# *De Novo* Design of Enzyme Inhibitors by Monte Carlo Ligand Generation

Daniel K. Gehlhaar,[†] Karl E. Moerder,[‡] Dominic Zichi,[§] Christopher J. Sherman,[†] Richard C. Ogden,[†] and Stephan T. Freer*,[†]

*Agouron Pharmaceuticals, Inc., 3565 General Atomics Court, San Diego, California 92121, Torrey Science and Technology, Inc., 9725 Scranton Road, San Diego, California 92121, and Nexagen Inc., 2860 Wilderness Place, Boulder, Colorado 80301*

*Received April 14, 1994*[⊗]

A new computational method for the *in situ* generation of small molecules within the binding site of a protein is described. The method has been evaluated using two well-studied systems, dihydrofolate reductase and thymidylate synthase. The method has also been used to guide improvements to inhibitors of HIV-1 protease. One such improvement resulted in a compound selected for preclinical studies as an antiviral agent against AIDS.

We describe MCDNLG (Monte Carlo De Novo Ligand Generator), a computer program for the *de novo* generation of ligands within the vacant binding site of a protein. Although there has been much recent work on *in situ* ligand design by computer programs, for example, see the work described by Rotstein and Murcko[1] and the references cited therein, it has yet to be shown that these techniques can be used to design new compounds that actually enhance binding. In this paper we demonstrate that our method can.

Protein structure-based drug design utilizes the structure of a therapeutically relevant protein target to direct the search for new drug leads and optimize their design.[2,3] Notwithstanding the diversity achieved with recent combinatorial chemistry approaches to generating compound libraries[4] coupled with high-throughput screens,[5] the use of structural information to discover leads and optimize their potency is a powerful and complementary endeavor. The detailed description of the binding cavity of a protein afforded by X-ray crystallography invites the application of computational methods to generate structural proposals for new ligands. In contrast to the database screening approach,[6] these methods generate customized solutions to the problem of discovering a ligand complementary to the target protein binding site. Current *de novo* design algorithms position chemical fragments and atoms at favorable locations within a binding cavity and then extend or join these components to generate basic design ideas. The evolution of the systems and their convergence to single fragment solutions is encouraged by a variety of means. Although MCDNLG was designed independently and its development was not influenced by any other computer program for ligand design, it is of interest in that it does contain features similar to those in other *de novo* design programs that have been described recently.[7]

The primary purpose for developing MCDNLG was to speed the discovery of structurally diverse lead compounds. In our experience, and compared to the structure-based elaboration of an existing lead compound, this activity is difficult and time-consuming, even when aided by useful computer programs such as GRID.[8] We have found that design proposals are frequently biased by the human designers' desire for familiar synthesis targets. MCDNLG avoids such prejudice by using a Monte Carlo simulation to carry out *de novo* design without the use of a fragment library or other bias regarding the type of molecule desired.

The *de novo* design algorithm implemented in MCDNLG anneals a "super molecule", a nonphysical arrangement of random atoms densely packed into the target protein's binding site, into a chemically stable and physically realizable molecule that is complementarity to the site. The super molecule is comprised of atoms that can far exceed typical chemical bond valences, participating in up to 12 bonds simultaneously. Metropolis Monte Carlo[9] and simulated annealing protocols are used to evolve the mechanical system from its densely packed original state into a low-energy molecule. Mechanical energies for these nonphysical molecular states provide the chemical forces that drive the evolution.

The MCDNLG program, which is written in C, first reads a file containing the coordinates of the atoms that make up the binding site, the coordinates of the close packed hexagonal array of atoms that is placed within the binding site at the start of the simulation, and the parameters associated with the various energy terms utilized in the simulation. It next carries out the Monte Carlo simulation to evolve the ligand, and, finally, writes out the ligand coordinates in PDB format. The Monte Carlo simulation (see Figure 1) is most easily described in terms of the molecular representation of the evolving ligand, the atomic interaction energies which drive ligand design, the ligand evolution process, and the ligand design protocol.

## Molecular Representation

The fundamental units of the evolving molecular system are atoms and bonds. Each atom has an atom type associated with it, and each bond has an order, either single or double, associated with it. Each atom has a set of parameters characterizing its interactions with the other atoms in the evolving ligand and its interactions with the protein target. The individual atom types represent three distinct atomic properties: element type, electronic hybridization state, and the number of implicit hydrogen atoms. Currently, only the elements carbon, nitrogen, and oxygen are used, but in principle, the program could be expanded to include any element desired. Only $sp^2$ and $sp^3$ hybridization states are incorporated in the present version of the program. Hydrogen atoms do not exist on their own but are

[†] Agouron Pharmaceuticals, Inc.
[‡] Torrey Science and Technology.
[§] Nexagen Inc.
[⊗] Abstract published in *Advance ACS Abstracts*, November 15, 1994.

