GroupBuild: A Fragment-Based Method for *De Novo* Drug Design

Sergio H. Rotstein and Mark A. Murcko*

Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, Massachusetts 02139-4211

Received February 9, 1993

A novel method for *de novo* drug design, GroupBuild, has been developed to suggest chemically reasonable structures which fill the active sites of enzymes. The proposed molecules provide good steric and electrostatic contact with the enzyme and exist in low-energy conformations. These structures are composed entirely of individual functional groups (also known as "building blocks" or "fragments") which the program chooses from a predefined library. User-selected enzyme seed atom(s) may be used to determine the area(s) in which structure generation begins. Alternatively, GroupBuild may begin with a predocked "inhibitor core" from which fragments are grown. For each new fragment generated by the program, several thousand candidates in a variety of locations and orientations are considered. Each of these candidates is scored based on a standard molecular mechanics potential function. The selected fragment and orientation are chosen from among the highest scoring cases. Tests of the method using HIV protease, FK506 binding protein, and human carbonic anhydrase demonstrate that structures similar to known potent inhibitors may be generated with GroupBuild.

Introduction

Techniques for the determination of protein structures are advancing rapidly,[1-5] as are homology-based methods for the prediction of protein structure.[6-9] Similarly, structure-based drug design continues to advance and to become more widely accepted.[10-13] As more protein structures become available, whether from crystallography, NMR spectroscopy, or homology modeling, the need grows for a set of computational tools which can analyze protein active sites and suggest compounds which may bind to these sites. Ideally, these methods would be graphical, interactive, and fast and produce a diverse set of chemically and biologically reasonable structures. To the greatest extent possible, they also should be free from arbitrary user bias. Although such "rational *de novo* drug design" methods do not yet exist, many promising approaches toward this goal have been reported in the recent literature.[14-50]

One of the best known methods for analyzing an active site and suggesting ligands with complementary steric properties is the program DOCK, which uses a fast sphere-matching algorithm to dock compounds from a user-supplied database in an enzyme active site.[14-17] DOCK has historically been hampered by its inability to handle ligand flexibility and electrostatic interactions. Recently, however, Leach and Kuntz[18] have developed the "directed DOCK" method, which uses additional information about the hydrogen-bonding characteristics of the enzyme active site to identify portions of ligands with complimentary characteristics and then employs a systematic search algorithm to explore the conformational preferences of the remainder of each ligand. The crystallographic binding orientation of methotrexate to dihydrofolate reductase and of netropsin to the DNA duplex d(CGCGATATCGCG) were both reproduced by this method.[18] A related method has been recently reported by Bacon and Moult in which a least-squares best-fit algorithm is used to maximize the surface overlap of enzymes and possible ligands.[19] Coulombic interactions are included in the scoring function. The method is fast enough that many conformations may be sampled by torsional sampling, helping to overcome the problem of ligand flexibility. In a number of test cases, the method consistently reproduced crystallographic results to within 1 Å.[19]

Other researchers have focused on placing individual groups properly in portions of the active site rather than simultaneously trying to fill the entire region. Goodford has developed GRID, which places small fragment probes at many regularly spaced grid points within the active site, determining the most favorable scores. When tested on a variety of enzyme–inhibitor complexes, GRID has been found to reproduce the positions of important hydrogen-bonding groups.[20,21] A related program is HSITE, which generates a map of the hydrogen-bonding regions of an enzyme active site, including the probability of H-bond formation at each point.[22,23] Miranker and Karplus[24] have modified CHARMM[25] so that molecular dynamics simulations may be performed without non-bonding 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. Finally, a variety of Monte Carlo and simulated annealing approaches have been described recently[26-29] which are well suited for determining low-energy conformations of molecular fragments and in some cases entire ligands.

One of the difficulties with all such methods is that it is a daunting problem to design a drug candidate that properly orients each of the fragments necessary for tight binding, is synthetically tractable, and is likely to be stable *in vivo*. This is an active field of research in computer-based drug design, and a number of methods to connect isolated fragments have been recently reported.

Dean and Lewis have developed methods for using "spacer skeletons" of appropriate size to match ligand atoms to the correct binding sites in enzyme active sites.[30,31] 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.[32] Lewis and co-workers have developed an elegant, efficient method that combines all the atoms from the highest-scoring molecules suggested by DOCK into an *irregular*

* Author to whom correspondence should be addressed.

0022-2623/93/1836-1700$04.00/0

lattice which can be used to connect distant atoms and/or fragments in chemically novel ways.[33] Finally, Lewis has developed a method to connect two isolated fragments with *linear chains* of atoms which can be assembled from a variety of atom types. The chains are generated by solving a series of trigonometric equations in torsion space.[34]

Methodologies for searching three-dimentional databases to test pharmacophore hypotheses and select compounds for screening have been developed by several groups.[35-37] Martin has recently done a very thorough job of reviewing this field.[38] A related program is CAVEAT, developed by Bartlett and co-workers.[39] 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.[40]

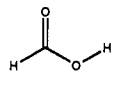
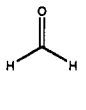
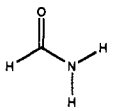
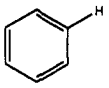
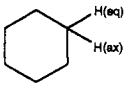
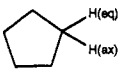
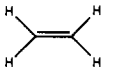
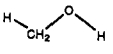
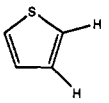
Three recently reported programs, CLIX,[41] LUDI,[42] and the "linked fragment approach"[43] all represent significant advances in fragment-based drug design. These programs have the same basic approach to drug design as that found in CAVEAT[39,40] and also the work by Dean[30,31] and Lewis[32-34]—the connection of separate, individual molecular fragments into a single viable molecule. Despite the similarity of the methods, however, there also are interesting differences in the approaches taken. CLIX[41] uses the output from GRID calculations,[20,21] 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". Then, this information is used to query a chemical database for candidate molecules which have good coincidence of individual fragments with members of the ensemble. Mildly repulsive interactions between candidates and the enzyme were relieved by allowing the candidate to relax slightly (without significantly reducing the overlap of the candidate with the ensemble of binding fragments). Also, CLIX is able to use the information from the GRID potential maps to suggest possible changes in the structures pulled out of the Cambridge Structural Database[44,45] 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[42] also accepts the output from GRID,[20,21] 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 contract pattern derived from the Cambridge Structural Database[44,45] 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" groups such as $-CH_2-$ 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. Finally, Verlinde and co-workers[43] have developed what they call the "linked-fragment approach". The active site of the enzyme is defined and broken into subregions, and the necessary properties of a ligand in each region are determined (shape, hydrophobicity, hydrogen bonding, *etc*). Then a large number of "building blocks" are evaluated in each subregion. Finally, a series of linkers are used to connect the highest-scoring fragments. Energetic analysis is used to guide both the selection and linking of fragments. The design of inhibitors of triosephosphate isomerase was used as a test case.

Two other groups have recently published exciting methods for determining the correct orientation of peptide substrates in enzyme active sites. Unlike the techniques discussed previously, these methods are designed to avoid the difficult program of connecting isolated fragments by using build-up procedures that linearly connect each fragment to the preceeding one. The tradeoff is that a much more limited set of fragments may be considered. Thornton and co-workers[46] have developed GEMINI, 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, which uses a build-up procedure to determine the best peptidal inhibitor or substrate for a given enzyme.[47] In some ways, GROW is reminiscent of the build-up procedure for determining protein conformation developed by Gibson and Scheraga.[48] 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 force field, 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 the structure of which has been determined crystallographically.

Nishibata and Itai have developed LEGEND,[49] 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 postprocessing program, LORE, may be used to select the more interesting structures for graphical analysis.

In our previous communication[50] we described GenStr, which uses sp^3 carbons to build up drug candidates. We highlighted the strengths and weaknesses of the method, and discussed our results from several test cases. In this paper we describe the GroupBuild method. GroupBuild uses a library of common organic templates and a complete force field description of the nonbonding interactions between ligand and enzyme to slowly "build up" drug candidates that have chemically reasonable structures, and have steric and electrostatic properties which are complimentary to the enzyme. We have tested our method with the well-characterized enzymes FKBP-12, human carbonic anhydrase, and HIV protease. As our test cases show, GroupBuild reproduces well-known binding motifs found in a variety of inhibitor classes for all three enzymes.

