PROTEINS: Structure, Function, and Genetics 11:314–328 (1991)

Computer Design of Bioactive Molecules: A Method for Receptor-Based de Novo Ligand Design

Joseph B. Moon and W. Jeffrey Howe
Computational Chemistry, Upjohn Laboratories, Kalamazoo, Michigan 49001

ABSTRACT **The design of molecules to bind specifically to protein receptors has long been a goal of computer-assisted molecular design. Given detailed structural knowledge of the target receptor, it should be possible to construct a model of a potential ligand, by algorithmic connection of small molecular fragments, that will exhibit the desired structural and electrostatic complementarity with the receptor. However, progress in this area of receptor-based, de novo ligand design has been hampered by the complexity of the construction process, in which potentially huge numbers of structures must be considered. By limiting the scope of the structure-space examined to one particular class of ligands—namely, peptides and peptide-like compounds—the problem complexity has been reduced to the point that successful, de novo design is now possible. The methodology presented employs a large template set of amino acid conformations which are iteratively pieced together in a model of the target receptor. Each stage of ligand growth is evaluated according to a molecular mechanics-based energy function, which considers van der Waals and coulombic interactions, internal strain energy of the lengthening ligand, and desolvation of both ligand and receptor. The search space is managed by use of a data tree which is kept under control by pruning according to the energy evaluation. Ligands grown by this procedure are subjected to follow-up evaluation in which an approximate binding enthalpy is determined. This methodology has proven useful as a precise model-builder and has also shown the ability to design bioactive ligands.**

INTRODUCTION

The ability of a molecule, such as a drug, to exert a desired biological effect is often related to its affinity for one or more endogeneous receptor molecules. For a ligand to interact optimally with a receptor, it must be able to attain a shape which is at least partly complementary to that of a binding location on the receptor. Additionally, other factors such as electrostatic interactions, hydrogen bonding, hydrophobic interactions, desolvation effects, and cooperative motions of ligand and receptor all influence the binding event and should be taken into account in attempts to design bioactive ligands. Processes such as distribution and metabolism, while they play a critical role in the delivery of the putative ligand to the receptor location, do not reflect a compound's "intrinsic activity" and lie outside the scope of the current discussion.

In principle, it should be possible to design molecules that will bind to a preselected site on a receptor. This is not a simple undertaking, since in most design situations little or no structural information exists to characterize the receptor. One can, however, use "indirect" methods[1] to exploit what *is* known about molecules that elicit the desired biological response (assuming that they interact with the same receptor) to generate a structural and electronic hypothesis of what the receptor recognizes or will accept. Various computer-based methods have been developed to assist in this kind of study.[1–8] Once the hypothesis has been generated it can be used to suggest molecular modifications to improve the activity of known ligands or to identify entirely new structural classes (lead compounds) for study as potential ligands. The latter can be accomplished via searches over large databases of 3D molecular structures to identify molecules which match the hypothesized requirements for activity.[9–16]

The increasing availability of biomacromolecule structures that have been solved crystallographically has prompted the development of "direct" computational methods for molecular design, in which the steric and electronic properties of receptor binding sites are used to guide the design of potential ligands.[1,11,12,17–19] Direct methods generally fall into two categories: (1) design by analogy, in which 3D structures of known molecules (such as from a crystallographic database) are placed in the receptor structure and scored for goodness-of-fit; and (2) de novo design, in which the ligand model is constructed piecewise in the receptor. The latter approach, in particular, offers considerable promise for the development of novel molecules, uniquely designed to bind to the target.

Received August 23, 1990; revision accepted March 15, 1991.
Address reprint requests to either author, The Upjohn Company, Computational Chemistry, 301 Henrietta St., Kalamazoo, MI 49001.

While examples of successful, computer-assisted, de novo design can be found,[20] there are no examples of automated, or computer-driven, de novo construction in the literature (although Wise et al.[21] have reported using the structure-building program GENOA[22] to generate molecules to match a requirements hypothesis). The term "automated de novo design" is used here to refer to the algorithmic construction of a putative ligand from small fragments, guided by steric and electronic constraints imposed by the receptor, plus appropriate consideration of solvation effects and internal strain energy of the ligand.

In a recent series of papers,[23–26] Dean and coworkers describe a four-step strategy for automated, de novo drug design. Although their goal has not yet been achieved, there has been considerable progress in algorithm development. Furthermore, their studies make clear the complexity of the de novo construction problem as well as the importance of developing noncombinatorial approaches. In our work, we have chosen to focus on one particular region of the large structure-space that is ultimately the design territory of such methods. By confining the search space to consider only amino acids and related fragments as the molecular building blocks, the construction problem has become quite tractable, and we are able to report the first examples of bioactive ligands designed by automated de novo methods. The putative ligands that result from this construction method are peptides and peptide-like compounds rather than the small organic molecules that are typically the goal of drug design research. The appeal of the peptide building approach is not that peptides are preferable to organics as potential pharmaceutical agents, but rather that: (1) they can be generated relatively rapidly de novo; (2) their energetics can be studied by well-parameterized force field methods; (3) they are much easier to synthesize than are most organics; and (4) they can be used in a variety of ways, for peptidomimetic inhibitor design, protein–protein binding studies, and even as shape templates in the more commonly used 3D organic database search approach described above. We also show that the method need not be restricted to just the 20 natural amino acids; it can easily be extended to include other related fragments of interest to the medicinal chemist.

METHODS

Description of the GROW Method

Overview

The de novo peptide design method has been incorporated in a software package called GROW. In a typical design session, standard interactive graphical modeling methods (using the Mosaic software system,[27] which is based on MacroModel[28]) are employed to define the structural environment in which GROW is to operate. The environment could be the active site cleft of an enzyme, or it could be a set of features on a protein surface to which the user wishes to bind a peptide-like molecule. The GROW program then operates independently of the user to generate a set of potential ligand molecules. Interactive modeling methods then come into play again, for examination of the resulting molecules, and for selection of one or more of them for further refinement.

The method is designed to construct peptide models from a user-selected starting position by iteratively piecing together amino acids in conformations which will interact most favorably with the atoms in the receptor site. For input, GROW operates on an atomic coordinate file generated by the user in the interactive modeling session, plus a small fragment (an acetyl group) positioned in the receptor to provide a starting point for peptide growth. These are referred to as "site" atoms and "seed" atoms, respectively. A second file provided by the user contains a number of control parameters to guide the peptide growth.

The operation of the GROW algorithm is conceptually fairly simple, and is summarized in Figure 1. GROW proceeds in an iterative fashion, to systematically attach to the seed fragment each amino acid template in a large preconstructed library of amino acid conformations. When a template has been attached, it is scored for goodness-of-fit to the receptor site, and then the next template in the library is attached to the seed. After all the templates have been tested, only the highest scoring ones are retained for the next level of growth. This procedure is repeated for the second growth level; each library template is attached in turn to each of the bonded seed/amino acid molecules that were retained from the first step, and is then scored. Again, only the best of the bonded seed/dipeptide molecules that result are retained for the third level of growth. The growth of peptides can proceed in the N-to-C direction only, the reverse direction only, or in alternating directions, depending on the initial control specifications supplied by the user. Successive growth levels therefore generate peptides that are lengthened by one residue. The procedure terminates when the user-defined peptide length has been reached, at which point the user can select from the constructed peptides those to be studied further. The resulting data provided by the GROW procedure include not only residue sequences and scores, but also atomic coordinates of the peptides, related directly to the coordinate system of the receptor site atoms. In the following sections we examine in more detail the individual components that comprise the basic procedure just described.