0022-2623/95/1838-0466$09.00/0 

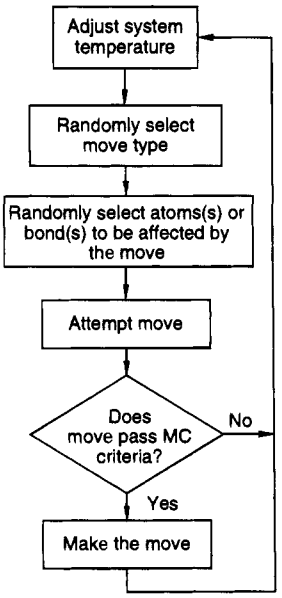


**Figure 1.** Flow diagram for the Monte Carlo simulation.

implied by the atom type. Target values for bond angles, torsional angles, and the number of valence bonds are indexed by atom type and together account for local features of the molecular geometry. During the simulation atoms are allowed to disappear and reappear, and hence an occupancy parameter is required. When the occupancy of an atom is set to "disappeared", that atom is no longer considered a part of the evolving ligand. A disappeared atom cannot be bonded to any atom. However, it remains part of the total evolving system, interacting with the protein and other disappeared atoms. Because a disappeared atom is not constrained by intramolecular potentials, it can be moved to a more geometrically favorable location in the binding site, where it may reappear to produce a ligand that better fits the binding site.

## Atomic Interaction Energies

The potential energy functions that govern the evolution and ultimately determine the structure of the evolved ligand are listed in Table 1. They fall into two distinct categories, the intramolecular potential energies among the ligand atoms and the intermolecular potential energies between the ligand and the target protein binding site. The intramolecular energy contributions primarily influence ligand geometry while intermolecular energy contributions ensure shape and chemical complementarity with the target binding site. The intramolecular interaction energies include a pairwise, radially symmetric atom–atom function based on a Morse potential. At separations greater than 6 Å, this function decays to zero. Bond angle strain energy is computed by a quadratic potential centered on the equilibrium bond angle for the vertex atom type. Torsional angle strain is computed for all combinations of three sequential bonds by a cosine function, with the number and height of the rotational barriers determined by the central bond atom types. Atomic valence strain energy is computed by a quadratic function of the deviation of the sum of the orders of its bonds from the expected number of valence bonds for each ligand atom. This term, absent from standard molecular mechanics potential energy functions, is necessary here because bond rearrangements are allowed to evolve through nonphysical valence states. At low temperatures, the valence energies ensure that chemically reasonable structures are obtained.

Intermolecular dispersion and repulsion interactions between the evolving ligand and the protein binding site are evaluated by a standard Lennard–Jones potential. To account for hydrogen bonding, MCDNLG has a quadratic potential for the donor–acceptor distance and a piecewise linear potential with minimum energy for donor–hydrogen–acceptor angles in the range of 120–180°. In addition, a desolvation penalty is given for each heteroatom in the ligand. This destabilizing energy, easily offset by hydrogen bonding to the protein, helps to ensure that heteroatoms are placed only where they interact favorably with the binding site or are necessary for the chemical stability of the ligand.

A variety of heuristic energy terms, listed in Table 2, influence the creation of rings, hydrocarbon chains, and heteroatom functionality and thereby evolve ligands with favorable molecular topology. Stabilization energies promote ring formation, local bond conjugation, and aromaticity. Destabilizing energies discourage macrocyclic ring formation, bridgehead formation, creation of fragmented ligands, and certain bonding configurations, e.g. O–O, unlikely to be found in drug molecules. These terms, also absent from traditional molecular force fields, are crucial in shaping the type of molecule generated. For example, all-carbon diamond lattices or extended planar rings are of little interest as drug candidates.

## Ligand Evolution

To evolve a ligand molecule, MCDNLG employs the standard Metropolis Monte Carlo algorithm[9] coupled with a simulated annealing protocol. As shown in Figure 1, each move is selected at random. The possible moves, and their set of allowed values, for an individual step in the Monte Carlo simulation are as follows: (1) change atom occupancy (appeared, disappeared), (2) change atom position (arbitrary small distance within cavity defined by boundary atoms), (3) change bond type (single, double), (4) change atom type (C $sp^2$, C $sp^3$, CH $sp^2$, CH $sp^3$, $CH_2$ $sp^2$, $CH_2$ $sp^3$, $CH_3$ $sp^3$, O $sp^2$, O $sp^3$, OH $sp^3$, N $sp^2$, N $sp^3$, NH $sp^2$, NH $sp^3$, $NH_2$ $sp^2$), (5) translate a fragment (arbitrary small distance within cavity defined by boundary atoms), (6) rotate a fragment about a random vector (arbitrary small angle), (7) rotate all atoms on one side of a selected bond about the bond (arbitrary small angle).

