew. Chem. Int. Ed. Engl.*, 26 (1987) 403.
- 5 Snyder, J.P., *Med. Chem. Rev.*, 11 (1991) 641.
- 6 Appelt, K., Bacquet, R.J., Bartlett, C.A., Booth, C.L.J., Freer, S.T., Fuhr, M.A.M., Gehring, M.R., Herrmann, S.M., Howland, E.F., Janson, C.A., Jones, T.R., Kan, C.-C., Katharadekar, V., Lewis, K.K., Marzoni, G.P., Matthews, D.A., Mohr, C., Moomaw, E.W., Morse, C.A., Oatley, S.J., Ogden, R.C., Reddy, M.R., Reich, S.H., Schoettlin, W.S., Smith, W.W., Varney, M.D., Villafranca, J.E., Ward, R.W., Webber, S., Webber, S.E., Welsh, K.M. and White, J., *J. Med. Chem.*, 34 (1991) 1925.
- 7 Baldwin, J.J., Ponticello, G.S., Anderson, P.S., Christy, M.E., Murcko, M.A., Randall, W.C., Schwam, H., Sugrue, M.F., Springer, J.P., Gautheron, P., Grove, J., Mallorga, P., Viader, M.-P., McKeever, B.M. and Navia, M.A., *J. Med. Chem.*, 32 (1989) 2510.
- 8 Navia, M.A. and Murcko, M.A., *Curr. Opin. Struct. Biol.*, 2 (1992) 202.
- 9 Fesik, S.W., *J. Med. Chem.*, 34 (1991) 2937.
- 10 Brunger, A.T., Kuriyan, J. and Karplus, M., *Science*, 235 (1987) 458.
- 11 Argos, P. and Mohana Rao, J.K., *Methods Enzymol.*, 30 (1986) 185.
- 12 Lathrop, R.H., Webster, T.A. and Smith, T.F., *Comm. ACM*, 30 (1987) 909.

- 13 Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E. and Thornton, J.M., *Nature*, 326 (1987) 347.
- 14 Sander, C. and Schneider, R., *Proteins Struct. Funct. Genet.*, 9 (1991) 56.
- 15 Kollman, P.A., In Horn, A.S. and Ranter, C.J.D. (Eds.), *X-Ray Crystallography and Drug Action*, Oxford University Press, London, 1984, pp. 63–82.
- 16 Dean, P.M., *Molecular Foundations of Drug-Receptor Interaction*, Cambridge University Press, Cambridge, 1987.
- 17 Fersht, A.R., Leatherbarrow, L.J. and Wells, T.N., *Biochemistry*, 26 (1987) 6030.
- 18 Quiocho, F.A., Wilson, D.K. and Vyas, N.K., *Nature*, 340 (1989) 404.
- 19 Sandorfy, C., Buchet, R., Hobza, P. and Ruelle, P., *J. Mol. Struct. (Theochem)*, 107 (1984) 251.
- 20 Teeter, M.M. and Whitlow, M.D., *Trans. ACA*, 22 (1986) 75.
- 21 Zuccarello, F. and Del Re, G., *J. Comp. Chem.*, 8 (1987) 816.
- 22 Tanford, C., *The Hydrophobic Effect.*, 2nd edn., Wiley, New York, 1980.
- 23 Gill, S.J., *Pure Appl. Chem.*, 61 (1989) 1009.
- 24 Lebl, M., Sugg, E.E. and Hruby, V.J., *Int. J. Pept. Protein Res.*, 29 (1987) 40.
- 25 Homer, J. and Mohammadi, M.S., *J. Chem. Soc. Faraday Trans. II*, 83 (1987) 1957.
- 26 Burkert, U. and Allinger, N.L., In *Molecular Mechanics (ACS Monograph Series, No. 177)*, American Chemical Society Press, Washington, 1982, pp. 39–52.
- 27 Still, W.C., Tempczyk, A., Hawley, R.C. and Hendrickson, T., *J. Am. Chem. Soc.*, 112 (1990) 6127.
- 28 Arald, J.-C., Nicholls, A., Sharp, K., Honig, B., Tempczyk, A., Hendrickson, T.F. and Still, W.C., *J. Am. Chem. Soc.*, 113 (1991) 1454.
- 29 Gilson, M.K. and Honig, B., *J. Comput.-Aided Mol. Design*, 5 (1991) 5.
- 30 Burley, S.K. and Petsko, G.A., *Adv. Prot. Chem.*, 39 (1988) 125.