Library construction

Because most amino acids are quite flexible, a large number of template structures must be tested during the growth procedure to ensure adequate

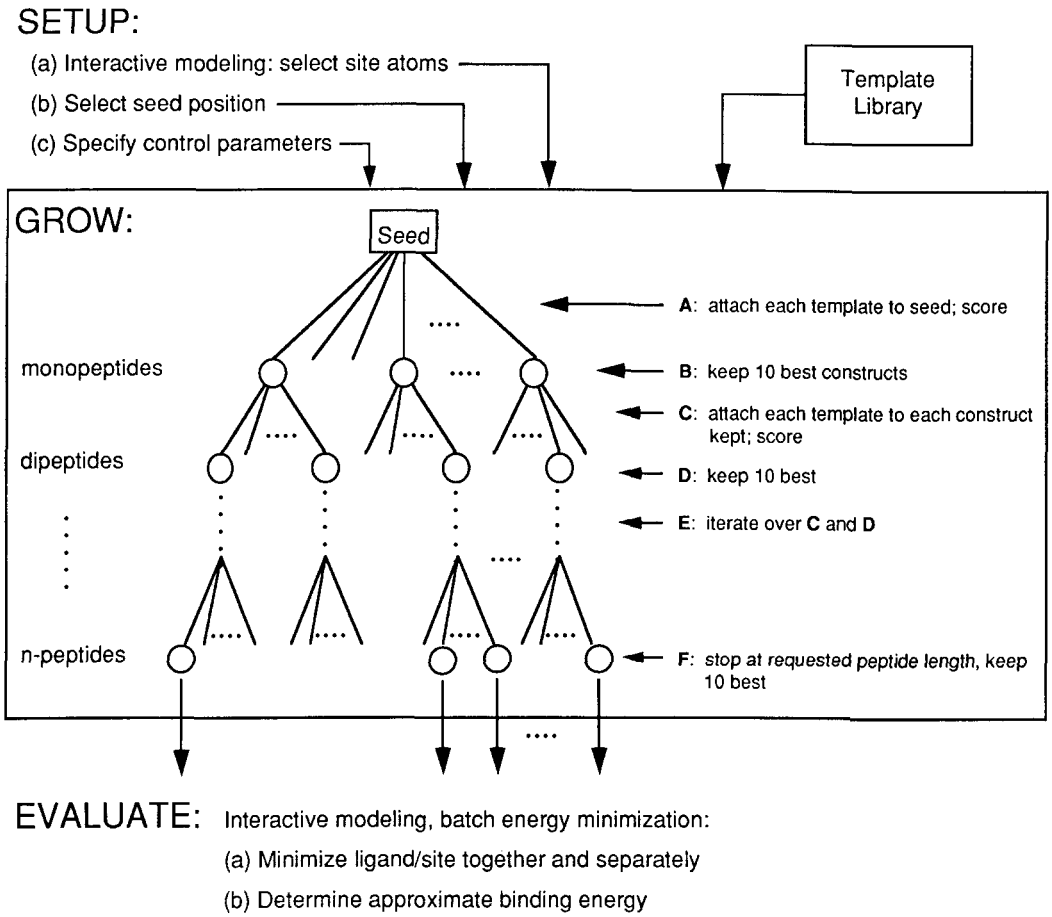

Fig. 1. Schematic overview of the operation of the GROW algorithm. The site-and seed coordinate file and the command file (described later) are provided to the GROW procedure by the user. Growth can be visualized as a tree process in which each library template is attached to the seed **(A)** and then evaluated by the scoring function. Of the resulting 6000+ constructs, the 10 best are kept for the next level **(B)**. 10 is the default retention; a command file keyword can be used to broaden the search at any stage. To each retained monopeptide/seed construct are attached all library templates, which are again scored **(C)**. After pruning **(D)**, the process is repeated **(E)** until the specified peptide length (specified in the command file, see Fig. 5) is reached **(F)**. In this tree diagram, circles represent those nodes selected (based on highest scores across the entire level) for further growth. Uncircled nodes are pruned. Horizontal dots denote continuation across all template additions, and vertical dots represent the iterative process of tree growth.

coverage of the conformational space accessible to each residue. The template library was generated with the Mosaic modeling program in conjunction with the MacroModel/BatchMin[28] (version 2.5) implementation of the AMBER[29] forcefield. The same forcefield implementation was used for all energy-related work described herein. Starting models of the 20 standard amino acids were constructed as *N*-acetyl-*N'*-methylamides (Fig. 2A), followed by energy minimization.* The models were then subjected to a search procedure in which conformers were generated by varying all flexible torsion angles in the amino acids by random increments. Any conformer which contained two nonbonded heavy atoms at a separation of <2.0 Å was discarded. After 3,000 to 5,000 viable conformations were produced for each amino acid, the structures were subjected to a *partial* energy minimization (15 iterations of block diagonal Newton–Raphson minimization) to relieve significant internal strain energies. At this point, each conformation was compared to every other conformation so that duplicate structures would be discarded. Two conformations were considered to be identical if no atomic positions differed by more than 0.3 Å when the structures were aligned by superpositioning of their N-terminal amide atoms. The remaining conformations were sorted in ascending energy order and were stored in the template library along with their energies. Templates of nonstandard amino acids, pseudodipeptides, and organic terminal groups were constructed in the same manner, em-

*Unless otherwise indicated, the convergence criterion used for all energy minimizations discussed in this paper was an rms gradient of <0.1 kcal/Å, with the BatchMin/MacroModel PRCG minimizer.

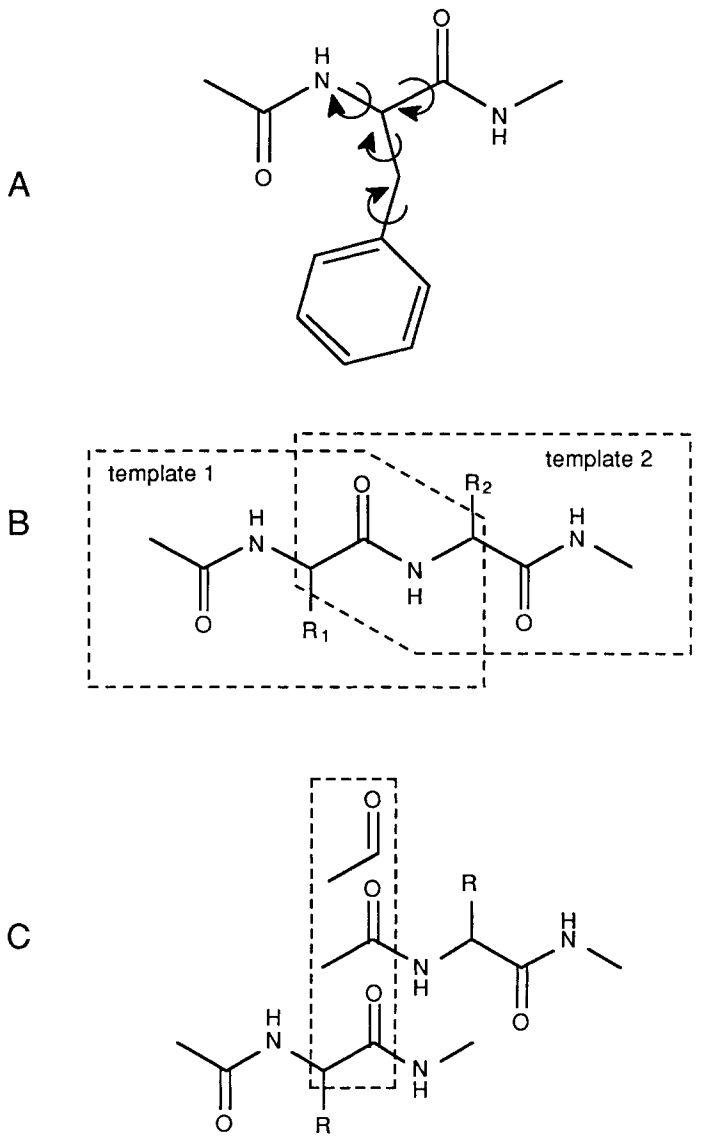

Fig. 2. **(A)** Template generation method using phenylalanine as an example: bonds marked with arrows are rotated by random increments to generate additional conformations. This is followed by contact filtering, partial minimization, and duplicate elimination. **(B)** Template connection method: amide end groups are superimposed to connect two templates together in the proper geometries to form peptides. **(C)** Template alignment method: the two alignments of a template with the seed group are shown. The alignment used depends on the direction in which the peptide is to be grown.

ploying the extended parameter set (in addition to the original[29] AMBER parameters) provided by the MacroModel/BatchMin implementation.