Table I. Current Fragment Library

 Acid	 Aldehyde	 Amide
 Amine	 Benzene	 Cyclohexane
 Cyclopentane	Ethane	 Ethylene
 Hydroxy	 Methoxy	Methane
 Sulfone	 Thiophene	

Method

In this section we describe the GroupBuild program. First we provide a brief overview, followed by a more detailed presentation of each part of the method.

GroupBuild is a fragment-by-fragment inhibitor generator. The term "fragment" is used to describe the fundamental building block used in the construction of the inhibitors. Each fragment consists of a single functional group such as a hydroxy or a carbonyl or a benzene ring. Our current list of fragments is given in Table I. In its simplest form, the method begins with a "core" fragment which has been predocked by the user into the enzyme active site. This core may be another known inhibitor, a portion of a known inhibitor, or any other structure. Alternatively, the user can specify one or more "seed atoms" from the enzyme and one or more "seed fragments" from the library, and GroupBuild will select an energetically reasonable starting fragment, which then becomes the core.

Once the core is in place, inhibitor generation begins. Each of the hydrogens in the core is replaced in turn by a candidate fragment, which is rotated around the newly created bond in predefined rotational increments. A number of rules are applied to avoid growing fragments in chemically unreasonable ways. The interaction energy between each rotomer and the enzyme is calculated. The effects of solvation also may be included in this process. Once all the possibilities have been scored, one is picked randomly from among all candidates whose score lies within 25% of the top score for that round. The inhibitor is then minimized briefly inside the active site and the whole process begins again, with the newly placed fragment now considered a part of the core. Fragment additions continue until one of several termination conditions is met. The finished inhibitor is stored, and generation of another inhibitor begins. The detailed explanation of the method is divided into the following sections: preparation for program execution, program invocation and user input, first fragment generation, addition of fragments, scoring, minimization, inhibitor termination, program output, and post-processing.

Preparation for Program Execution. Prior to executing GroupBuild, the user must run a preprocessing program, GRID_-SETUP, to generate two special grids. This step is necessary only once for a given enzyme. The enzyme is stored in standard formats.[51] The user must specify the boundaries of a rectangular box containing the enzyme active site. The enzyme and active site boundaries are then used by the preprocessing program to create the grids. The *contacts grid* contains a list of all enzyme atoms that are within 5 Å of each gridpoint. This information simplifies enzyme–inhibitor bump checking because instead of evaluating the interactions of the inhibitor with every enzyme atom, we only need to check the enzyme atoms that are listed in the local gridpoints. A 2-Å grid spacing is used. The second grid is the *potentials grid*. At each gridpoint, we sum the contributions made by each nearby enzyme atom to the attractive and repulsive van der Waals potentials and to the Coulombic potential. Currently, an 8-Å cutoff is used, and the grid spacing is 0.5 Å. The use of these grids leads to a dramatic increase in the speed with which fragment–enzyme interactions may later be computed during the GroupBuild run.

In order to take advantage of some of the features of GroupBuild, it may be necessary to run a second preprocessing program. As with GRID_SETUP, this program, called BUCKETS, must only be run once for a given enzyme. BUCKETS places individual fragments from the fragment library at regular gridpoints within the active site, rotating the fragments in all three dimensions and scoring all the orientations. The resulting output provides a description of the region(s) of the active site where each fragment interacts well with the enzyme. This list can be used by GroupBuild in two different ways. First, statistics from the BUCKETS runs can be used to *check scores* or *normalize scores* during fragment scoring. Second, the actual positions of the functional groups that have good interaction with the enzyme can be used by the program to generate its own starting "core" fragment in those cases in which the user has no predocked core. These options are described below in greater detail.

Since the inhibitors are built fragment by fragment, GroupBuild requires a *fragment_list* of all the functional groups that will be used during the execution. For each fragment, the list of hydrogen atoms which can be used as connection points to other fragments must be provided. Also, if the user wishes to use score checking or normalization during the fragment scoring process, then the statistics about each fragment in the library (generated by the BUCKETS program) must be stored in the list. By editing this list of fragments, the user can determine which fragments will be used during a particular GroupBuild run. Currently our fragment library consists of the 14 functional groups listed in Table I. In order to obtain reasonable inter-fragment bond lengths, bond angles, and torsion angles, as well as to perform inter- and intramolecular bump checking, GroupBuild also requires input files which contain information about the charge, van der Waals radius, and nonbonding parameters of each atom type, as well as information about preferred bond lengths and angles. Standard values are used in all cases.[51]

If an "inhibitor core" is to be used to begin the generation process, it first must be predocked into the enzyme active site. This core may be obtained, for example, from the crystal structure of an enzyme–inhibitor complex. All or part of the inhibitor may be used as the core. Any number of core files may be created and one is selected at run time.

Table II contains a list of variables that the user may wish to customize before execution.

Program Invocation and User Input. When the program is invoked in the stand-alone mode, the user may customize many characteristics of inhibitor generation. We summarize those options here. More details about each are found later in this section.

Solvation effect. Optionally, the effect of desolvating the enzyme may be calculated.

Minimization. The inhibitor may be optimized within the active site as each fragment is added. Various options are available.

Maximum Size of the Inhibitor. The user may specify a maximum allowable size for the inhibitors, expressed in various ways.

Customization. The number of inhibitors to be generated, and the names of the output inhibitor structures, are specified.

Distance Constraints. The user may define a distance constraint criterion preventing the inhibitor from growing too far from a specified enzyme or core atom.

Normalization. The user has the option of using the average score for each fragment, as obtained in the BUCKETS preprocessing run, to "normalize" the scores of each candidate fragment.

Table II. User-Adjustable Program Variables

MAX_NUM_FRAGMENTS	100	maximum number of fragments in the fragment library
MAX_FRAG_ATOMS	60	maximum number of atoms of a fragment
MAX_CORE_ATOMS	300	maximum number of atoms of a core
MAX_INHIB_ATOMS	500	maximum number of atoms of a generated inhibitor
MAX_ENZYME_ATOMS	5000	maximum number of atoms of an enzyme
CONTACTS_GRID_INCR	2.0 Å	resolution of the contacts grid
POTENTIALS_GRID_INCR	0.5 Å	resolution of the potentials grid
H_BOND_LOWER	1.2 Å	lower bounds for hydrogen bonding (H···X distance)
H_BOND_UPPER	2.6 Å	upper bounds for hydrogen bonding (H···X distance)
ROTOMER_INCREMENT	10.0°	increment for bond rotation as each fragment is tested
UNACCEPTABLE_SCORE	-5.0	5.0 kcal/mol repulsion
PERCENT_OF_TOP_SCORE	25.0	a candidate fragment must be within this percentage of the highest-scoring fragment in order to be considered acceptable
VDW_SLOP_INTER	0.5 Å	allowance for intermolecular overlap of atoms
VDW_SLOP_INTRA	0.25 Å	allowance for intramolecular overlap of atoms
OUTPUT_BASENAME		name used for all output pdb files
NUM_INHIBS		number of inhibitors to generate
MAX_ATOMS		maximum number of atoms per inhibitor
MAX_FRAGS		maximum number of fragments per inhibitor
MAX_WEIGHT		maximum molecular weight per inhibitor
CONSTRAINT_ATOM		atom used for distance constraints
MAX_DISTANCE		maximum distance allowed from core atom
MIN_OPTION		1 = whole inhib min; 2 = core fixed; 3 = all but last residue fixed
NUM_MIN_CYCLES		number of cycles of minimization
SCORE_CHECK_OPTIONS		1 = higher than mean for this fragment; 2 = higher than mean + 1 SD; 3 = within the top 10%; 4 = within the top 25%
NUM_SEEDS		number of enzyme seed atoms for first fragment generation
SEED_LIST		list of enzyme seed atoms

Score Checking. This option allows the user to reject any fragment whose score falls below a predefined threshold. The threshold is different for each fragment type and is obtained from the statistics of the BUCKETS preprocessing run.

Core Generation. If this feature is turned on, then the user must enter a list of enzyme seed atoms and allowed fragments. The starting fragment of each inhibitor is then generated in the neighborhood of the specified enzyme seed atom(s), using the information obtained from the BUCKETS preprocessing run. If the core generation feature is turned off, then the user must choose among the available predocked cores.