Bonds are allowed to change type, but bond existence is a function of interatomic distance only. In the simulation a bond exists between all appeared atoms that lie within 2.1 Å of each other. If, in a positional move, one atom in a bond moves more than 2.1 Å from the other, the bond is deleted. The system temperature is defined by a user-provided piecewise linear profile, an example of which is shown in Figure 2. The relative frequency among different move types is also an adjustable input parameter. Other parameters govern the relative contribution of each energy type (see Table 1). A single ligand is generated each time the simulation is run, but because the potential energy surface for the system is complex with many local minima, multiple runs that differ only by the value of the initial random

**Table 1.** Force Field Expressions with Typical Parameter Values[a]

| description | functional form | typical values |
|---|---|---|
| hydrogen-bonding protein–ligand interactions 2.3 Å < $r$ < 3.5 Å | $E_{HB} = k_{HB}(r - r^0_{HB})^2 - E^0_{HB}$ | $k_{HB} = 8.0$<br>$r^0_{HB} = 2.9$ Å<br>$E^0_{HB} = 2.5$ |
| non-bonded protein–ligand interactions | $E_{LJ} = 4k_\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$ | $k_\epsilon = 0.4$<br>$\sigma = 3.75$ Å |
| ligand intramolecular interactions | $E_{intra} = k_{intra}(1 - e^{-k_\beta(r - r^0_{intra})})^2 - k_{intra}$ | $k_{intra} = 7.5$<br>$k_\beta = 3.0$<br>$r^0_{intra} = 1.45$ Å |
| bond angle strain in degrees | $E_{angle} = k_{angle}(\theta - \theta^0_{hyb})^2 - E^0_{angle}$ | $k_{angle} = 0.025$<br>$E^0_{angle} = 10.0$ |
| deviation from standard valence | $E_{valence} = k_{valence}(\nu - \nu^0_{type})^2 - E^0_{valence}$ | $k_{valence} = 4.0$<br>$E^0_{valence} = 5.0$ |
| $sp^2$ torsional strain | $E^{sp2}_{torsion} = k^{sp2}_{torsion}[1 - \cos(2\theta)]$ | $k^{sp2}_{torsion} = 8.0$ |
| $sp^3$ torsional strain | $E^{sp3}_{torsion} = k^{sp3}_{torsion}[1 - \cos(3\phi - \pi)]$ | $k^{sp3}_{torsion} = 2.0$ |

[a] The values given are independent of atom type. The total energy is the sum of these components, with each component summed over its respective interactions, plus contributions from the heuristic energy terms given in Table 2. Ideal bond angles, $\theta^0_{hyb}$, are either 109.5° for $sp^3$ or 120° for $sp^2$ hybridization; standard valences, $\nu^0_{type}$, depend on atom type, e.g., 1 for OH or $CH_3$, 2 for O or $CH_2$, etc.; $r$ is an interatomic distance; $\theta$ is a ligand bond angle in degrees; $\nu$ is the valence of a ligand atom; and $\phi$ is a dihedral angle. All force constants, $k$, are unitless and all energy terms, $E$, are in arbitrary energy units. The ligand–ligand intramolecular energy potential is used to model both bonding and nonbonding interactions, with a bond being defined by a distance cutoff (see text); the attractive nonbonded potential encourages ligand atoms to move within bonding distance of each other.

**Table 2.** Heuristic Energy Terms Used To Encourage the Formation of Chemically Reasonable Structures[a]

| situation | type | description |
|---|---|---|
| atom type/bond type inconsistency | penalty | discourage, for example, an $sp^3$ atom with a double bond |
| two double bonds from the same atom | penalty | discourage allene formation |
| macrocycle formation | penalty | discourage formation of rings with more than seven atoms |
| bridged ring system formation | penalty | discourage overrepresentation |
| spiro-fused ring system formation | penalty | discourage overrepresentation |
| bonds between heteroatoms | penalty | discourage formation of potentially unstable groups, e.g. peroxides |
| multiple fragments | penalty | encourage output of a single, connected structure |
| heteroatom existance | penalty | desolvation penalty that discourages placement of heteroatoms in unfavorable locations |
| formation of a conjugated bond | reward | encourage conjugation, which may lead to aromaticity |
| small ring formation | reward | encourage five-, six-, and seven-membered rings |
| aromatic ring formation | reward | encourage formation of aromatic systems |

[a] Magnitudes of the atom type/bond type inconsistency and conjugated bond formation are typically 5.0, the desolvation penalty is typically 3.0, and the remainder are typically 10.0, in arbitrary energy units.

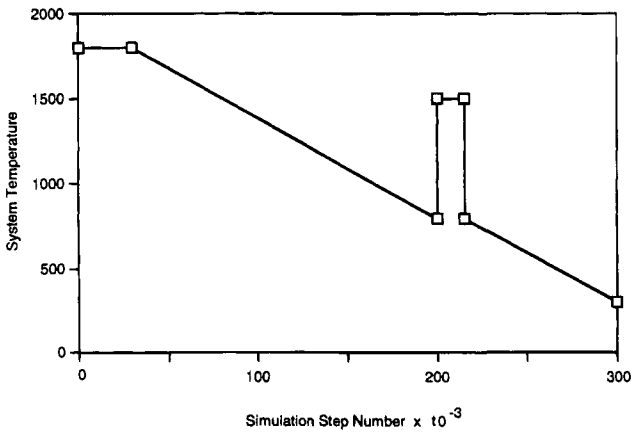


**Figure 2.** Typical temperature profile for a 300 000 step simulation. The temperature increase at approximately step 200 000 was found (generally) to produce ligands with lower energy.

number seed will produce quite different ligands and thereby offer a rich variety of molecular solutions to the same binding site topography.