Figure 3 lists the contents of the template library and the number of unique conformations stored for each residue. During a GROW run, from 300 to 1,000 lowest energy conformations are typically utilized for each amino acid; the default is 300. For comparison, values in parentheses indicate the number of *initial* conformations generated for the residues during library construction. Of the 2,000 trial conformations of alanine, for example, partial energy minimization and duplicate elimination reduced the set to 171 unique conformations. As might be expected, this type of reduction in the number of conformations was not seen with the pseudodipeptides and certain of the other residues, due to their extreme flexibility. The implications of template flexibility will be discussed in a later section.

Application of a *partial* energy minimization during library construction produces structures that lie near, but not generally at, energetic minima. Since energetic minima of a bound ligand will not necessarily correspond to minima of an unbound ligand, restriction of templates to unbound minimum-energy conformations represents an unwarranted

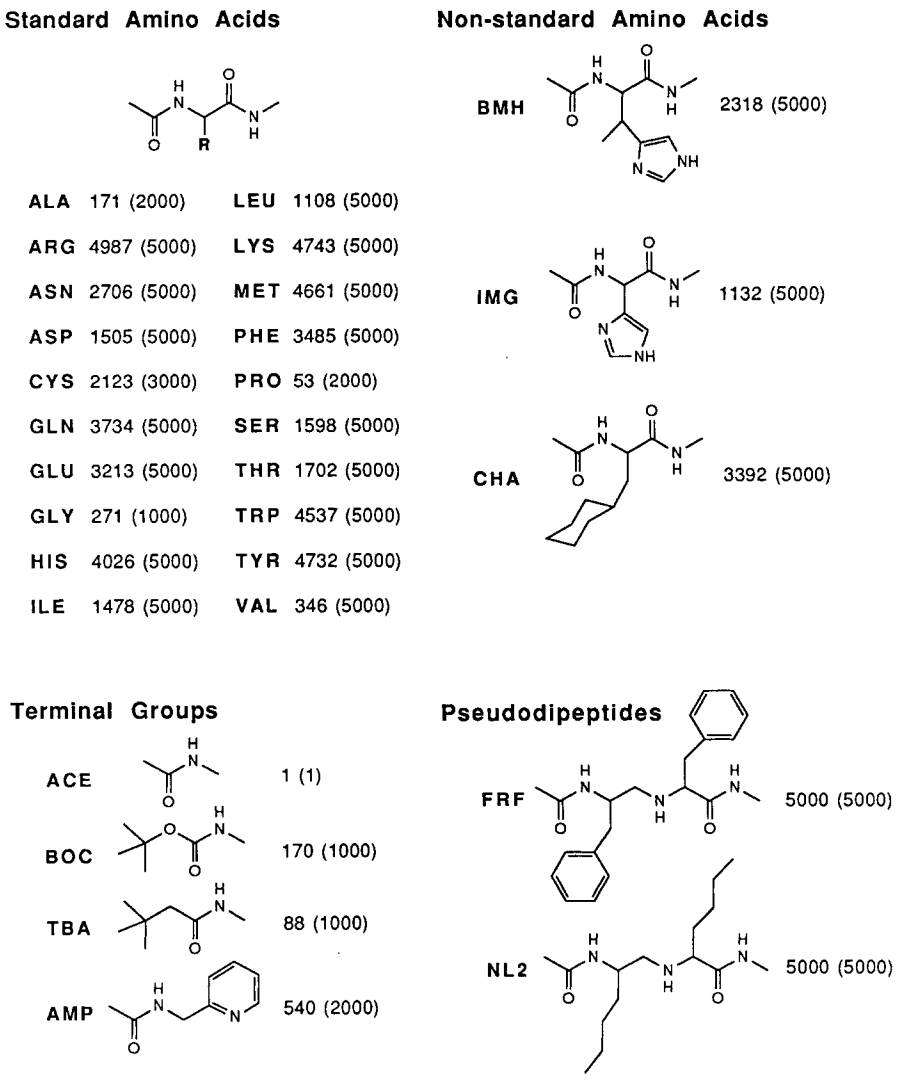

Fig. 3. Contents of the template library. At present, the template library contains standard L- and D-amino acids, several non-standard residues, organic terminators, and pseudodipeptides, some of which are shown here. The table indicates, for each fragment, its 3-character identifier, which can be specified in the control file for running GROW in restricted mode, a parenthesized value which indicates the number of initial conformations generated for that fragment during library construction, and an unparenthesized value which indicates the number of conformations that survived the partial minimization and duplicate elimination steps during library construction. Data shown for standard amino acids apply equally for L- and D-forms.

constraint. The collection of amino acid templates that resulted from the procedure just outlined represents a broad sampling over low-energy conformational space. The assumption made is that such fragments can be connected together to form peptides with low internal conformational energy; adverse interactions *between* residues are dealt with at a later stage.

The acetyl and amide end groups placed on the amino acid models serve two purposes. First, they produce some of the conformational restriction experienced by individual amino acids when they are connected in a polypeptide chain. They also provide a convenient way to connect the templates during peptide construction; two templates can be joined together simply by superimposing the N-terminal amide of one template onto the C-terminal amide of another (Fig. 2B).

Seed fragment positioning

The placement of the seed fragment, while separate from the GROW method itself, has a great influence on the outcome of a GROW procedure. A poorly positioned seed can prevent designed peptides from reaching important interaction sites in the receptor. Because of this sensitivity, we have examined a number of techniques for choosing reasonable seed positions. In the few cases in which an X-ray crystallographic structure of a bound ligand is available, atoms within the ligand can be used to form a

seed position. If the ligand is a peptide, choosing a seed group from the ligand backbone will greatly improve the chances of producing meaningful results. Without an X-ray structure of a receptor–ligand complex, however, other methods must be used to generate seed positions. It is possible with most modeling systems to manually dock an acetyl-containing seed fragment into a receptor site. However, identifying the optimal placement of seed fragments is not a trivial problem. Work is in progress to develop additional methods for this purpose.

The technique that we have found most promising in this situation is to employ an automatic fragment docking algorithm which positions numerous copies of a small amide-containing fragment in the specified site, at locations calculated to provide the strongest interactions between the fragment and site atoms. Originally based on the shape-matching DOCK methodology developed by Kuntz and coworkers,[11,12] the algorithm has been extended to score each potential docking orientation based on van der Waals and electrostatic interactions between a docked fragment and the site atoms. Orientations which score below a specified threshold are discarded. A large number of overlapping copies of the amide-containing fragment generally results from this procedure, tracing out pathways in the site where ligand interactions with the receptor would be strongest. From this map one or more seed positions can be obtained from the location of the amide bonds within the docked fragments. Each seed selected defines a separate GROW run.

Scoring

Before describing the method by which peptides are iteratively constructed during a GROW operation, we first examine the scoring function that is applied as each new library template is attached to a growing peptide. The scoring function is based on potentials from the AMBER force field (as implemented in MacroModel/BatchMin v2.5), with the addition of a solvation treatment developed by Scheraga.[30] When a template has been properly oriented to connect to a growing peptide, a score is calculated which is the sum of five terms:

$$\text{SCORE} = -[E_{\text{vdw}} + E_{\text{es}} + E_{\text{conf}} + E_{\text{solv}}(\text{templ}) + E_{\text{solv}}(\text{rec})].$$

The more negative the individual energy terms, the greater the estimated binding affinity of the template, and the higher the score.

E_{vdw} is the van der Waals energy calculated between the positioned template and the receptor atoms, and between the template and peptide atoms already grown in the receptor site. This is calculated using a modified Lennard–Jones 6–12 potential with BatchMin's AMBER parameters, for atoms within 6 Å of the template. A small amount of "free play" (default amount = 0.1 Å) has been incorporated into the repulsive term of this potential, so that van der Waals penetrations of less than that amount are not penalized. This softening of the repulsion is necessary to compensate for the use of template sets which are not continuous through conformational space. The degree of softening is controllable by the user, via a parameter passed in the control file.