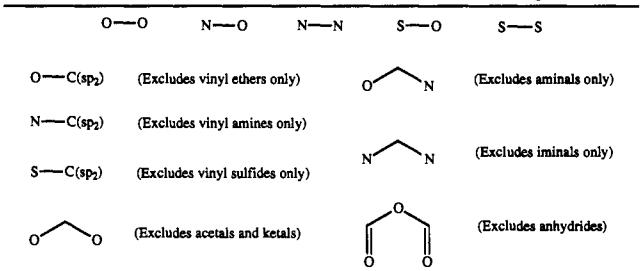
Graphics. GroupBuild may be executed graphically, allowing the user to watch the inhibitor actually being constructed in real-time. (All graphics are done with standard GL library routines running on Silicon Graphics hardware.) Turning off the graphical interface greatly speeds up program execution.

More recently, we have interfaced GroupBuild with InsightII. When run in this manner, all user prompts are handled from dialog boxes and pull-down menus. This interface has proven to be easier for the novice to use.

First Fragment Generation. When no predocked core is used, the user is asked to specify "seed" atom(s) from the enzyme and allowed fragment(s) from the fragment library. GroupBuild must then generate a starting fragment and location. This is done by using the information from the BUCKETS program. The lists of fragment orientations within the active site, and the scores of each, are used to select instances of the specified fragments that are within 5 Å of the specified seed atoms. Once all such fragments have been collected, they are sorted according to their scores. The top 25% or the top 100 fragments (whichever is smaller) of this list are taken and stored in a list of core fragments. When generation of a new inhibitor begins, one of these fragments is chosen at random from the list and used as the starting core. It should be noted that whichever normalization or score checking scheme is chosen by the user also applies to core generation. However if the user specifies no such scheme, the program defaults to considering candidate fragments that score above the average for that fragment type, since we want the starting fragment to interact well with the enzyme. Typically, five seed atoms scattered around the active site are selected, and a small number of simple fragments (e.g., methyl, hydroxy, methoxy, carbonyl) are chosen.

Addition of Fragments. In order to add a new fragment to an inhibitor the program selects a hydrogen from the inhibitor and a hydrogen from the fragment. The bonds between the hydrogens and their corresponding heavy atoms are used to align the new fragment with the structure in such a way that a bond can be created between these heavy atoms. The length of the new bond is determined using the two non-hydrogen atom types and a simple lookup table. Once the fragment has been oriented correctly, the hydrogens are deleted and the new bond is created.

Table III. Chemical Rules (Disallowed Bonds and Angles)



The program always attempts to grow all available fragments from each of the hydrogens in the inhibitor. However, in every inhibitor there are some hydrogens which are already so close to the enzyme that no substituent can be added. Such hydrogens are marked as "dead ends" and no further attempts are made to grow fragments from them. To maintain this information, a special *growth list* is maintained. This list contains information about which fragments can be grown from which inhibitor atoms. When a fragment is chosen from all the candidates and added to the growing inhibitor, all entries of the growth list that contain the deleted hydrogen are removed. Any entries in the list that would result in chemically unreasonable bonds also are removed. For example, we do not want GroupBuild to form peroxides or vinyl amines. The complete list of rules is found in Table III.

Each candidate fragment is attached to the structure and rotated around the new bond in increments of 10°. First, a simple set of rules is used to reject rotamers with torsion angles which are known to be high in energy (for example, a sequence of four consecutive sp^3 carbons with a torsion angle of 0°). Next, for each rotamer, we look for repulsive interactions between the new fragment and the enzyme, and between the new fragment and the rest of the inhibitor. It is best to use a loose definition of "repulsive" to allow more possibilities to be tested. Currently, two atoms are considered to be too close to each other if their distance is less than the sum of their van der Waals radii minus a user-controlled overlap allowance. The current default is 0.5 Å. In other words, if two atoms interpenetrate by more than 0.5 Å, we throw away that fragment conformer. An exception to the above rule is made if the candidate fragment is capable of hydrogen bonding to the interpenetrating enzyme atom. For

these fragments, when we check for steric repulsion to the enzyme, we also ensure that each conformer is able to make at least one hydrogen bond to the enzyme. If hydrogen-bond formation is impossible, then that fragment is rejected outright. We use a loose definition of the hydrogen-bonding geometry to allow for geometric optimization of the inhibitor after the fragment has been added. Even with the loose definition, we have found that more than half of the cases are rejected using this criterion, resulting in more reasonable structures and saving significant computation time.

Scoring. If the fragment conformation passes all of the rule-based tests described above, we move on to actually scoring the binding energy of the fragment. As previously described, we precalculate a potentials grid which contains the sum of all enzyme van der Waals and Coulombic potentials at each grid point.[51] The potentials grid is used to approximate the van der Waals and Coulombic interactions between the fragment and enzyme. The degree of error depends directly on the granularity of the potentials grid (i.e. the gridpoint separation). We have determined that a gridpoint spacing of 0.5 Å gives adequate results. For each atom in the new fragment, we identify the nearest gridpoint and retrieve the potentials information. Then we multiply those values by the van der Waals and charge parameters for the fragment atom.

We also would like the fragment selection process to include some accounting for solvation effects. The current approach is straightforward. At the beginning of a GroupBuild run, we use the SASA algorithm[52] to calculate the exposed solvent-accessible surface for every non-hydrogen atom in the enzyme. This value is stored. We also store the solvent-accessible surface for each non-hydrogen atom of each fragment. Later, for each fragment we score, we calculate the exposed surface area of the non-hydrogen atoms of the enzyme–fragment complex and subtract it from the values of the new fragment and the enzyme separately. This gives us the change in exposed surface area, ΔA, for each non-hydrogen atom of both the enzyme and the fragment. We then multiply ΔA by −0.003 if the atom is "nonpolar" (defined as having an absolute charge of less than 0.2). The basic idea is to provide an additional energetic reward for burying nonpolar atoms, which make good hydrophobic contacts and exclude water. The value of 3 cal/Å^2 is based on recent literature work quantifying hydration free energies.[53-57]

There is one final difficulty which must be addressed. Different fragment types will achieve very different enzyme interaction energies. For example, a *poorly* docked water fragment can have an interaction energy of −5 kcal/mol while a well-placed benzene ring may only be −2 kcal/mol. To overcome this problem we have developed two different techniques: *score checking* and *normalization*. Normalization is done by dividing the score for each candidate fragment by the average score of that fragment as previously determined in the BUCKETS preprocessing simulations. Normalization helps to give hydrogen-bonding and non-hydrogen-bonding groups, as well as large and small hydrophobic groups, approximately equal weighting. In this way, the fragments which are positioned well in a statistical sense—that is, in comparison to other orientations of the same fragment—have the highest likelihood of being selected. Score checking simply forces GroupBuild to discard candidate fragments which have scores below a pre-established level. This level is based on the statistics for each fragment as determined from the BUCKETS preprocessing simulation. There are four possible criteria the user may select. A fragment can be rejected if its score is (a) lower than the average score for that fragment type, (b) lower than its average plus one standard deviation, (c) not a member of the top 10% of the scores for that fragment type, or (d) not a member of the top 25%. Option a is the most generous, as it discards only about half of the scored fragments. Option c is the most restrictive, as 90% of the scored fragments are rejected.

Finally, the list of scores is sorted and a candidate is selected randomly from among the top 25%. The randomness is introduced here because of the implicit imperfections of the scoring function, and also to generate a more varied set of inhibitors. Once we have selected a candidate, we connect it permanently to the structure.

Minimization. GroupBuild can perform a geometric optimization of the entire inhibitor in the enzyme active site after the addition of each fragment. Currently, this is done by writing a script file suitable for Discover.[51] The enzyme is fixed. A distance-dependent dielectric and the conjugant gradient minimizer are used. The number of steps of minimization is selected by the user; the default is 50. The user can choose whether to minimize the whole inhibitor every time, minimize the inhibitor, keeping the starting core fixed, or minimize just the last fragment that was added, keeping the rest of the inhibitor fixed.

Inhibitor Termination. An inhibitor is complete when any of the following conditions is met: (a) The user-specified size of the inhibitor has been exceeded; the number of atoms, molecular weight, and number of fragments all may be specified as criteria. (b) The active site is filled, i.e., any additional fragments will be placed outside the user-specified active site boundaries. This includes any distance constraints selected by the user. (c) Every proposed new fragment has a score below the user-specified tolerance limit, if any.

Program Output. Each inhibitor created by GroupBuild is written out in PDB format. Each fragment has a unique residue name and number. The temperature factor column is used to hold the scores of each fragment. Additional information about the run (user name, date, enzyme, and the like) are stored in HEADER and REMARK records. A script is written which can direct the InsightII program to read all the generated inhibitors, along with the enzyme and other reference compounds (such as known inhibitors).