## MCDNLG Design Protocol

MCDNLG starts with a nonphysical hexagonal close packed array of atoms with 1.5 Å interatomic spacing placed in a random orientation within the binding site of the target protein. The atom types are randomly selected unless the occupancies of all array atoms are set to "disappeared", in which case the default atom type, $sp^3$ OH, is used. A separate computer program with a graphical interface caps the binding site with boundary atoms, representing solvent, extracts this site from the remainder of the protein and fills it with the randomly oriented hexagonal close packed array of atoms. The only inputs required are a Protein Data Bank atomic coordinate file for the protein within whose binding site the ligands are to be generated and interactive direction from the user to locate and cap the site. Polar hydrogen atoms are frequently included with the protein in order to facilitate calculation of ligand–protein hydrogen bond geometry. A typical ligand design simulation consists of 300 000 Monte Carlo steps, with no checks for system convergence, and requires about 15 min of runtime on one processor of a Convex C240 vectorizing supercomputer. For some systems, longer runs have been shown to increase the yield of reasonable ligands.

The MCDNLG program has been used routinely for structure-based design projects in our laboratories. During its use we have investigated the program's ability to generate a variety of novel structures, its ability to build onto existing substructures, and its ability to link two fixed structural fragments together.

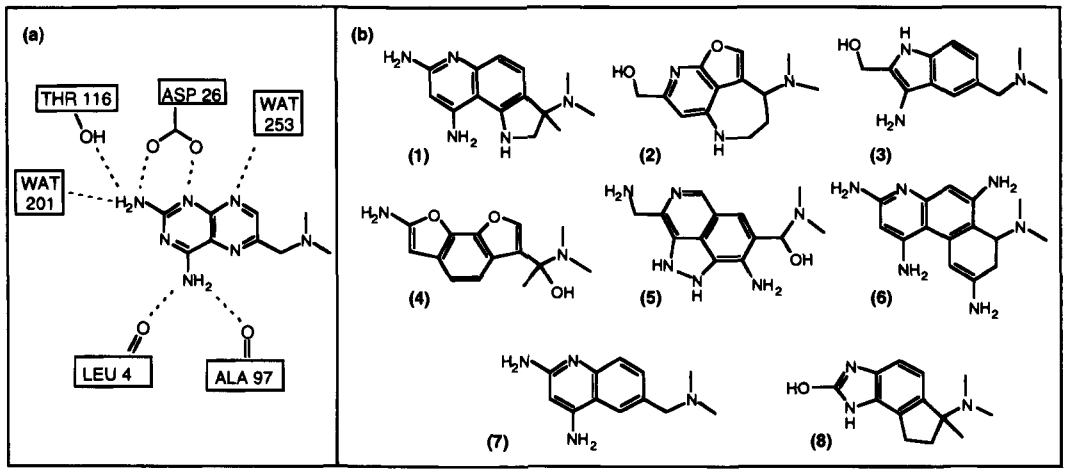


**Figure 3.** Schematic representation of design ideas produced for dihydrofolate reductase. (a) Binding mode of the pteridine ring portion of methotrexate to DHFR; dotted lines represent hydrogen bonds. (b) Representative sample of design ideas generated by MCDNLG to fill the pteridine binding pocket of DHFR. The orientation of the structures is with reference to that in part a. Note: in order to make ligand generation independent of assumed pH, MCDNLG allows the oxygen atoms of Asp and Glu side chains to be either hydrogen-bond donors or acceptors.

We describe below three applications of MCDNLG which illustrate some of these capabilities.

## Dihydrofolate Reductase

As one of the earliest paradigms of how detailed three-dimensional structural information can reveal the origins of potency and selectivity in a pharmacophore, dihydrofolate reductase (DHFR) has become a test system for many computational studies of molecular recognition.[10] We have run MCDNLG in this system to investigate how much structural diversity can be generated in a binding cavity where the vast majority of published ligands contain a 2,4-diaminopteridine or pyrimidine ring system. The intent of this experiment was to find out if MCDNLG could generate ideas for ligands that resemble the pteridine portion of known inhibitors, such as methotrexate, as well as ligands that differ but still make many of the same interactions with the protein.

Figure 3b shows a sample of the output in the same reference frame as methotrexate, which is shown schematically in Figure 3a. These structures were generated using the published structure of DHFR (Brookhaven file 3dfr). A portion of the ligand, methotrexate, containing the pteridine ring and C9 was removed, and N10 was selected as the point from which all ligands were grown. Eight representative compounds generated by MCDNLG are shown in Figure 3, exactly as they were designed by the program. Not surprisingly, some of these structures may not be chemically viable (for example, number **4**) and others may be too synthetically challenging to be useful as speculative leads. It should be stressed that the MCDNLG program is an idea generator and that knowledge of tautomer stability, $pK_a$, or even low-energy conformation is not built into it at this time. It is noteworthy that none of the structures contain the equivalent of N3 in the pteridine ring. This atom does not make any direct polar interactions with the site and hence its appearance in the output would be accompanied by an uncompensated desolvation energy penalty. The importance of N3 in methotrexate is indirect in that it alters the electronic properties of the ring and its substituents, allowing for more favorable coulombic interactions at N1. Further, several of the structures (numbers **4**–**6**) make plausible, additional interactions which are not present with methotrexate. Many of the structures in Figure 3 are unreported in the literature and therefore probably have not been evaluated as potential DHFR inhibitors.