E_{es} is the coulombic electrostatic interaction between the template atoms and those in both the receptor site and the portion of the ligand already grown. Again, a 6 Å cutoff is employed for the sake of speed. E_{conf} is the conformational strain energy of the template. This is a precalculated value which represents the energy difference between the template's conformation and the lowest energy conformation found, among the 3,000–5,000 conformations examined, for that particular amino acid during library creation. The value is retrieved from the library during ligand growth. $E_{\text{solv}}(\text{templ})$ and $E_{\text{solv}}(\text{rec})$ are solvation terms calculated with Scheraga's method[30]: they depend on the changes in solvent-accessible surface area caused by moving the template and receptor from their fully hydrated unbound state to their partially hydrated bound state. The solvation terms favor the hydration of polar atoms, and penalize hydration of hydrophobic atoms, thereby simulating a hydrophobic binding effect. Solvent-accessible surface areas are determined by the approximate method of Still,[31] because of its tremendous speed advantage over analytical calculations. To further enhance the speed of the scoring procedure, the unbound solvation energies of the templates and of the receptor are calulated only once, the former during library construction and the latter at the beginning of the GROW operation. Only the changes in these energies are calculated during scoring.

Ligand growth

Having described the nature of the template library and the method by which templates are evaluated during ligand construction, we can now examine the ligand growth procedure, per se. To begin, GROW aligns all of the library templates in turn with the seed fragment. This superimposition can be accomplished in either of two ways, as indicated in Figure 2C. For the sake of this discussion, if one considers just the 20 standard amino acids, with on average 300 library conformations for each, a total of 6,000 templates must be evaluated. These are retained on a "data tree" which, by this point, contains one node for the seed and 6,000 branches, each one representing a template (see Fig. 1). Each node is scored according the method described in the preceding section. If GROW were to continue in this manner, an exhaustive search for the best *dipeptide* would involved 3.6×10^7 evaluations, and $1.3 \times$

10^{15} evaluations would be needed to find the best *tetrapeptide*. Instead, the program retains and grows from only a certain number of the highest scoring branches, usually from 10 to 100 (the default is 10), after each tree-level of growth. All other branches are pruned from the tree. This cycle of attaching templates, scoring templates, and pruning structures is repeated until peptides of the desired length are built. Because of the pruning step, the conformational search performed by GROW covers a very small fraction of the total number of possible conformations, but is heavily biased toward peptides which have low conformational energy and which will interact most favorably with the receptor site. An attractive feature of this tree structure is that increasing the number of templates used by the program, or the length of the peptides grown, results in only a linear increase in the processor time required for the calculations.

The ligands grown by this procedure may have the original seed location at either the C- or N-terminal end, or somewhere in the middle of the peptides, depending on which directional control option was specified by the user. In follow-up visual examinations to determine how well the peptides appear to fit the target site, it has been our experience that the match is extremely good, often rivalling or surpassing the visual fit in X-ray structures of inhibitors and enzymes. It has also been our experience, however, that visual fit can be highly misleading. For that reason we routinely subject the results of a GROW analysis to more detailed evaluation.

Evaluation of Results

Although the scoring method used by GROW is very effective for guiding the pruning process during peptide growth, it is still only a rough estimator of binding affinity. Before the synthesis and testing of any structures designed by GROW are warranted, it is necessary to further prune the results by more detailed energy-based methods. For this we use a two-part process. First, the peptide/receptor complex, the unbound receptor, and the unbound peptide are subjected to energy-based optimizations with Batchmin's AMBER implementation and Scheraga's solvation model. From these energies, an estimation of the energy of binding can be made:

$$E(\text{binding}) = E(\text{complex}) - E(\text{unbound receptor}) - E(\text{unbound peptide}).$$

If this estimated binding energy is unfavorable, the structure can be rejected and the next one tested, or a new seed position can be selected for another GROW procedure. Otherwise, a second step is carried out, in which conformational analysis of the unbound peptide is performed to attempt to find its lowest energy conformation. This can be accomplished relatively quickly using simulated annealing methods.[32] If conformations are found which are so low in energy that the estimated binding energy is no longer favorable, the structure is rejected. In other words, if too great an energy penalty must be paid to take the ligand from its low-energy solution conformation(s) to a solution conformation which is close to that of its bound state, then this energy cost can more than offset any energy gained due to binding.

This follow-up energy evaluation is not actually part of the GROW algorithm and could be replaced by more lengthy calculations of binding free energy. Regardless of the method used, however, it is essential that it take into account the effects of solvation on peptide/receptor interaction. These effects are quite large, and in most cases represent a major driving force for binding. It is for the same reason that the solvation terms have been included in GROW's scoring function.

Growth Options and Control

If the GROW algorithm is allowed to consider on an equal basis each member of the template library during template selection then it is effectively sampling very broad regions of the conformational space accessible to each residue. This method of operation is referred to as "unrestricted growth." There are also instances in which it is desirable to restrict the ligand design process, to guide the outcome to satisfy certain constraints. For example, if the structure of an enzyme-peptide complex is known, it has proven useful to have GROW design new ligands which have the same general conformation as the known structure, but with different amino acid sequences. This is referred to as "restricted growth." The appropriate restriction parameters (in this case, the backbone ϕ and ψ angles of the known structure) are specified through the use of keywords in the command file passed to the program along with the site-and-seed coordinate file mentioned earlier. During peptide growth, only those templates that satisfy the specified constraints are selected for scoring and attachment to the evolving ligand. If an active peptide ligand's sequence is known, but its bound conformation is not, it can be useful to specify that sequence and let GROW generate feasible binding conformations for the peptide. In this case, GROW functions more as a ligand model-building tool than as a ligand design program. In both of the preceding examples of restricted growth, the procedure usually takes less than 5 min (on a VAX 8800 computer) to generate peptides of length 6–8 residues. Unrestricted growth, such as would be appropriate in the common situation where the structures of bound ligands are not known, generally takes 40–50 min for similar-length peptides.

The user may also control (through command file keywords) the shape of the search tree that defines the number of peptides to be retained at each stage

Command	Parameter Description
DIEL d	Dielectric constant to be used in the electrostatic energy calculations.
NRES n	Length of the peptide (number of residues) to be grown.
ORDR $p_1,p_2,...p_n$	A numerical sequence specifying the order in which the residues are to be connected.
RES $r_1,r_2,...r_n$	Residue sequence to be grown. r_i is the name of a residue to be considered at position *i*. Multiple lines may be used to specify alternate residue choices for each position.
SUB $s_1,s_2,...s_n$	Substitute residues allowed for each position during an annealing run. Same format as for RES keyword. Used only in conjunction with ANNL.
CONF $c_1,c_2,...c_n$	Specific conformation numbers to be used at each position. Used in conjunction with the RES keyword to force the program to incorporate a particular template/conformation into the growing peptide.
DIR x	Growth direction. x=1: grow from N to C direction. x=-1: grow from C to N direction. x=2: grow in alternating directions. These three growth directions can also be specified using the ORDR keyword.
NINB	Causes intramolecular non-bonded interactions to be omitted from the scoring.
PHIS $\phi_1,\phi_2,...\phi_n$	Designates allowable backbone phi angles for each position in the peptide.
PHTL $\delta\phi_1,\delta\phi_2,..\delta\phi_n$	Tolerances for backbone phi angles. Used in conjunction with PHIS keyword.
PSIS $\psi_1,\psi_2,...\psi_n$	Similar to PHIS, but specifies allowable backbone psi angles.
PSTL $\delta\psi_1,\delta\psi_2,..\delta\psi_n$	Backbone psi angle tolerances.
BRNC $b_1,b_2,...b_n$	Number of branches (peptides) to be retained by the program after each position is grown.
NROT res,nconf	Directs the program to use *nconf* conformations of residue *res* from the template library (e.g. NROT ARG, 1000).
NBFF f	Value in Angstroms by which the Lennard-Jones potential function is softened during the scoring.
ANNL n,t_1,t_2,t_m	Directs the program to anneal the highest scoring grown peptide. n = number of perturbations to be made at each temperature. t_1 = initial temperature, t_2 = final temperature, t_m = factor by which the temperature is reduced after each step.

Fig. 4. List of command file keywords and their parameters. The command file is created by the user to control the operation of GROW. Examples of two different command files are shown in Figure 5.

of growth. The more branches that are retained at each stage, the more thorough is the search for peptides, and retaining a large number of branches often provides the program with the ability to keep templates which initially score poorly, but which ultimately allow higher scoring peptides to be formed in subsequent stages. As desirable as this improved look-ahead may be, however, it must be balanced against the increased length of time required for the search. A strategy which seems to work well is to keep a large number of branches in the early stages of growth, when the look-ahead capability is most crucial, then decrease the number of branches retained in subsequent stages. This serves as an effective compromise between thoroughness and speed.