Postprocessing. One of the common attributes of *de novo* drug design programs is the large number of inhibitors they can design. GroupBuild, for example, can generate hundreds of suggested compounds in the course of an overnight run on a typical workstation. This capability requires facile methods for analysis of the output.

First, we read all the structures into InsightII and view them sequencially with our *browse* module.[58] This module, developed with Open Interface,[59] allows the chemist to quickly examine each inhibitor bound in the active site. They can be compared with other known compounds, and a subset of the GroupBuild inhibitors may be selected for more detailed study.

A second postprocessing tool is our pattern recognition software.[50] We superimpose all structures generated by GroupBuild and search for common "themes", for example, a carbonyl group which is always placed in a certain location in order to take advantage of a hydrogen bond or a collection of hydrophobic fragments which are all located near a certain greasy side chain.

Finally, we would like to know whether any of the structures generated by GroupBuild are identical or similar to known compounds. We are developing tools which allow the chemist to easily run a "similarity search" against the Cambridge Structural Database (CSD) of small molecules[44,45] or the Fine Chemicals Directory (FCD) of commercially available compounds.[60] This is of considerable importance to the medicinal chemist, who may be intrigued by the ideas generated by a method such as GroupBuild but who (quite reasonably) would like some independent validation of the design concepts before engaging in a lengthy synthesis.

Results

GroupBuild may be used either for completely *de novo* drug design or the modification of known drugs using a core structure. We have evaluated GroupBuild in each kind of situation. As test systems, we have considered FK506 binding protein (FKBP-12), HIV-1 aspartyl proteinase (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 GroupBuild would suggest candidates reminiscent of known drugs.

1. **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.[56-64] We performed two types of runs with FKBP. In the first, we began with the piperidine ring docked in

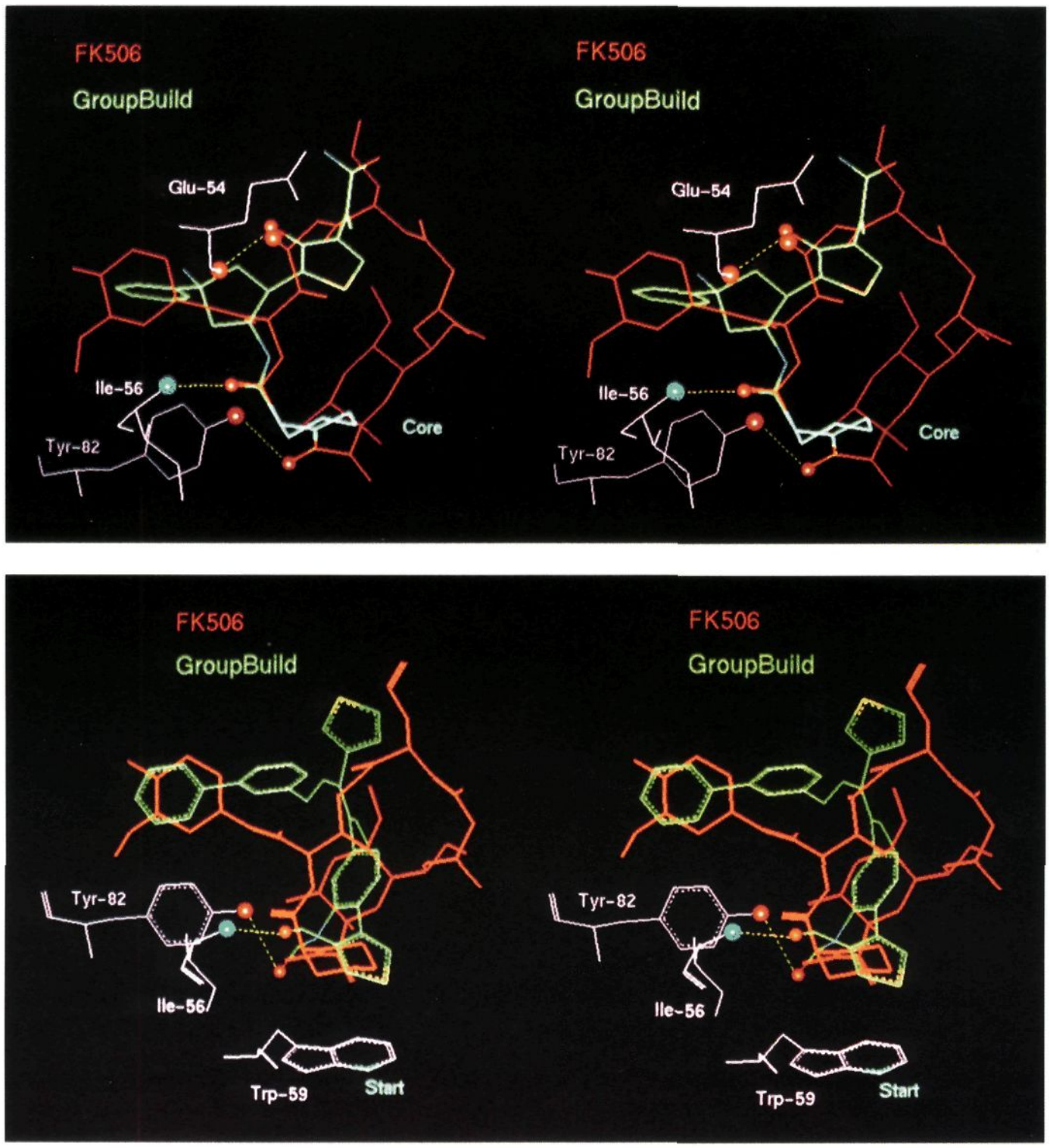


Figure 1. (a, top) FKBP-12 inhibitor generated by the program using only the piperidine ring of FK506 as a staring core. The piperidine ring is shown in light blue, and the rest of FK506 is red. The generated inhibitor is shown using standard atom colors (carbon = green, nitrogen = blue, oxygen = red, sulfur = yellow). Hydrogen bonds between the enzyme and the generated inhibitor are shown as dashed lines, and the enzyme residues are labeled. The atoms of the drug and enzyme involved in those hydrogen bonds are shown as colored spheres. Dark blue is used for nitrogen and red for oxygen. (b, bottom) Completely *de novo* FKBP-12 inhibitor created by GroupBuild. Generation began in the vicinity of Trp-59. FK506 is shown in red. The generated inhibitor is shown using standard atom colors (carbon = green, nitrogen = blue, oxygen = red, sulfur = yellow). Hydrogen bonds between the enzyme and the generated inhibitor are shown as dashed lines, and the enzyme residues are labeled. The atoms of the drug and enzyme involved in those hydrogen bonds are shown as colored spheres. Dark blue is used for nitrogen and red for oxygen.

its crystallographic orientation. A large majority of the structures generated from this core reproduce the ester, the pyruvamide, and the C-24 hydroxyl group present in FK506; a representative example is shown in Figure 1a. In the second run, we did not use a core inhibitor, instead selecting several enzyme atoms from residues Ile-56 and Trp-59 as seeds. The inhibitors generated from this run show greater diversity, as expected. However, a large majority of them reproduce the key structural elements found in FK506 and rapamycin: the greasy binding core near Trp-59, the hydrogen bonding interactions with Ile-56 and Tyr-82, and the greasy group in the "northwest corner" of the active site. A representative example is shown in Figure 1b. Clearly, GroupBuild is providing reasonable suggestions for novel classes of compounds.

2. HIV Protease. As a test with HIV protease, we began with a predocked "core" inhibitor composed of the known inhibitor Ro 31-8959.[65-67] In the first series of runs, we removed everything on the N-terminal side of the P2 asparagine α carbon (we also removed the P2 side chain) and allowed GroupBuild to fill the P2 and P3 pockets. A representative result is shown in Figure 2a. Here, an amide was placed in the P2 side chain, positioned so as to make quite similar hydrogen-bonding interactions with Asp-29 and Asp-30. Also, an amide was placed along the backbone, perfectly mimicking the known inhibitor. Finally, a thiophene ring was placed in the P3 pocket quite close to the piperidine ring of the known inhibitor. In the second series of runs, we removed everything on the C-terminal side of the central hydroxy group and allowed GroupBuild to suggest P1′ and P2′ modifications. A representative result is shown in Figure 2b. A ring is placed in the P1′ position, followed by an amide, then an isopropyl—in each case, a very close match to the known inhibitor. Interestingly, an amide is placed at the end of the molecule, interacting with Asp-29 and Asp-30 in the P2′ pocket, quite analogous to the asparagine side chain interactions in the identical P2 pocket.