## Thymidylate Synthase

The protein structure used for this retrospective design exercise was the ternary complex of *Escherichia coli* thymidylate synthase (TS) containing covalently bound 5-fluoro-2′-deoxyuridylate, but from which the bound antifolate, 10-propargyl-5,8-dideazafolate (CB3717), has been removed. Key water molecules known to bridge between most potent ligands and the active site residues were retained. The binding site was capped to limit ligand generation to the quinazoline region of the folate binding pocket. This structure was initially solved by Matthews and co-workers[11] and has been described previously as a suitable starting place for *de novo* design.[2]

Several groups have reported the discovery of antifolates which bind to this protein, inhibit its activity, and have the potential to be clinically useful agents.[12] Figure 4A is a schematic representation of the interactions made between the TS active site and CB3717. Figure 4B shows, on the left, four of the substructures that have been elaborated into potent enzyme inhibitors in our laboratories,[13] including two (**1a** and **4a**) which are drug leads currently under clinical investigation. Adjacent to each substructural class are examples of output from the MCDNLG program which are most suggestive of each class. Again, no filtering of the output to take account of ease of synthesis, stability, or tautomerism has been performed. The generation of the set of 100 structures from which these examples were chosen took approximately 20 h of CPU time on single processor of a Convex C240 computer. Although no exact replicas of the known structures were found in this run, it is gratifying that linear bicyclic (**1b**) and tricyclic (**2c**) as well as fused tricyclic ring systems (**4b**) were found.

## HIV-1 Protease

The above examples represent applications of the program to designing novel substructures in well-

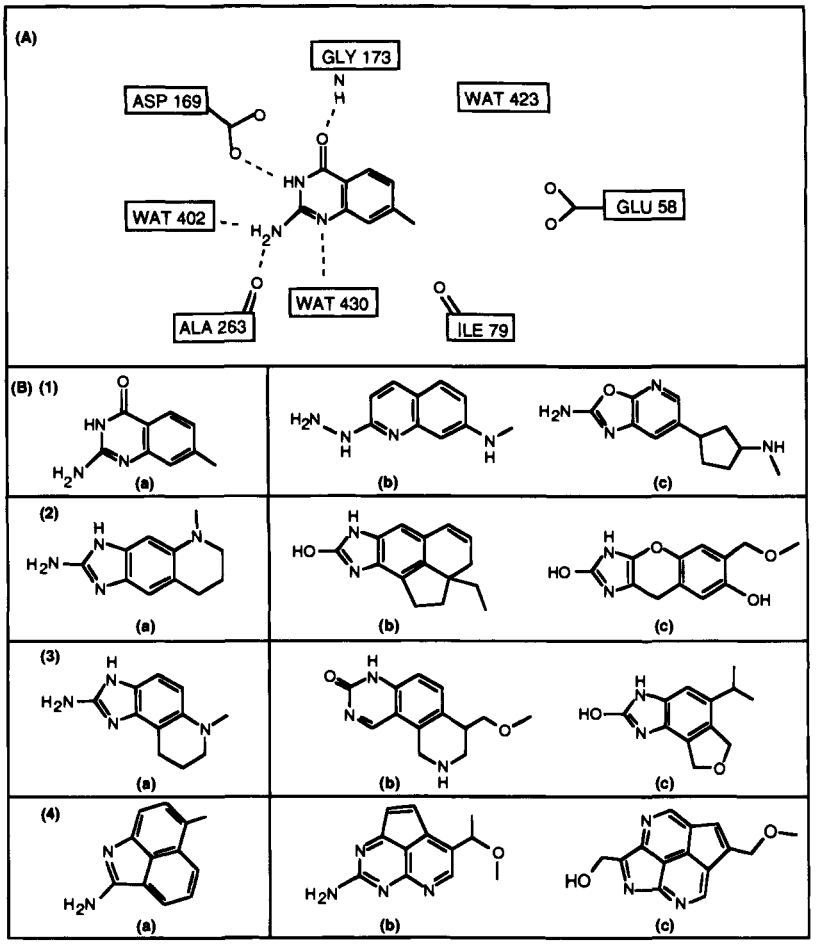


**Figure 4.** Schematic representation of design ideas produced for thymidylate synthase: (A) binding mode of the quinazoline portion of CB3717, (B) representative sample of design ideas generated by MCDNLG to fill the quinazoline binding pocket of thymidylate synthase. On the left (**1a**–**4a**) are structurally diverse portions of inhibitors synthesized in our laboratories. On the right, for each class of inhibitor, are closely related design ideas generated by MCDNLG to fill the corresponding portion of the binding site.