The complete set of GROW's command file keywords and associated parameters is shown in Figure 4. Figure 5 presents two sample command files to illustrate how the more commonly used commands can be set up for an unrestricted growth and a restricted growth, respectively.

RESULTS AND DISCUSSION

Although the ultimate validation of this method is, and will continue to be, its ability to generate active compounds de novo, we have also sought other means of validation that are more conducive to detailed structural examination. One such method is to employ GROW in its model-building mode (restricted growth) to attempt to duplicate the crystal structure conformations of inhibitors bound to enzymes. The first two examples below not only demonstrate GROW's ability to closely match experimentally determined binding conformations but also point out some of the difficulties inherent in the use of highly flexible pseudodipeptide templates.

A:

```
INFILE
OUTFILE
NRES    4
RES     ANY   ANY   ANY   ANY
DIEL    10
ORDR    4     3     1     2
BRNC    10    10    100   10
```

B:

```
INFILE
OUTFILE
NRES    4
DIEL    10
DIR     1
RES     HIS   PHE   CYS   ANY
RES           TRP
RES           TYR
RES           LEU
PHIS    0     -115  0     0
PHTL    180   30    180   180
PSIS    0     120   0     0
PSTL    180   30    180   180
```

Fig. 5. Two sample command files. Both start with file names which define the source file containing the site and seed coordinates and the destination file into which are written the grown peptide atoms and coordinates. Both also specify a dielectric constant to be used in the scoring function. **(A)** Sample control file for an unrestricted run. NRES specifies that tetrapeptides are to be grown, RES specifies that any (ANY is shorthand for any of the standard 20 L-amino acids) residue can occupy each of the positions, and ORDR specifies the growth order. The column positions in the ORDR row match the column positions in the RES row. The "1" in the third column of the ORDR row indicates that in the R1–R2–R3–R4 peptides to be grown, growth is to start at the third residue, or R3. That is, templates are first attached at the (default) C-terminal side of the seed. The "2" means that the next one gets added to the C-terminus end of the seed/monopeptide constructs. The "3" and "4" indicate that subsequent growth goes toward the N-direction, from the seed position. The resulting tetrapeptides will all have the seed location at the R2–R3 peptide bond. The BRNC row also matches columns in the RES row. It specifies that the highest scoring 100 constructs are to be retained for the first residue grown, and thereafter the best 10 are kept at each subsequent level. **(B)** Control file for a partially restricted GROW operation. The four RES rows specify that 10 (default) tetrapeptides (NRES 4) with sequences His-X-Cys-Y are to be generated, where X = Phe, Trp, Tyr, or Leu and Y = any residue. Templates selected for the second residue must have a backbone dihedral angle pair (ϕ,ψ) of (-115 ± 30, 120 ± 30). Growth will occur in the N-to-C direction from the seed (DIR 1).

The remaining three examples illustrate unrestricted growth.

Growth Examples

The Brookhaven Protein Data Bank[33] (PDB) entry 3APR is a 1.8 Å structure[34] of the aspartyl protease enzyme rhizopuspepsin bound to the inhibitor (D)His-Pro-Phe-His-PheΨ[CH_2NH]Phe-Val-Tyr. The PheΨ[CH_2NH]Phe insert is a "reduced amide" pseudodipeptide derived from Phe–Phe, in which the peptide carbonyl group is replaced by a methylene, rendering the C′–N bond invulnerable to hydrolysis. For the initial computational experiment, a PheΨ[CH_2NH]Phe template set was generated (as described above) and added to the template library. GROW was then operated in restricted growth mode, using the sequence of the inhibitor [minus the terminal (D)His and Tyr residues, which were missing from the X-ray structure]. The peptide linkage between the inhibitor's His and Phe was provided as the seed. Initially, GROW was unable to find any conformations of the pseudodipeptide template which could fit into the active site. On further examination it was discovered that of the 5,000 conformations generated for this fragment during template construction, none corresponded to the conformation observed in the X-ray structure. The fragment contains 9 variable torsion angles and the probability of generating a particular conformation by random assignment of 9 angles is apparently too small. This is likely to be a general problem with the use of extremely flexible structures as library templates. A more effective strategy under development is to break highly flexible templates into smaller units before including them in the library.

The experiment was repeated, with the crystal structure conformation of the PheΨ[CH_2NH]Phe

Fig. 6. Stereo diagrams showing the first three stages in the sequence-restricted growth of the peptide Pro-Phe-His-PheΨ[CH_2NH]Phe-Val (blue) in the rhizopuspepsin active site cleft. For clarity, enzyme atoms are not shown. The crystal structure of the bound inhibitor (red) has been overlaid to provide a visual comparison with the conformations selected by GROW. No knowledge of the red molecule, other than its sequence, is used during the growth procedure. Each diagram shows the 10 highest scoring peptides that result from each stage of growth. **(A)** After the first stage, GROW has found three regions for placement of the 10 His conformations. All other conformations sampled have been filtered out in the scoring. One of the regions corresponds to that observed in the crystal structure. **(B)** By the end of the second stage only one of the His placements has survived. All ten of the Phe conformations are close to that seen in the X-ray structure, although GROW also finds sufficient room in the enzyme's cleft to accommodate a perpendicular orientation of the phenyl ring. **(C)** The tight clustering of the Pro placements is apparently due to steric congestion in the enzyme site and, therefore, to greater selectivity. The 10 highest scoring tripeptides by this point include the single His conformation and two Phe conformations. Growth continues in this manner, to select a PheΨ[CH_2NH]Phe template, followed by Val conformations at the other end of the peptide. The minimized, fully grown peptide is shown in Figure 7.

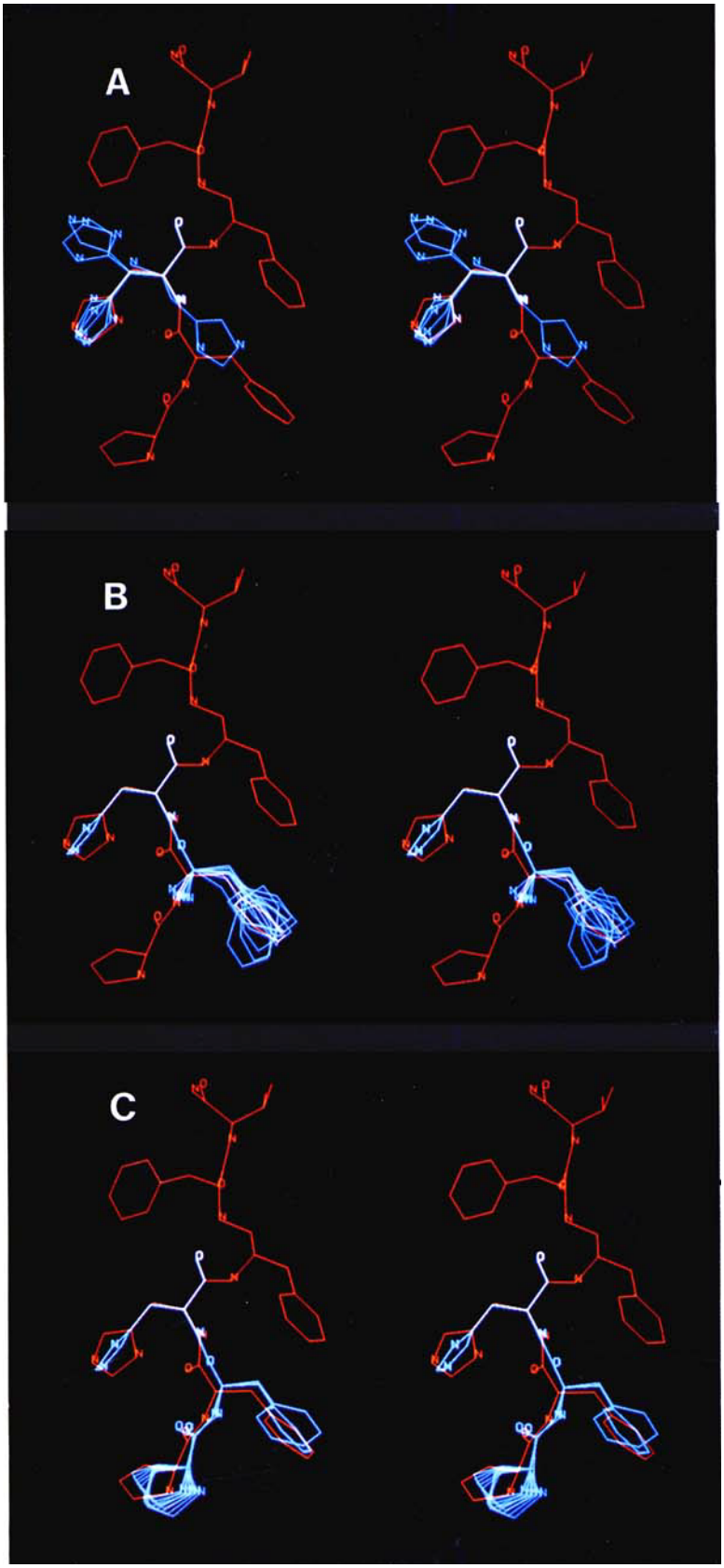

Fig. 6 A–C. Legend appoears on page 322.