- 31 Thomas, K.A., Smith, G.M., Thomas, T.B. and Feldmann, R.J., *Proc. Natl. Acad. Sci. USA*, 79 (1982) 4843.
- 32 Náray-Szabó, G., *J. Mol. Graphics*, 7 (1989) 76.
- 33 Hayes, D.M. and Kollman, P.A., *J. Am. Chem. Soc.*, 98 (1976) 3335.
- 34 Bellido, M.N. and Rullmann, J.A.C., *J. Comput. Chem.*, 10 (1989) 479.
- 35 Goodford, P.J., *J. Med. Chem.*, 28 (1985) 849.
- 36 Boobbyer, D.N.A., Goodford, P.J., McWhinnie, P.M. and Wade, R.C., *J. Med. Chem.*, 32 (1989) 1083.
- 37 Danziger, D.J. and Dean, P.M., *Proc. R. Soc. London Ser. B, Biol. Sci.*, 236 (1989) 101.
- 38 Danziger, D.J. and Dean, P.M., *Proc. R. Soc. London Ser. B, Biol. Sci.*, 236 (1989) 115.
- 39 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., *J. Comput. Chem.*, 4 (1983) 187.
- 40 Miranker, A. and Karplus, M., *Proteins Struct. Funct. Genet.*, 11 (1991) 29.
- 41 Goodsell, D.S. and Olsen, A.J., *Proteins Struct. Funct. Genet.*, 8 (1990) 195.
- 42 Yue, S.-Y., *Prot. Eng.*, 4 (1990) 177.
- 43 Wilson, S.R. and Guarneri, F., *Tetrahedron Lett.*, 32 (1991) 3601.
- 44 Lee, R.H. and Rose, G.D., *Biopolymers*, 24 (1985) 1613.
- 45 Lesk, A.M., *Acta Crystallogr.*, A42 (1986) 83.
- 46 Lewis, R.A. and Dean, P.M., *Proc. R. Soc. London Ser. B, Biol. Sci.*, 236 (1989) 125.
- 47 Lewis, R.A., Dean, P.M., *Proc. R. Soc. London Ser. B, Biol. Sci.*, 236 (1989) 141.
- 48 Lewis, R.A., *J. Comput.-Aided Mol. Design*, 4 (1990) 205.
- 49 Lewis, R.A., Roe, D.C., Huang, C., Ferrin, T.E., Langridge, R. and Kuntz, I.D., *J. Mol. Graphics*, 10 (1992) 66.
- 50 Bartlett, P.A., Shea, G.T., Telfer, S.J. and Waterman, Spec. Publ., Royal Chem. Soc., 78 (1989) 182.
- 51 CAVEAT: A Program to Facilitate the Design of Organic Molecules. Lauri, G., Shea, G.T., Waterman, S., Telfer, S.J., Bartlett, P.A., Version 2.0, University of California at Berkeley.
- 52 Van Drie, J.H., Weininger, D. and Martin, Y.C., *J. Comput.-Aided Mol. Design*, 3 (1989) 225.
- 53 Martin, Y.C., *Tetrahedron Comp. Methodol.*, 3 (1990) 15.
- 54 Sheridan, R.P., Rusinko, III, A., Nilakantan, R. and Venkataraghavan, R., *Proc. Natl. Acad. Sci. USA*, 86 (1989) 8165.
- 55 Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E., *J. Mol. Biol.*, 161 (1982) 269.
- 56 DesJarlais, R.L., Sheridan, R.P., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., *J. Med. Chem.*, 29 (1986) 2149.
- 57 DesJarlais, R.L., Sheridan, R.P., Seibel, G.L., Dixon, J.S., Kuntz, I.D. and Venkataraghavan, R., *J. Med. Chem.*, 31 (1989) 722.

- 58 DesJarlais, R.L., Seibel, G.L., Kuntz, I.D., Furth, P.S., Alvarez, J.C., De Montellano, P.R.O., DeCamp, D.L., Babé, L.M. and Craik, C.S., Proc. Natl. Acad. Sci. USA, 87 (1990) 6644.
- 59 Lawrence, M.C. and Davis, P.C., Proteins Struct. Funct. Genet., 12 (1992) 31.
- 60 Bohm, H.-J., J. Comput.-Aided Mol. Design, (1992) 61.
- 61 Bernstein, F.C., Koetzle, T.F., Williams, G.T.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., (1977) 535.
- 62 Allen, F.H., Kennard, O. and Taylor, R., Acc. Chem. Res., 16 (1983) 146.