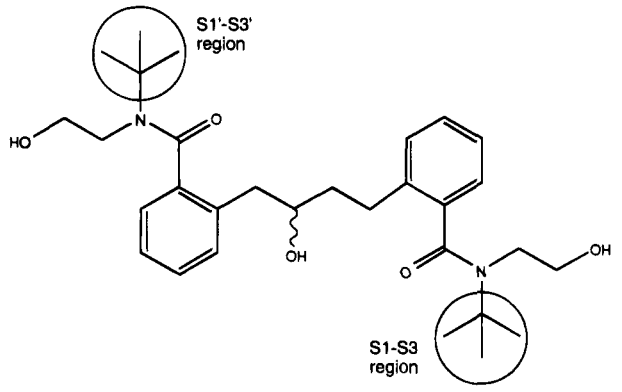


**Figure 5.** HIV-1 Protease inhibitor used as a starting point for MCDNLG design. *tert*-Butyl groups that bind in the S1'-S3' and S1-S3 pockets are circled. The crystal structure of the complex used for design contained the *R* enantiomer. Inhibition constants for both the starting compound and the compound suggested by MCDNLG were measured on racemic mixtures.

defined subsites of an enzyme active site. In certain enzymes, of which the aspartyl protease from HIV-1 is a good example, the active sites lack such obvious large subsites but rather, as a consequence of the extended conformation of their polypeptide substrates, consist of elongated binding sites with one or more specificity pockets that accommodate certain substrate side chains.

In the course of a research project to discover nonpeptidic inhibitors of HIV-1 protease for potential use as orally active antiviral drugs, our colleagues synthesized one series of molecules containing two ortho-substituted *N*-*tert*-butyl-*N*-(2-hydroxyethyl)benzamide moieties, one of which is shown in Figure 5. The *tert*-butyl groups were known from X-ray crystallographic studies to bind in the S1-S3, and S1'-S3' pockets.[14] As one potential way to increase the potency of the lead compound in this series, MCDNLG was run in these subsites of the enzyme active site. Prior to the run, the sites were capped by about 20 boundary atoms placed in the vicinity of residues Phe 53, Arg 8, and Pro 81 in both subunits of the enzyme, and the methyl groups forming the *tert*-butyl substituent were removed from the bound ligand structure, shown in Figure 5. The remaining carbon atom was defined as the point of attachment for the replacement fragment. This atom was not allowed to disappear, but otherwise acted as any other ligand atom, i.e., new bonds could be formed to it, and it was allowed to move and change atom type. Two independent series of runs were carried out, into the capped S1'-S3' region and into the symmetry related S1-S3 region. We had observed in our crystallographic studies that ligand binding could induce different conformations of the residues that make up these subsites and therefore decided to treat them independently.

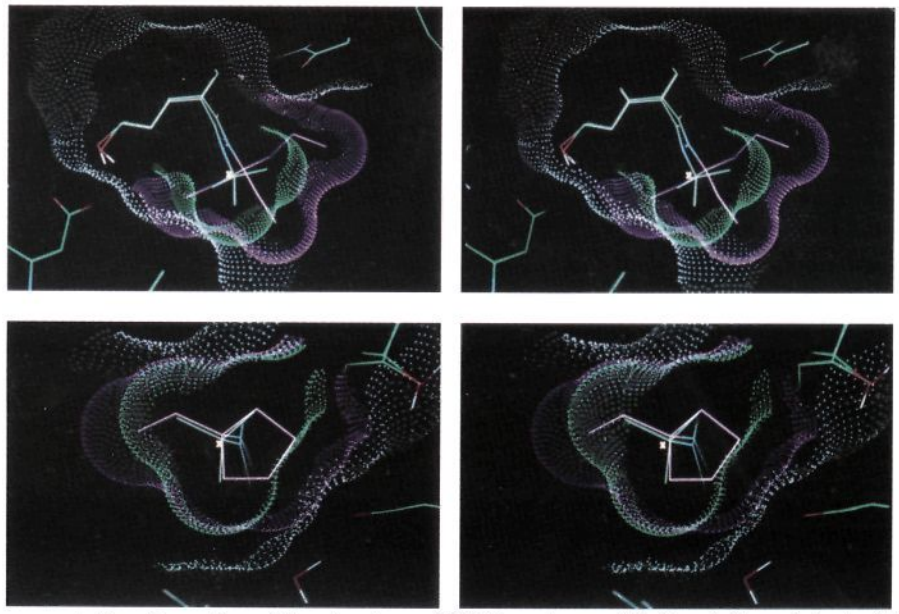

**Figure 6.** Stereo representation of a portion of the structure of HIV-1 protease with ligands bound. (a, top) the S1′-S3′ region of the active site. Crystallographically determined positions of the parent compound with the *tert*-butyl group in green and the derivative with the ethyl and *n*-propyl groups in purple are shown together with the respective Connolly surfaces for those groups and the surface for the protein in white. The origin atom for the MCDNLG run is marked with an X. (b, bottom) the S1-S3 region of the active site. Crystallographically determined positions of the parent compound with the *tert*-butyl group in green and the derivative with the cyclopentyl and ethyl groups in purple are shown with the respective Connolly surfaces for those groups and the surface for the protein in white.