fragment explicitly included in the library. GROW selected the appropriate template for the pseudodipeptide and generated 10 conformations of the specified hexapeptide, the highest scoring of which closely reproduced the observed inhibitor conformation. The successive stages of this growth operation are illustrated in Figure 6A–C. The rms deviation of the highest scoring structure from the X-ray structure (excluding the pseudodipeptide coordinates) was 0.87 Å. When both structures were subjected to energy minimization in the enzyme active site, the rms deviation was reduced to 0.60 Å (Fig. 7). Importantly, all hydrogen bonding interactions between the inhibitor and the protein observed in the X-ray structure were duplicated in the grown inhibitor.

A similar study was carried out with the 2.3 Å crystallographic structure[35] of the aspartyl enzyme HIV-1 protease, bound to the reduced amide inhibitor MVT-101 (Ac-Thr-Ile-NleΨ[CH_2NH]Nle-Gln-Arg-NH_2, Nle = norleucine). Because of the lower resolution of the crystallographic data, we subjected the inhibitor and all enzyme atoms within 10 Å of the inhibitor to energy minimization before starting a GROW operation. As in the case of the PheΨ[CH_2NH]Phe fragment above, the NleΨ[CH_2NH]Nle group, with 11 variable torsion angles, proved too flexible for making templates. The conformation of this fragment was taken from the minimized structure and used as a template. GROW was given the coordinates of the Ile-Nle peptide bond as a seed position, and was restricted to the MVT-101 sequence, with no conformational restrictions. The highest scoring conformation that resulted had an rms deviation of 1.26 Å from the minimized X-ray structure. Energy minimization of the grown inhibitor reduced the rms deviation to 0.84 Å (Fig. 8). Again, all enzyme–inhibitor hydrogen bonding interactions were preserved in the grown inhibitor.

The fact that this methodology can closely duplicate experimentally observed results when operating in restricted mode points to its potential utility as a precise, automatic model-building tool. In fact, we have found the method to be particularly useful for this purpose in the design of enzyme inhibitors, given one crystal structure of an inhibitor-bound enzyme. With the known inhibitor acting as a seed source, GROW (plus the associated energy evaluation) has been used in a prospective manner to suggest new lead structures for synthesis and in a retrospective manner to rationalize the receptor interactions of previously synthesized molecules, both active and inactive.

Its ability to function well in *restricted* mode provides a degree of assurance that GROW can generate useful structures in *unrestricted* mode, as well. This has been demonstrated indirectly in a study of residue variability at one particular position in HIV-1 protease substrates. An unrestricted GROW operation was performed on the MVT-101/HIV-1 protease structure, and the program was allowed to replace the inhibitor's P_2 isoleucine residue with any of the 20 standard amino acids (in *N*-acetyl form). The seed position and site atoms employed in the previous restricted-mode experiment were also used here. The 30 highest-scoring constructs were retained, and included multiple conformations of just eight different residues (listed in Fig. 9).

For comparison, Figure 9 also lists the 11 P_2 residues that have been observed in the more than 40 substrate sequences known[36–39,46] to be cleaved by the protease in a variety of viral and nonviral protein substrates. Seven of the eight "predicted" residues have also been observed at that position in substrates, and of the 12 "nonpredicted" (low-scoring) residues, eight have never been observed experimentally at the P_2 position. The contingency table constructed from the data in Figure 9 yields[40] a χ^2 value of 5.69. Application of the χ^2 test indicates that the probability of obtaining this degree of association by chance is less than 2%. A more detailed study is under way to examine the HIV-1 protease substrate/inhibitor residue preferences at all six positions in the P_3–P_3' range.

In another experiment with the 3APR structure discussed earlier, GROW was applied to the design of a peptide ligand that would fit the rhizopuspepsin active site. The seed position was the same as was used earlier and a peptide length of 8 was specified. No sequence or conformational restrictions were applied. In approximately 40 min, GROW generated 10 peptides, each of which demonstrated excellent shape complementarity to the binding site. The peptides were then subjected to the energy evaluation procedure and the octapeptide Ac-Asn-Gly-Trp-His-Phe-Thr-Asn-Asn-NH_2 was ranked as the highest (with the Phe-Thr bond at the scissile location). Molecular graphics analysis revealed that the conformation generated by GROW for this peptide was able to form 9 hydrogen bonds to the enzyme active site. The peptide was subsequently synthesized, and did show ability to bind to the enzyme, as evidenced by substrate activity.[‡] Because multiple cleavages occurred, additional studies are planned to determine whether, and at what point in the multiplicity of events, the predicted Phe-Thr hydrolysis took place.

GROW has also been applied to the design of peptides to inhibit the aspartyl protease renin. At the time of the experiment, no X-ray crystallographic structure of renin was available, so the study employed a structural model developed by Carlson et al.[41] through homology-based model building methods. Seed selection made use of a model of the renin inhibitor Ac-Pro-Phe-His-PheΨ[CH_2NH]Phe-NH_2 which had previously been constructed in the active site using molecular graphics and molecular dynamics techniques.[42,43] A seed position taken from the Pro–Phe amide of the modeled inhibitor was pro-

[‡]The peptide was synthesized using standard solid-phase chemistry methods. L. L. Maggiora (unpublished results).

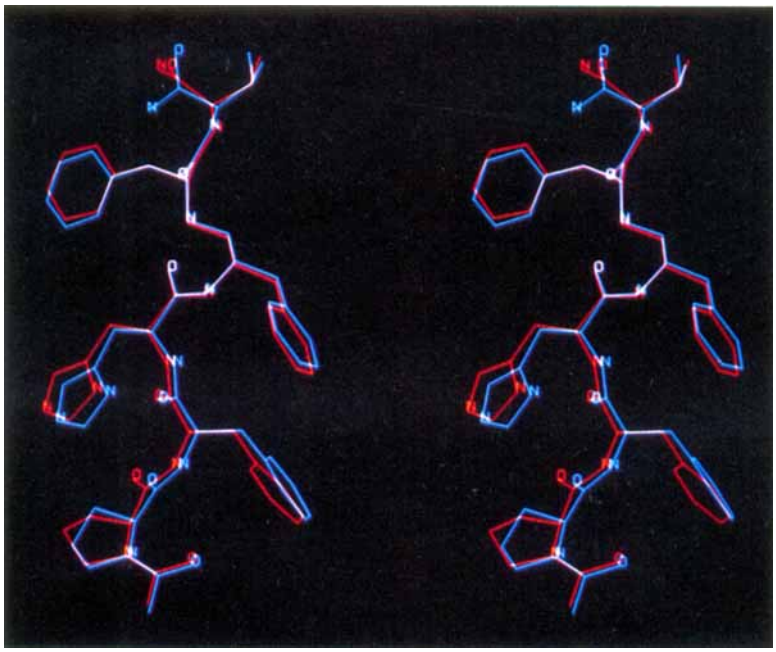

Fig. 7. Stereo diagram comparing the highest-scoring peptide conformation generated by GROW (blue) with the conformation of the same rhizopuspepsin inhibitor obtained from the 3APR X-ray structure (red). Both molecules have been energy minimized in the enzyme active site. Some differences between the minimized and unminimized versions can be noted (see also Fig. 6C), especially in the conformation of the His and Pro residues. The rms deviation between the X-ray and grown structures was 0.87 Å before minimization and 0.60 Å in the minimized forms shown here.