- 63 Nishibata, Y. and Itai, A., Tetrahedron, 47 (1991) 8985.
- 64 Singh, J., Saldanha, J. and Thornton, J.M., Prot. Eng., 4 (1991) 251.
- 65 Moon, J.B. and Howe, W.J., Proteins Struct. Funct. Genet., 11 (1991) 314.
- 66 Wiberg, K.B. and Murcko, M.A., J. Am. Chem. Soc., 110 (1988) 8029.
- 67 Yamashita, M.M., Wesson, L., Eisenman, G. and Eisenberg, D., Proc. Natl. Acad. Sci. USA, 87 (1990) 5648.
- 68 InsightII, version 2.1, Biosym Technologies, San Diego, CA.
- 69 Connolly, M.L., Science, 221 (1983) 709.
- 70 Perry, B.K. and Murcko, M.A., unpublished results.
- 71 Moore, J.M., Peattie, D.A., Fitzgibbon, M.J. and Thomson, J.A., Nature, 351 (1991) 248.
- 72 Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L. and Clardy, J., Science, 252 (1991) 839.
- 73 Michnick, S.W., Rosen, M.K., Wandless, T.J., Karplus, M. and Schreiber, S.L., Science, 252 (1991) 836.
- 74 Van Duyne, G.D., Standaert, R.F., Schreiber, S.L. and Clardy, J., J. Am. Chem. Soc., 113 (1991) 7433.
- 75 CHARMM, version 21, Molecular Simulations, Waltham, MA.
- 76 Swain, A.L., Miller, M.M., Green, J., Rich, D.J., Schneider, J., Kent, S.B.H. and Wlodawer, A., Proc. Natl. Acad. Sci. USA, 87 (1990) 8805.
- 77 Roberts, N.A., Martin, J.A., Kinchington, D., Broadhurst, A.V., Craig, J.C., Duncan, I.B., Galpin, S.A., Handa, B.K., Kay, J., Krohn, A., Lambert, R.W., Merrett, J.H., Mills, J.S., Parkes K.E.B., Redshaw, S., Ritchie, A.J., Taylor, D.A., Thomas, G.H. and Machin, P.J., Science, 248 (1990) 358.
- 78 Krohn, A., Redshaw, S., Ritchie, J.C., Graves, B.J. and Hatada, M.H., J. Med. Chem., 34 (1991) 3340.
- 79 Maren, T.H., Drug Dev. Res., 10 (1987) 255.
- 80 Eriksson, A.E., Jones, T.A. and Liljas, A., Proteins Struct. Funct. Genet., 4 (1988) 274.
- 81 Eriksson, A.E., Kylsten, P.M., Jones, T.A. and Liljas, A., Proteins Struct. Funct. Genet., 4 (1988) 283.
- 82 Merck & Company Annual Report, Rahway, NJ, 1990, p. 36.
- 83 Murcko, M.A., unpublished results.
- 84 Privalov, P.L. and Gill, S.J., Pure Appl. Chem., 61 (1989) 1097.
- 85 Privalov, P.L. and Gill, S.J., Adv. Prot. Chem., 39 (1988) 191.
- 86 Kellis, Jr., J.T., Nyberg, K., Sali, D. and Fersht, A.R., Nature, 333 (1988) 784.
- 87 Rashin, A.A., Iofin, M. and Honig, B., Biochemistry, 25 (1986) 3619.
- 88 Eriksson, A.E., Baase, W.A., Zhang, X.-J., Heinz, D.W., Blaber, M., Baldwin, E.P. and Matthews, B.W., Science, 255 (1992) 178.
- 89 Reynolds, J.A., Gilbert, D.B. and Tanford, C., Proc. Natl. Acad. Sci. USA, 71 (1974) 2925.
- 90 Hermann, R.B., J. Phys. Chem., 76 (1972) 2754.
- 91 Sharp, K.A., Nicholls, A., Friedman, R. and Honig, B., Biochemistry, 30 (1991) 9686.
- 92 Kellis, Jr., J.T., Nyberg, K., Sali, D. and Fersht, A.R., Nature, 333 (1988) 784.
- 93 Chothia, C., J. Mol. Biol., 105 (1976) 1.
- 94 Kellis, J.T., Nyberg, K. and Fersht, A.R., Biochemistry, 28 (1989) 4914.
- 95 Shortle, D., Stites, W.E. and Meeker, A.K., Biochemistry, 29 (1990) 8033.
- 96 Connelly, P.R. and Thomson, J.A., Proc. Natl. Acad. Sci. USA, 89 (1992) 4781.