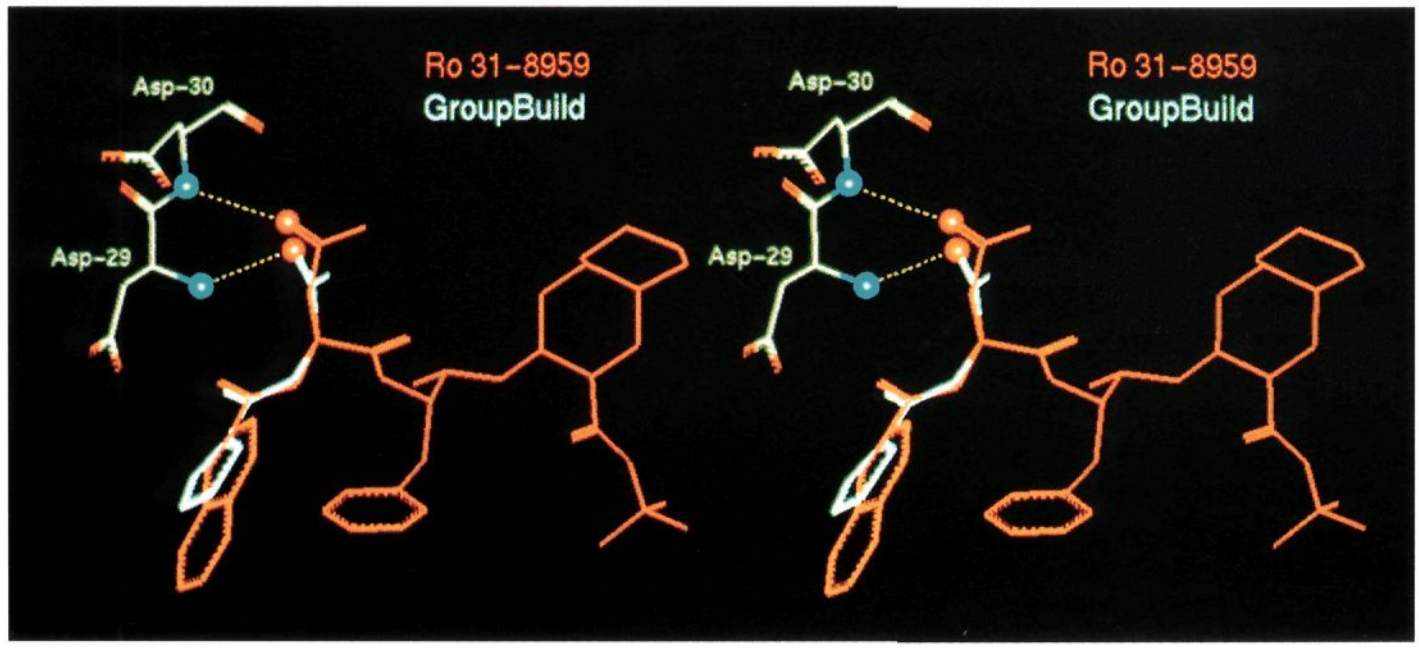


Figure 2. (a, top) HIV inhibitor generated by GroupBuild using a part of the Roche compound, Ro 31-8959 as a starting core. The Roche inhibitor is shown in red and the generated inhibitor is shown in light blue. The program was allowed to generate into the P_2 and P_3 regions of the enzyme active site. Hydrogen bonds between the enzyme and the generated inhibitor are shown as dashed lines, and the enzyme residues are labeled. The atoms of the drug and enzyme involved in those hydrogen bonds are shown as colored spheres. Dark blue is used for nitrogen and red for oxygen. (b, bottom) A Second example of an HIV inhibitor generated by GroupBuild using part of the Roche compound as a starting core. The Roche compound is shown in red and the generated inhibitor is shown in light blue. The program was allowed to generate into the P_1' and P_2' regions of the enzyme active site. Hydrogen bonds between the enzyme and the generated inhibitor are shown as dashed lines, and the enzyme residues are labeled. The atoms of the drug and enzyme involved in those hydrogen bonds are shown as colored spheres. Dark blue is used for nitrogen and red for oxygen.

3. Carbonic Anhydrase. It is well-known that aromatic sulfonamides are excellent inhibitors of human carbonic anhydrase (HCA-II).[68] The sulfonamide group has been shown to bind to the catalytic Zn^{2+} ion.[69,70] We began with a sulfonamide docked in the crystallographically determined orientation[12,69,70] as our core. Several representative examples are shown in Figure 3a,b. For comparison, the structure of MK-417, a potent HCA-II inhibitor that has been shown to lower intraocular pressure in man, is also shown. The structure of MK-417 is very closely related to another HCA-II inhibitor, MK-507, currently in Phase III clinical trials[71] for the treatment for glaucoma. In the vast majority of structures generated by GroupBuild, an aromatic ring is directly attached to the sulfonamide group, as is found in MK-417. Further, the hydrogen bond between the side chain of Gln-92 and the sulfone oxygen is nicely reproduced. Finally, hydrophobic interactions with Phe-131, Pro-201, and other residues that line the bottom of the active site are present in most of the generated inhibitors.

Discussion

Our previous paper described GenStr, which built inhibitors using only sp^3 hybridized carbon atoms to build inhibitor structures from scratch. Decisions about the placement of atoms were based on a crude measure of van der Waals contact with the enzyme, and simple rules were used to ensure that reasonable low-energy conformations were selected. This method has proven remarkably useful at identifying useful structural motifs for inhibitor design.[50] However, GenStr suffers from the inability to handle heteroatoms and atomatic systems. We are exploring ways to add these capabilities to GenStr, maintaining its atom-by-atom approach to ligand construction, but at the same time, we have some concerns that the efficiency of this approach will be unacceptably low. Accordingly, we have developed GroupBuild, which uses a fragment-based approach and a much more realistic representation of enzyme–ligand interaction energies.

Critical Features of the GroupBuild Algorithm. There are several critical aspects to the drug design process used by GroupBuild. These include the scoring of various types of fragments, the selection of each fragment, and the avoidance of chemically unreasonable structures.

Fragment Scoring. Since this method produces inhibitors composed of many fragments, ranging from purely hydrophobic to highly polar and capable of forming multiple hydrogen bonds, it is difficult to imagine that a