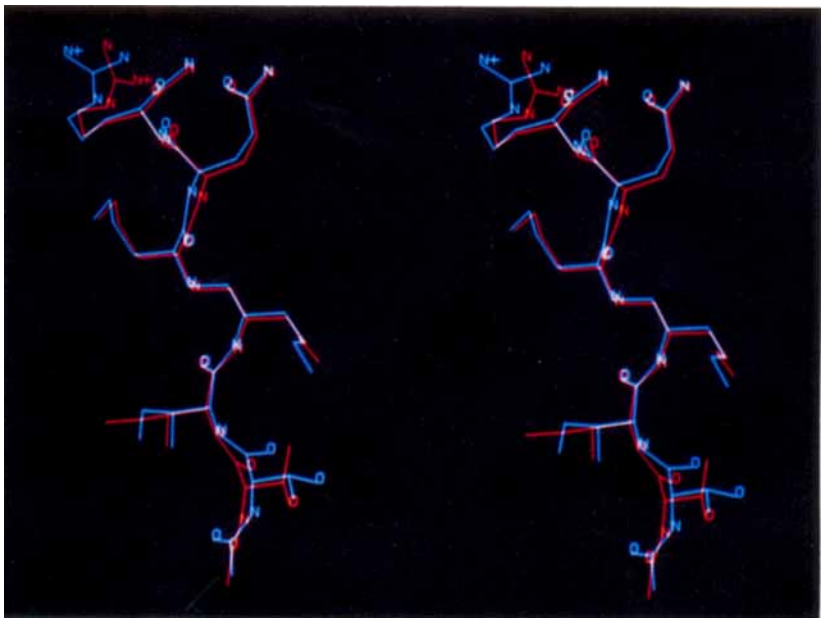

Fig. 8. Stereo diagram comparing the highest scoring conformation of the HIV-1 protease inhibitor, MVT-101, generated by GROW (blue), and the crystal structure of the same molecule in its enzyme-bound form (red). Both have been energy minimized in the enzyme's active site, resulting in an rms deviation of 0.84 Å.

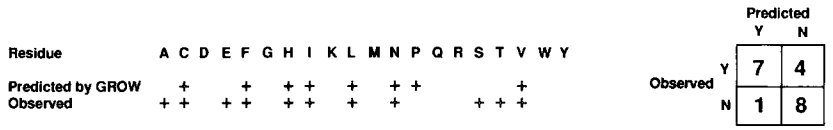

Fig. 9. Residue variability accommodated by the HIV-1 protease enzyme structure at the P_2 position, using the MVT-101 P_2–P_1 amide as the seed (see text). Top row: the eight residues selected by GROW (in multiple conformations) in the 30 top-scoring constructs. Bottom row: all residues that have been observed to date at the P_2 position in protein substrates of the HIV-1 protease.[36–39,46] Right side: 2 × 2 contingency matrix summarizing the "predicted vs. observed" data.

vided to GROW for an unrestricted growth procedure. The highest scoring sequence produced was Ac-Pro-Trp-Trp-Phe-Arg-Arg-NH_2 (the Ac-Pro fragment was not "grown," but was retained from the model). This compound was subsequently found to inhibit renin with $K_i = 30$ μM.[§] It should be noted that this all-peptide-bond inhibitor was designed in under an hour from an approximate enzyme model, without application of existing structure–activity data, and still demonstrated a binding affinity comparable to that seen in molecules patterned after the human renin substrate. By way of comparison, the synthetic tetradecapeptide based on human angiotensinogen has a K_m of 21 μM.[45]

Limitations and Extensions

Although the techniques described here have shown marked potential for de novo ligand design, there are clearly some parts of the procedure in which improvements could be made. One of the limitations of the method lies in the use of a finite number of discrete template conformations to represent the continuous space accessible to flexible molecules. Increasing the number of template conformations and/or the number of branches retained in GROW's data tree will generally improve the results of a GROW operation. Ideally, both of these numbers should be large. However, there is a limit in the extent to which these parameters may be increased without slowing the program's execution to an unacceptable speed. To address this limitation we have recently developed, and are currently investigating, a form of simulated annealing[32] to further optimize the fit of peptides designed by GROW. Rather than simply varying the conformation of the GROW'n molecule, the annealing procedure perturbs the molecule by replacing residues in the molecule with randomly selected templates from the library. Gradual lowering of the "temperature" used to compute the Boltzmann acceptance factor constrains the search into increasingly higher scoring regions of peptide-sequence space. This extension offers at least a partial way around the look-ahead problem, as it shows substantial improvements in score, especially for ligands larger than four residues constructed with the unrestricted growth option.

Another area which is receiving close attention is the scoring procedure, the accuracy and inclusiveness of which have a direct bearing on the quality of the results. Since current molecular mechanics techniques provide only moderate accuracy for estimating ligand binding energies, and since other factors that influence binding have been neglected, such as entropy and the cooperative motions of ligand and receptor, the present scoring function provides only a very approximate assessment of binding capability. More efficient methods for handling highly flexible templates, such as pseudodipeptides and organic fragments, are also clearly desirable. The technique used by the program for connecting templates by superimposing amide end groups can be generalized to allow connection of templates by various other functional groups. Work is currently under way to develop template libraries based on small organic fragments. As this work progresses to include fragments more "exotic" than the simple (and relatively well-parameterized) ones discussed here, it will be even more important to have accurate and reliable scoring function components.

CONCLUSIONS

Despite these current limitations, however, sufficient evidence has accumulated during our use of the techniques described here to establish the methodology as an extremely useful adjunct to other computational techniques for structural studies and for the generation of bioactive agents. It should be restated that, within the larger framework of de novo molecular design methods, the GROW approach currently addresses only one region of "structure space," as it is restricted at present to peptides, pseudopeptides, and related structures containing organic end groups. Methods for exploring the much larger, and potentially much more important areas of structure space that include nonpeptidic organic molecules, are currently being studied as extensions of the approach presented here. And finally, although any de novo construction approach requires a detailed description of the receptor molecule, such information is becoming available at an increasing

[§]L. L. Maggiora and J. Hinzmann, unpublished results. The peptide was assayed for renin inhibition according to the method described in ref. 44.

rate. This, in turn, will continue to increase the opportunities for application of receptor-based design methods.

ACKNOWLEDGMENTS

We thank L. L. Maggiora, T. K. Sawyer, and R. A. Poorman for helpful discussions related to synthesis and enzyme kinetics, J. R. Blinn for development of some of the energy methods used in these studies, M. A. Johnson for statistical suggestions, and R. L. Heinrikson for discussions on substrate specificities. We also thank A. Wlodawer for X-ray crystallographic coordinates of the synthetic HIV-1 protease/MVT-101 inhibitor complex, prior to their availability in the PDB.