The first run, in the S1′-S3′ region, produced several interesting functional groups that appeared to make superior contact with the hydrophobic residues forming the subsite as compared to the original *tert*-butyl group. Following a routine gas phase minimization of the most reasonable generated structures,[15] they were re-evaluated both conformationally and from the point of view of synthetic accessibility. The structure containing the diethyl, *n*-propyl substitution pattern off the origin atom was selected for synthesis. Figure 6a shows the crystallographically determined structure of the new bound ligand in the region of the new substituent, compared to the structure of the parent ligand. The point of origin for the MCDNLG runs is indicated (X) and the solvent accessible surfaces[16] shown are for the parent ligand (green), the new ligand (purple), and the protein (white). The additional contacts made by the four extra carbon atoms are clearly apparent in Figure 6a and probably account for the increased binding observed for the new ligand, which is 9-fold greater than parent ligand in inhibiting the enzyme.

The same type of MCDNLG run was performed in the S1-S3 pocket, capped in a similar manner and initiated from the equivalent carbon atom. Again, several structural leads worth pursuing further were obtained and, in part because of the small site differences and in part because of the nature of the Monte Carlo simulation, there was no overlap of the output with the previous experiment. Figure 6b shows the structure of an inhibitor containing cyclopentyl and ethyl substituents off the origin carbon, again superimposed on the structure of the parent *tert*-butyl amide. This compound, directly generated by the program, was synthesized and shown to bind in the protein with the cyclopentyl and ethyl groups in two possible rotational orientations, one of which was identical to that generated by MCDNLG. The binding of this ligand was improved 10-fold over the parent molecule against the isolated enzyme.[14] As before, additional protein−ligand contact made by the additional ligand atoms may account for the increased binding.

## Discussion

The results of the MCDNLG applications described above demonstrate that the program is a useful addition to small molecule design protocols. When used as an idea generator, in conjunction with input from modelers, it can rapidly accelerate design of new ligands and suggest improvements to existing ligands. The majority of MCDNLG output structures require some modification before they become synthetically viable targets, but on occasion, as with some of the HIV-1 protease runs, the program generates readily synthesizable compounds.

We are working on ways to enhance the performance of MCDNLG and to extend its application. Part of our current efforts is directed toward optimizing the user-adjustable parameters, which have a varying degree of influence on the output ligands. Since genetic algorithms are ideally suited to finding optimal combinations of linked and unlinked parameters, we are investigating their application to this problem. In its present form, MCDNLG does not distinguish whether a newly designed ligand is truly novel, and we are developing techniques for automatic postrun filtering of generated structures to eliminate ligands overly similar to ones previously generated. We are also trying to develop methods for reliably ranking the remaining compounds to select those most promising. A fundamental change in the design protocol would be to employ fragment positioning, such as that found in the LUDI program,[17] to extend the fundamental unit of manipulation in MCDNLG to include a selected set of chemical fragments. This approach may have the effect of allowing

known pharmacophores or known conformational preferences of functional groups to appear in the eventual output. A weakness in the current version of MCDNLG is the assumption that the binding site is rigid; i.e., there is no way to account for motion of protein backbone and side chain atoms; we hope to develop methods to account for flexibility of the binding cavity in a subsequent version of the program. We are investigating application of MCDNLG to computational database searching methods for optimization of hits found by them and to suggest appropriate database queries to direct the search itself. In addition, we are developing portions of the MCDNLG code to use for ligand docking and for ranking design ideas.