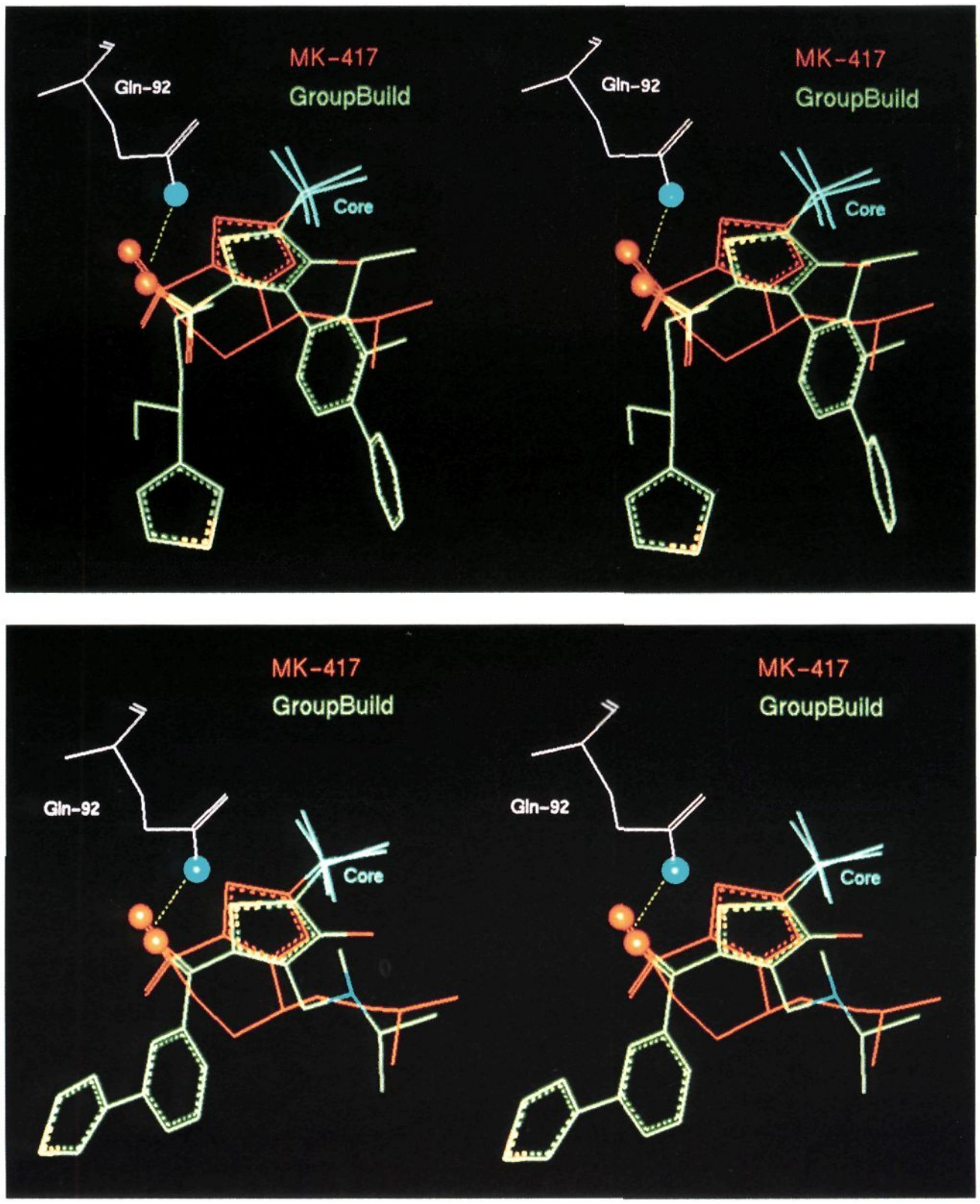


Figure 3. (a, top) An inhibitor of human carbonic anhydrase type II (HCA-II) generated by GroupBuild. Only the sulfonamide portion of the drug MK-417 was used as the starting core. MK-417 is shown in red, except for the sulfonamide core, which is shown in light blue. The generated inhibitor is shown using standard atom colors (carbon = green, nitrogen = blue, oxygen = red, sulfur = yellow). Hydrogen bonds between residue Gln-92 and the inhibitor are shown as dashed lines, and the enzyme residue is labeled. The atoms of the drug and enzyme involved in those hydrogen bonds are shown as colored spheres. Dark blue is used for nitrogen and red for oxygen. (b, bottom) A Second example of an HCA-II inhibitor generated by GroupBuild. Only the sulfonamide portion of the drug MK-417 was used as the starting core. MK-417 is shown in red, except for the sulfonamide core which is shown in light blue. The generated inhibitor is shown using standard atom colors (carbon = green, nitrogen = blue, oxygen = red, sulfur = yellow). Hydrogen bonds between residue Gln-92 and the inhibitor are shown as dashed lines, and the enzyme residue is labeled. The atoms of the drug and enzyme involved in those hydrogen bonds are shown as colored spheres. Dark blue is used for nitrogen and red for oxygen.

crude method will be sufficiently accurate to make the scoring meaningful. Instead, we use the "forcefield on a grid" approach, as described in the Methods section. This gives us interaction energies which are quite reasonable so long as a fine enough grid is used. We have found that 0.5 Å is sufficient, while 0.25 Å provides little additional advantage at great additional cost in computer time. However, the fragment selection process is still nontrivial, due to the fact that different kinds of fragments tend to have quite different scores. For example, even a poorly positioned water can have a better interaction enthalpy than a well-positioned benzene. When the free energy of transfer of the group from bulk solvent to bound conformation is considered, this discrepancy should disappear. However, although the consideration of solvation and entropic effects which we have implemented will help overcome this problem, the methods necessary for the rigorous calculation of these effects are still impractical. Therefore we have concluded that we need some manner of "equating" the respective groups, and that this would be true regardless of the manner in which the groups are positioned in the active site and their binding energies calculated. We can perform this adjustment in several different ways. In the first method, called normalization, we divide the score for each candidate fragment by the average score seen for that fragment in the BUCKETS run carried out prior to starting GroupBuild. This tends to make the selection of fragments more independent of size, which we consider desirable because it ensures that large bulky rings are only added where they are truly beneficial in terms of binding. Miranker and Karplus used an estimate of the solvation free energy for each fragment

to achieve a similar weighting factor.[24] In the second method, called score checking, we simply discard candidate fragments which score below a user-specified threshold. As was explained in the Methods section, the user may select several criteria, all based on the statistics coming from the BUCKETS runs for each fragment type.

Fragment Selection. Once all the possibilities have been scored and (if desired) normalized, the actual fragment selection must be made. In our earlier atom-based method, GenStr, we selected randomly from the list of high-scoring candidate atoms. In practice, this meant that out of 300–400 candidates, approximately 20 were among the best scoring, and one of these was selected. In GroupBuild, we select among the fragment rotamers with scores within 25% of the best score.

Chemically Reasonable Structures. A third issue is the chemical validity of the resulting structures. Ideally, GroupBuild (and all other ligand design methods) should suggest compounds of low molecular weight, in low-energy conformations, with few stereocenters, no chemically unreasonable bonds, and no complex ring systems. We have attempted to address most of these concerns. First, the size of the inhibitors may be limited by molecular weight, number of atoms, number of fragments, and maximum allowed distance from a specified atom. Second, as described in the Methods section, we avoid chemically or biochemically unreasonable bonds. For example, peroxides are not allowed because these structures tend to be unstable in vivo. A list of the disallowed connections is given in Table III. Third, we include intramolecular bump checking and some torsional rules to help avoid high-energy conformations. Fourth, the current method does not allow the formation of complex ring systems. This is only a partial solution because some "complex" ring systems are actually quite easy to synthesize. Unfortunately, it is quite difficult to establish accurate rules capable of recognizing which ring systems are reasonable and which are not. For similar reasons, we have chosen for version 1.0 not to address the issue of stereochemistry. In practice, this has turned out not to be a serious issue, as the typical GroupBuild structure contains only two or three stereocenters, which is not excessive. It is also important to bear in mind that GroupBuild is primarily used as an "idea generation" tool.

Approaches for Fragment-Based *De Novo* Drug Design. Two quite distinct approaches for fragment-based *de novo* drug design seem to be emerging in the recent literature. The first involves the connection of isolated well-placed functional groups with "linker" or "skeleton" fragments. This approach is exemplified by CAVEAT,[39,40] CLIX,[41] LUDI,[42] the linked-fragment approach,[43] the work of Dean,[30,31] the methods of Lewis,[32–34] and the 3D database methods.[35–38] The location of the fragments may be determined within the program itself, as with LUDI,[42] or may come from other preprocessing programs such as GRID.[18,19] The second, exemplified by GEMINI,[46] GROW,[47] LEGEND,[49] GenStr,[50] and now GroupBuild, involves the build-up of a drug candidate by the sequencial addition of new fragments. Each approach has distinct strengths and weaknesses. The differences between the two approaches is shown graphically in Figure 4, and described below.

Connection of Isolated Fragments. This method relies on the fact that a small number of well-placed fragments, each making "key" interactions with the enzyme, may provide a significant portion of the overall binding energy. Pharmacophore models usually include only a small number of binding elements, and this "connect-the-fragments" approach to ligand design allows the researcher to construct a set of proposed inhibitors which fit the model and then either synthesize them or locate them in a library of available compounds. An advantage of this strategy is that information about favorable fragment locations may be obtained from any source. A second advantage, in principle, is that structural rigidity may be incorporated into the proposed inhibitor by using libraries of cyclic "scaffolds". There also are several disadvantages to this approach. First, any connector scaffold, no matter how good it may be in other respects, may be rejected because a single atom overlaps with the enzyme (Figure 4a). This risk can be managed or minimized by the use of "forgiving" scoring functions,[41] but cannot completely be eliminated. Second, there is often no chemical way to make the molecule so constructed. Finally, the linker pieces are often rather large, adding considerably to the molecular weight but not necessarily providing much in the way of binding interactions.