REFERENCES

1. Cohen, N.C., Blaney, J.M., Humblet, C., Gund, P., Barry, D.C. Molecular modeling software and methods for medicinal chemistry. J. Med. Chem. 33:883–894, 1990.
2. Marshall, G.R., Barry, C.D., Bosshard, H.E., Dammkoehler, R.A., Dunn, D.A. The conformational parameter in drug design: The active analog approach. In: "Computer-Assisted Drug Design," Amer. Chem. Soc. Symp. Series. Vol. 112. Olsen, E.C., Christoffersen, R.E. eds. Washington: American Chemical Society, 1979: 205–226.
3. Cramer, R.D., III, Patterson, D.E., Bunce, J.D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc. 110:5959–5967, 1988.
4. Doweyko, A.M. The hypothetical active site lattice. An approach to modelling active sites from data on inhibitor molecules. J. Med. Chem. 31:1396–1406, 1988.
5. Crippen, G.M. Distance geometry analysis of the benzodiazepine binding site. Mol. Pharmacol. 22:11–19, 1982.
6. Mabilia, M., Pearlstein, R.A., Hopfinger, A.J. Computer graphics in molecular shape analysis. In: "Topics in Molecular Pharmacology," Vol. 3. Burgen, A.S.V., Roberts, G.C.K., Tute, M.S. eds. Amsterdam: Elsevier, 1986: 158–182.
7. Dean, P.M. "Molecular Foundations of Drug-Receptor Interaction." Cambridge: Cambridge Univ. Press, 1987.
8. Sheridan, R.P., Venkataraghavan, R. New methods in computer-aided drug design. Acc. Chem. Res. 20:322–329, 1987.
9. Gund, P., Wipke, W.T., Langridge, R. Computer searching of a molecular structure file for pharmacophoric patterns. In: "Computers in Chemical Research, Education, and Technology," Vol. 3. 1974: 5–21.
10. Y. C. Martin, M. G. Bures, P. Willett. In: "Reviews in Computational Chemistry," Vol. 1. Lipkowitz, K., Boyd, D. eds. VCH Publ., 1990.
11. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R., Ferrin, T.E. A geometric approach to macromolecule-ligand interactions. J. Mol. Biol. 161:269–288, 1982.
12. DesJarlais, R. L., Sheridan, R. P., Seibel, G.L., Dixon, J.S., Kuntz, I.K. 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. 31:722–729, 1988.
13. Sheridan, R.P., Venkataraghavan, R. Designing novel nicotinic agonists by searching a database of molecular shapes. J. Comp.-Aided Mol. Design 1:243–256, 1987.
14. Brint, A.T., Willett, P. Pharmacophoric pattern matching in files of 3-D chemical structures: comparison of geometric searching algorithms. J. Mol. Graphics 5:49–56, 1987.
15. Sheridan, R.P., Rusinko, A., III, Nilakantan, R., Venkataraghavan, R. Searching for pharmacophores in large coordinate data bases and its use in drug design. Proc. Natl. Acad. Sci. U.S.A. 86:8165–8169, 1989.
16. vanDrie, 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-dimensional molecular structures. J. Comp.-Aided Mol. Design 3:225–240, 1989.
17. 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," Spec. Publ., Roy. Soc. Chem. 78:182–196, 1989.
18. Goodford, P.J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J. Med. Chem. 28:849–857, 1985.
19. 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. 29:2149–2153, 1986.
20. Ripka, W.C., Sipio, W.J., Blaney, J.M. Molecular modeling and drug design: strategies in the design and synthesis of phospholipase A_2 inhibitors. Lect. Heterocycl. Chem. IX: S95–S104, 1987.
21. Wise, M., Cramer, R.D., Smith, D., Exman, I. Progress in three-dimensional drug design: the use of real time color graphics and computer postulation of bioactive molecules in DYLOMMS. In: "Quantitative Approaches to Drug Design," J. C. Dearden, ed. Amsterdam: Elsevier, 1983: 145–146.
22. Carhart, R.E., Smith, D.H., Gray, N.A.B., Nourse, J.G., Djerassi, C. GENOA: A computer program for structure elucidation utilizing overlapping and alternative substructures. J. Org. Chem. 46:1708–1718, 1981.
23. 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 Ser. B 236:101–113, 1989.
24. Danziger, D.J., Dean, P.M. Automated site-directed drug design: The prediction and observation of ligand point positions at hydrogen-bonding regions on protein surfaces. Proc. R. Soc. London Ser. B 236:115–124, 1989.
25. Lewis, R.A., Dean, P.M. Automated site-directed drug design: The concept of spacer skeletons for primary structure generation. Proc. R. Soc. London Ser. B 236:125–140, 1989.
26. Lewis, R.A., Dean, P.M. Automated site-directed drug design: The formation of molecular templates in primary structure generation. Proc. R. Soc. London Ser. B 236:141–162, 1989.
27. Howe, W.J., Blinn, J.R., Moon, J.B., Hagadone, T.R., White, G.J., Schultz, M.W., to be submitted.
28. Mohamadi, F., Richards, N.G.J., Guida, W.C., Liskamp, R., Lipton, M., Caufield, C., Chang, G., Hendrickson, T., Still, W.C. MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. J. Comp. Chem. 11:440–467, 1990.
29. Weiner, S.J., Kollman, P.A., Case, D.A., Chandra Singh, U., Ghio, C, Alagona, G., Profeta, S., Jr., Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
30. Ooi, T., Oobatake, M., Nemethy, G., Scheraga, H.A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc. Natl. Acad. Sci. U.S.A. 84:3086–3090, 1987.
31. Hasel, W., Hendrickson, T.F., Still, W.C. A rapid approximation to the solvent accessible surface areas of atoms. Tetrahedron Comput. Methodol. 1:103–116, 1988.
32. Kirkpatrick, S., Gelatt, C.D., Jr., Vecchi, M.P. Optimization by simulated annealing. Science 220:671–680, 1983.
33. Bernstein, F.C., Koetzle, T.F., Williams, G.J.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. 112:535–542, 1977.
34. Suguna, K., Padlan, E.A., Smith, C.W., Carlson, W.D., Davies, D.R. Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: Implications for a mechanism of action. Proc. Natl. Acad. Sci. U.S.A. 84:7009–7013, 1987.
35. Miller, M., Schneider, J. Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H., Wlodawer, A. Structure of a complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science 246:1149–1152, 1989.
36. Tomasselli, A.G., Hui, J.O., Sawyer, T.K., Staples, D.J.,

Bannow, C., Reardon, I.M., Howe, W.J., DeCamp, D.L., Craik, C.S., Heinrikson, R.L. Specificity and inhibition of proteases from human immunodeficiency viruses 1 and 2. J. Biol. Chem. 265:14675–14683, 1990.

37. Hui, J.O., Tomasselli, A.G., Zürcher-Neely, H.A., Heinrikson, R.L. Ribonuclease A as a substrate of the protease from human immunodeficiency virus-1. J. Biol. Chem. 265:21386–21389, 1990.
38. Tomasselli, A.G., Howe, W.J., Hui, J.O., Sawyer, T.K., Reardon, I.M., DeCamp, D.L., Craik, C.S., Heinrikson, R.L. Calcium-free calmodulin is a substrate of proteases from human immunodeficiency viruses 1 and 2. Proteins: Struct. Funct. Genet. 10:1–9, 1991.
39. Shoeman, R.L., Höner, B., Stoller, T.J., Kesselmeier, C., Miedel, M.C., Traub, P., Graves, M.C. Human immunodeficiency virus type 1 protease cleaves the intermediate filament proteins vimentin, desmin, and glial fibrillary acidic protein. Proc. Nat. Acad. Sci. U.S.A. 87:6336–6340, 1990.
40. Conover, W.J. "Practical Nonparametric Statistics," 2nd ed., Ch. 4. New York: John Wiley, 1980.
41. Carlson, W., Karplus, M., Haber, E. Construction of a model for the three-dimensional structure of human renal renin. Hypertension 7:13–26, 1985.
42. Sawyer, T.K., Pals, D.T., Mao, B., Staples, D.J., deVaux, A.E., Maggiora, L.L., Affholter, J.A., Kati, W., Duchamp, D. Hester, J.B., Smith, C.W., Saneii, H.H., Kinner, J. Handschumacher, M., Carlson, W. Design, structure activity, and molecular modeling studies of potent renin inhibitory peptides having N-terminal N^{in}-For-Trp(Ftr): angiotensinogen congeners modified by P_1-P_1' Phe-Phe, Sta, LeuΨ[CH(OH)CH_2]Val or LeuΨ[CH_2NH]Val substitutions. J. Med. Chem. 31:18–30, 1988.
43. Sawyer, T.K., Mao, B. Molecular modeling logic in the design of renin-targeted cardiovascular therapeutics. Chimica*oggi* (Chem. Today) 8:53–58, 1990.
44. Heinrikson, R.L., Poorman, R.A. The biochemistry and molecular biology of recombinant human renin. In: "Hypertension: Pathophysiology, Diagnosis, and Management," Laragh, J.H., Brenner, B.M. eds. New York: Raven Press, 1990: 1179–1196.
45. Cumin, F., Le-Nguyen, D., Castro, B., Menard, J. Corvol, P. Comparative enzymatic studies of human renin acting on pure natural or synthetic substrates. Biochim. Biophys. Acta 913:10–19, 1987.
46. Poorman, R.A., Tomasselli, A.G., Heinrikson, R.L., Kezdy, F.J. A cumulative specificity model for proteases from human immunodeficiency virus types 1 and 2, inferred from statistical analysis of an extended substrate database. J. Biol. Chem., in press, 1991.