**Acknowledgment.** We appreciate the help provided by our Agouron colleagues who reviewed the manuscript.

## References

(1) Rotstein, S.; Murcko, M. GroupBuild: A Fragment-Based Method for *De Novo* Drug Design. *J. Med. Chem.* **1993**, *36*, 1700–1710.
(2) 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.; 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.; and White, J. Design of Enzyme Inhibitors Using Iterative Protein Crystallographic Analysis. *J. Med. Chem.* **1991**, *34*, 1925–1934.
(3) Erickson, J. W.; Fesik, S. W. Macromolecular X-Ray Crystallography and NMR as Tools for Structure-based Drug Design. *Annu. Rep. Med. Chem.* **1992**, *27*, 271–289.
(4) Jung, G.; Beck-Sickinger, A. G. Multiple Peptide Synthesis Methods and Their Applications. *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 367–383. Moos, W. H.; Green, G. D.; Pavia, M. R. Recent Advances in the Generation of Molecular Diversity. *Annu. Rep. Med. Chem.* **1993**, *28*, 315–324.
(5) Burch, R. M.; Kyle, D. J. Mass Receptor Screening for New Drugs. *Pharm. Res.* **1991**, *8*, 141–147.
(6) 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. 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**, *78*, 182–196. Humblet, C.; Dunbar, J. B., Jr. Database Searching and Docking Strategies. *Annu. Rep. Med. Chem.* **1993**, *28*, 275–284.
(7) Pearlman, D. A.; Murcko, M. A. CONCEPTS: New Dynamic Algorithm for *De Novo* Drug Suggestion. *J. Comput. Chem.* **1993**, *14*, 1184–1193. Nishibata, Y.; Itai, I. Confirmation of Usefulness of a Structure Construction Program Based on Three-Dimensional Receptor Structure for Rational Lead Generation. *J. Med. Chem.* **1993**, *36*, 2921–2928.
(8) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. *J. Med. Chem.* **1993**, *36*, 140–147. Goodford, P. J. A Computational Procedure for Determining Enegetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
(9) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1093.
(10) Tschinke, V.; Cohen, N. C. The NEWLEAD Program: A New Method for the Design of Candidate Structures from Pharmacophoric Hypotheses. *J. Med. Chem.* **1993**, *36*, 3863–3870. Nishibata, Y.; Itai, A. Automatic Creation of Drug Candidate Structures Based on Receptor Structure. Starting Point for Artificial Lead Generation. *Tetrahedron.* **1991**, *47*, 8985–8990.
(11) Matthews, D. A.; Appelt, K.; Oatley, S. J.; Xuong, N. H. Crystal Structure of *Escherichia coli* Thymidylate Synthase Containing Bound 5-Fluoro-2′-deoxyuridylate and 10-Propargyl-5,8-dideazafolate. *J. Mol. Biol.* **1990**, *214*, 923–936.
(12) Jones, T. R.; Calvert, A. H.; Jackman, A. L.; Brown, S. J.; Jones, M.; Harrap, K. R. A Potent Antitumour Quinazoline Inhibitor of Thymidylate Synthetase: Synthesis, Biological Properties and Therapeutic Results in Mice. *Eur. J. Cancer* **1981**, *17*, 11–17. Marsham, P. R.; Hughes, L. R.; Jackman, A. L.; Hayter, A. J.; Oldfield, J.; Wardelworth, J. M.; Bishop, J. A. M.; O'Connor, B. M.; Calvert, A. H. Quinazoline Antifolate Thymidylate Synthase Inhibitors: Heterocyclic Benzoyl Ring Modifications. *J. Med. Chem.* **1991**, *34*, 1594–1605. McNamara, D. J.; Berman, E. M.; Fry, D. W.; Werbel, L. M. Potent Inhibition of Thymidylate Synthase by Two Series of Nonclassical Quinazolines. *J. Med. Chem.* **1990**, *33*, 2045–2051.
(13) Varney, M. D.; Marzoni, G., P.; Palmer, C. L.; Deal, J. G.; Webber, S.; Welsh, K. M.; Bacquet, R. J.; Bartlett, C. A.; Morse, C. A.; Booth, C. L. J.; Herrmann, S. M.; Howland, E. F.; Ward, R. W.; and White, J. Crystal-Structure-Based Design and Synthesis of Benz[*cd*]indole-Containing Inhibitors of Thymidylate Synthase. *J. Med. Chem.* **1992**, *35*, 663–676. Reich, S. H.; Fuhry, M. A.; Nguyen, D.; Pino, M. J.; Welsh, K. M.; Webber, S.; Janson, C. A.; Jordan, S. R.; Matthews, D. A.; Smith, W. W.; Bartlett, C. A.; Booth, C. L. J.; Herrmann, S. M.; Howland, E. F.; Morse, C. A.; Ward, R. W.; White, J. Design and Synthesis of Novel 6,7-Imidazotetrahydroquinoline Inhibitors of Thymidylate Synthase Using Iterative Protein Crystal Structure Analysis. *J. Med. Chem.* **1992**, *35*, 847–858. Webber, S. E.; Bleckman, T. M.; Attard, J.; Deal, J. G.; Kathardekar, V.; Welsh, K. M.; Webber, S.; Janson, C. A.; Matthews, D. A.; Smith, W. W.; Freer, S. T.; Jordan, S. R.; Bacquet, R. J.; Howland, E. F.; Booth, C. L. J.; Ward, R. W.; Herrmann, S. M.; White, J.; Morse, C. A.; Hilliard, J. A.; and Bartlett, C. A. Design of Thymidylate Synthase Inhibitors Using Protein Crystal Structures: The Synthesis and Biological Evaluation of a Novel Class of 5-Substituted Quinazolinones. *Med. Chem.* **1993**, *36*, 733–746.
(14) Reich, S. H.; Melnick, M.; Davies, J. F., II; Appelt, K.; Lewis, K. K.; Fuhry, M. A.; Pino, M.; Trippe, A. J.; Nguyen, D.; Dawson, H.; Wu, B.; Musick, L.; Shetty, B.; Kosa, M.; Khalil, D.; Webber, S.; Gehlhaar, D. K; and Andrada, D. Protein Structure-Based Design of Potent, Orally Bioavailable Nonpeptidic HIV Protease Inhibitors. In press.
(15) 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.
(16) Connolly, M. L. Analytical Molecular Surface Calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558. Connolly, M. L. Computation of Molecular Volume. *J. Am. Chem. Soc.* **1985**, *107*, 1118–1124.
(17) Böhm, H.-J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61–78.

JM940228V