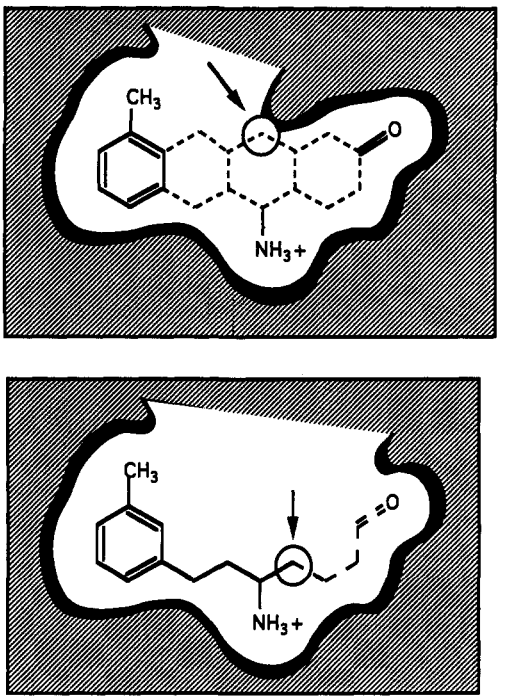


Figure 4. (a, top) Possible problem with the "connect-the-fragment" approach to ligand design. A proposed tricyclic "linker scaffold" is shown in dashed lines. It can connect the aromatic ring, the amino group, and the carbonyl. However, one atom of the tricyclic linker (circled) is severely bumping the enzyme and the entire scaffold may be rejected. See text for further discussion. (b, bottom) Possible problem with the "sequential build-up" approach to ligand design. Here, the binding elements shown in dashed lines may never be selected by this method, since one of the atoms (circled) is relatively far away from the enzyme, and probably will be given a low score. In other words, the method will fail to "traverse" the active site from one region of tight binding to the other. See the text for further discussion.

Sequential Fragment Build-Up. In this approach, one builds up the inhibitor structure fragment by fragment; each new group interacts favorably with the enzyme. This method, as described in the current literature, differs from the fragment-connection algorithms in a subtle but crucial way. No information about "critical binding regions" is used in the beginning to identify disconnected regions of the active site which must be filled. In the event that we must cross an "open region" of the active site in order to get from one desirable "binding pocket" to another, there

is a good chance that one of the sites will be missed (Figure 4b). This is because in a sequential build-up approach, fragments which do not form good interactions with the enzyme are not scored highly, and consequently are never chosen. There are several methods which might be employed to overcome this deficiency. We might add a suitable weighting factor so as to "nudge" the growth process toward the desired regions. (This, of course, presupposes that such information is available, perhaps from analysis of previous runs.) We might start from distinct places, filling each "subsite", and in a separate postprocessing step connect those isolated fragments later.

Other Enhancements to GroupBuild. There are a number of other enhancements that may be made to GroupBuild.

Spanning the Active Site. As previously discussed, it is sometimes difficult for the GroupBuild algorithm to "span" an active site, i.e., to traverse an open space between two regions where fragments can bind tightly. We are currently exploring a variety of ways to overcome this limitation, as mentioned above.

Potential Functions. We are continuing to experiment with our potential functions, including a variety of ways to incorporate estimates of solvation and hydrophobic effects in the drug-design process.

Additional Functional Groups. We are adding more fragments, especially terminal fragments (fluoro, chloro, nitro, etc.) and several more heterocycles to broaden the structural diversity of generated compounds.

Acknowledgment. We would like to thank Scott LeGrand and Ken Merz of Penn State University for giving us access to their SASA surface area program. We thank Drs. Chris Lepre, Govinda Rao, Pat Connolly, and David Pearlman for critical review of the manuscript.

References

(1) Fesik, S. W. NMR Studies of Molecular Complexes as a Tool in Drug Design. *J. Med. Chem.* **1991**, *34*, 2937–2945.
(2) Brunger, A. T.; Kuriyan, J.; Karplus, M. Crystallographic *R* Factor Refinement by Molecular Dynamics. *Science* **1987**, *235*, 458–460.
(3) Markley, J. L. Two-Dimentional Nuclear Magnetic Resonance Spectroscopy of Proteins: An Overview. *Methods Enzymol.* **1989**, *176*, 12–66.
(4) Clore, G. M.; Gronenborn, A. M. Structures of Larger Proteins in Solution: Three- and Four-Dimentional Heteronuclear NMR Spectroscopy. *Science* **1991**, *252*, 1390–1399.
(5) Brunger, A. T. Free R Value: A Novel Statistical Quantity For Assessing the Accuracy of Crystal Structures. *Nature* **1992**, *355*, 472–475.
(6) Argos, P.; MohanaRao, J. K. Prediction of Protein Structure. *Methods Enzymol.* **1986**, *30*, 185–207.
(7) Lathrop, R. H.; Webster, T. A.; Smith, T. F. Ariadne: Pattern-Directed Inference and Hierarchichal Abstraction in Protein Structure Recognition. *Commun. ACM* **1987**, *30*, 909–921.
(8) Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M. Knowledge-Based Prediction of Protein Structures and the Design of Novel Molecules. *Nature* **1987**, *326*, 347–352.
(9) Sander, C.; Schneider, R. Database of Homology-Derived Protein Structures and the Structural Meaning of Sequence Alignment. *Proteins: Struct., Funct., Genet.* **1991**, *9*, 56–68.
(10) Cohen, N. C.; Blaney, J. M.; Humblet, C.; Gund, P.; Barry, D. C. Molecular Modeling Software and Methods for Medicinal Chemistry. *J. Med. Chem.* **1990**, *33*, 883–894.
(11) Appelt, K.; Bacquet, R. J.; Bartlett, C. A.; Booth, C. L. J.; Freer, S. T.; Fuhry, M. A. M.; Gehring, M. R.; Herrmann, S. M.; Howland, E. F.; Janson, C. A.; Jones, T. R.; Kan, C.-C.; Kathardekar, 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.; White, J. Design of Enzyme Inhibitors Using Iterative Protein Crystallographic Analysis. *J. Med. Chem.* **1991**, *34*, 1925–1934.
(12) 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.; Navia, M. A. Thienothiopyran-2-sulfonamides: Novel Topically Active Carbonic Anhydrase Inhibitors for the Treatment of Glaucoma. *J. Med. Chem.* **1989**, *32*, 2510–2513.
(13) Navia, M. A.; Marcko, M. A. The Use of Structural Information in Drug Design. *Curr. Opinions Struct. Biol.* **1992**, *2*, 202–210.
(14) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A Geometric Approach to Macromolecule–Ligand Interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
(15) DesJarlais, R. L.; Sheridan, R. P.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Docking Flexible Ligands to Macromolecular Receptors by Molecular Shape. *J. Med. Chem.* **1986**, *29*, 2149–2153.
(16) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using Shape Complementarity as an Initial Screen in Designing Ligands for a Receptor Binding SIte of Known Three-Dimensional Structure. *J. Med. Chem.* **1989**, *31*, 722–729.
(17) 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.; Craik, C. S. Structure-Based Design of Nonpeptide Inhibitors Specific for the Human Immunodeficiency Virus 1 Protease. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644–6648.
(18) Leach, A. R.; Kuntz, I. D. Conformational Analysis of Flexible Ligands in Macromolecular Receptor Sites. *J. Comput. Chem.* **1992**, *13*, 730–748.
(19) Bacon, D. J.; Moult, J. Docking by Least-Squares Fitting of Molecular Surface Patterns. *J. Mol. Biol.* **1992**, *225*, 849–858.
(20) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
(21) Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New Hydrogen-Bond Potentials for Use in Determining Energetically Favorable Binding Sites in Molecules of Known Structure. *J. Med. Chem.* **1989**, *32*, 1083–1094.
(22) Danziger, D. J.; Dean, P. M. Automated Site-Directed Drug Design: A General Algorithm for Knowledge Acquisition about Hydrogen-Bonding Regions at Protein Surfaces. *Proc. R. Soc. London* **1989**, *B 236*, 101–113.
(23) Danziger, D. J.; Dean, P. M. Automatic Site-Directed Drug Design: The Prediction and Observation of Ligand Point Positions at Hydrogen-Bonding Regions on Protein Surfaces. *Proc. R. Soc. London* **1989**, *B 236*, 115–124.
(24) Miranker, A.; Karplus, M. Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29–34.
(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
(26) Goodsell, D. S.; Olsen, A. J. Automated Docking of Substrates to Proteins by Simmulated Annealing. *Proteins: Struct., Funct., Genet.* **1990**, *8*, 195–202.
(27) Yue, S.-Y. Distance-Constrained Molecular Docking by Simulated Annealing. *Protein Eng.* **1990**, *4*, 177–184.
(28) Wilson, S. R.; Guarneri, F. Calculation of Rotational States of Flexible Molecules using Simulated Annealing. *Tetrahedron Lett.* **1991**, *32*, 3601–3604.
(29) Hart, T. N.; Read, R. J. A Multiple-Start Monte-Carlo Docking Method. *Proteins: Struct., Funct., Genet.* **1992**, *13*, 206–222.
(30) Lewis, R. A.; Dean, P. M. Automated Site-Directed Drug Design: The Concept of Spacer Skeletons for Primary Structure Generation. *Proc. R. Soc. London* **1989**, *B 236*, 125–140.
(31) Lewis, R. A.; Dean, P. M. Automated Site-Directed Drug Design: The Formation of Molecular Templates in Primary Structure Generation. *Proc. R. Soc. London* **1989**, *B 236*, 141–162.
(32) Lewis, R. A. Automated Site-Directed Drug Design: Approaches to the Formation of 3D Molecular Graphs. *J. Comput. Aided Mol. Des.* **1990**, *4*, 205–210.
(33) Lewis, R. A.; Roe, D. C.; Huang, C.; Ferrin, T. E.; Langridge, R.; Kuntz, I. D. Automated Site-Directed Drug Design using Molecular Lattices. *J. Mol. Graphics* **1992**, *10*, 66–78.
(34) Lewis, R. A. Automated Site-Directed Drug Design: A Note on the Generation of General 3D Molecular Graphs. *J. Mol. Graphics*, In press.
(35) Sheridan, R. P.; Rusinko, A. III; Nilakantan, R.; Venkataraghavan, R. Searching for Pharmacophores in Large Coordinate Databases and its Use in Drug Design. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8165–8169.
(36) Van Drie, J. H.; Weininger, D.; Martin, Y. C. ALADDIN: An Integrated Tool for Computer-Assisted Molecular Design and Pharmacophore Recognition From Geometric, Steric, and Substructure Searching of Three-Dimentional Molecular Structures. *J. Comput-Aided Mol. Des.* **1989**, *3*, 225–251.
(37) Jakes, S. E.; Willett, P. Pharmacophoric Pattern MAtching in Files of 3-D Chemical Structures: Selection of Interatomic Distance Screens. *J. Mol. Graphics* **1986**, *4*, 12–20.
(38) Martin, Y. C. 3D Database Searching in Drug Design. *J. Med. Chem.* **1992**, *35*, 2145–2154.

(39) Bartlett, P. A.; Shea, G. T.; Telfer, S. J.; Waterman, S. CAVEAT: A Program to Facilitate the Structure-Derived Design of Biologically Active Molecules. In *Molecular Recognition in Chemical and Biological Problems*; Special Pub., Royal Chem. Soc.: London, 1989, Vol. 78, pp 182–196.
(40) 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.
(41) Lawrence, M. C.; Davis, P. C. CLIX: A Search Algorithm for Finding Novel Ligands Capable of Binding Proteins of Known Three-Dimensional Structure. *Proteins: Struct. Funct., Genet.* **1992**, *12*, 31–41.
(42) Bohm, H.-J. The Computer Program LUDI: A New Method for the De Novo Design of Enzyme Inhibitors. *J. Comput.-Aid. Mol. Des.* **1992**, *6*, 61–78.
(43) Verlinde, C. L. M. J.; Rudenko, G.; Hol, W. G. J. In Search of New Lead Compounds For Trypanosomiasis Drug Design: A Protein Structure-Based Linked-Fragment Approach. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 131–147.
(44) Bernstein, F. C.; Koetzle, T. F.; Williams, G. T. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535–539.
(45) Allen, F. H.; Kennard, O.; Taylor, R. Systematic Analysis of Structural Data as a Research Technique in Organic Chemistry. *Acc. Chem. Res.* **1983**, *16*, 146–153.
(46) Singh, J.; Saldanha, J.; Thornton, J. M. A Novel Method for the Modeling of Peptide Ligands to their Receptors. *Prot. Eng.* **1991**, *4*, 251–261.
(47) Moon, J. B.; Howe, W. J. Computer Design of Bioactive Molecules: A Method for Receptor-Based *de Novo* Ligand Design. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 314–328.
(48) Gibson, K. D.; Scheraga, H. A. Revised Algoriths For the Build-up Procedure For Predicting Protein Conformations by Energy Minimization. *J. Comput. Chem.* **1987**, *8*, 826–834.
(49) Nishibata, Y.; Itai, A. *Tetrahedron* **1991**, *47*, 8985.
(50) Rotstein, S. H.; Murcko, M. A. GenStr 1.0: A Method For De Novo Drug Design. *J. Comput.-Aided Mol. Des.* **1993**. *7*, 23–43.
(51) InsightII, version 2.2, and Discover, version 2.8, Biosym Technologies, San Diego, CA.
(52) Le Grand, S.; Merz, K. A Rapid Analytic Approximation to Molecular Surface Area via the Use of Boolean Logic and Look-Up Tables. *J. Comput. Chem.* In press.
(53) Privalov, P. L.; Gill, S. J. The Hydrophobic Effect: A Reappraisal. *Pure Appl. Chem.* **1989**, *61*, 1097–1104.
(54) Privalov, P. L.; Gill, S. J. Stability of Protein Structure and Hydrophobic Interaction. *Adv. Protein Chem.* **1988**, *39*, 191–234.
(55) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. Extracting Hydrophobic Free Energies from Experimental Data: Relationship to Protein Folding and Theoretical Models. *Biochemistry* **1991**, *30*, 9686–9697.
(56) Murphy, K. P.; Privalov, P. L.; Gill, S. J. *Science* **1989**, *247*, 559–561.
(57) Connolly, P. R.; Varadarajan, R.; Sturtevant, J. M.; Richards, F. M. Thermodynamics of Protein–Peptide Interactions in the Ribonuclease S System Studied by Titration Calorimetry. *Biochemistry* **1990**, *29*, 6108–6114.
(58) Griffith, J. P. Unpublished results.
(59) Open Interface, version 2.1, Biosym Technologies, San Diego, CA.
(60) Fine Chemicals Directory, Molecular Design Limited, San Leandro, CA.
(61) Moore, J. M.; Peattie, D. A.; Fitzgibbon, M. J.; Thomson, J. A. Solution Structure of the Major Binding Protein for the Immunosuppressant FK506. *Nature* **1991**, *351*, 248–250.
(62) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. Atomic Structure of FKBP-FK506, an Immunophilin-Immunosuppressant Complex. *Science* **1991**, *252*, 839–842.
(63) Michnick, S. W.; Rosen, M. K.; Wandless, T. J.; Karplus, M.; Schreiber, S. L. Solution Structure of FKBP, a Rotomase Enzyme and Receptor for FK506 and Rapamycin. *Science* **1991**, *252*, 836–839.
(64) Van Duyne, G. D.; Standaert, R. F.; Schreiber, S. L.; Clardy, J. Atomic Structure of the Rapamycin Human Immunophilin FKBP-12 Complex. *J. Am. Chem. Soc.* **1991**, *113*, 7433–7434.
(65) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. J.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. X-ray Crystallographic Structure of a Complex Between a Synthetic Protease of Human Immunodeficiency Virus 1 and a Substrate-Based Hydroxyethylamine Inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.
(66) 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.; Machin, P. J. Rational Design of Peptide-Based HIV Proteinase Inhibitors. *Science* **1990**, *248*, 358–361.
(67) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. Novel Binding Mode of Highly Potent HIV-Proteinase Inhibitors Incorporating the (R)-Hydroxyethylamine Isostere. *J. Med. Chem.* **1991**, *34*, 3340–3342.
(68) Maren, T. H. General Perspectives and Advances in Glaucoma Research. *Drug Dev. Res.* **1987**, *10*, 255–276.
(69) Eriksson, A. E.; Jones, T. A.; Liljas, A. Refined Structure of Human Carbonic Anhydrase II at 2.0Å Resolution. *Proteins: Struct., Funct. Genet.* **1988**, *4*, 274–282.
(70) Eriksson, A. E.; Kylsten, P. M.; Jones, T. A.; Liljas, A. Crystallographic Studies of Inhibitor Binding Sites in Human Carbonic Anhydrase II: A Pentacoordinated Binding of the SCN^- Ion to the Zinc at High pH. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 283–293.
(71) Merck & Company, Rahway, NJ, 1990 Annual Report, p 36.