CONCERTS: Dynamic Connection of Fragments as an Approach to *de Novo* Ligand Design

David A. Pearlman* and Mark A. Murcko

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

Received October 26, 1995[⊗]

We have implemented and tested a new approach to *de novo* ligand design, CONCERTS (creation of novel compounds by evaluation of residues at target sites). In this method, each member of a user-defined set of fragments is allowed to move independently about a target active site during a molecular dynamics simulation. This allows the fragments to sample various low-energy orientations. When the geometry between proximal fragments is appropriate, bonds can be formed between the fragments. In this fashion, larger molecules can be built. The bonding arrangement can subsequently be changed—breaking bonds between chosen fragment pairs and forming them between other pairs—if the overall process creates lower energy molecules. We have tested this method with various mixes of fragments against the active sites of the FK506 binding protein (FKBP-12) and HIV-1 aspartyl protease. In several cases, CONCERTS suggests ligands which are in surprisingly good agreement with known inhibitors of these proteins.

Introduction

Although recent developments in the area of combinatorial chemistry hold promise for accelerating the lead identification and optimization cycle of drug design,[1] at present this process remains very time-consuming. And even if the promise of such methods is borne out, it is expected that they will be expensive to implement and that there will be systems of interest for which they will not be amenable. In addition, it is certain that with such methods we will still be limited to synthesizing only a small fraction of the molecules that could be reasonably tested against a particular target. For these reasons, there is tremendous interest in devising theoretical methods that can help suggest, *de novo* and relatively quickly, compounds which bind to macromolecular targets of interest. These predictions can be used directly as ligand suggestions, as templates for directed database searches for potential inhibitors, or to complement combinatorial approaches by helping to narrow the range of molecules which are synthesized.

In the past few years, a substantial number of papers have appeared which describe techniques that attempt to address this interest.[2] Many of the methods appear promising, though all suffer various drawbacks, particularly in terms of how they deal with the "combinatorial explosion" in the number of choices that must be made in designing a ligand. Most of these methods rely on chained or tree search techniques, which fail when inappropriate choices are made early in the chain. Unfortunately, what appears to be the optimal choice at any link in the chain is frequently not the best choice when information further down the chain is considered.

In an attempt to circumvent this problem, we recently developed and described a new method, CONCEPTS (creation of novel compounds by evaluation of particles at target sites).[3] In the CONCEPTS approach, an active site is filled with atoms, which adopt low-energy configurations during a molecular dynamics (MD) simulation.[4] Bonds between atoms can be formed and subsequently broken based on geometric and energetic considerations. This ability to both form and break bonds, and the fact that the system is constantly evolving under molecular dynamics while these changes to the topology are occurring, help differentiate the CONCEPTS technique from other *de novo* approaches. Application of CONCEPTS to the FK506 binding protein FKBP-12 and to HIV-1 aspartyl protease demonstrated that the method is capable of generating useful *de novo* leads suggestive of known inhibitors of these proteins.[3]

While the results obtained using CONCEPTS are quite encouraging, and while this method offers, at least in principle, a potential improvement on other *de novo* approaches, CONCEPTS also suffers from several drawbacks. In particular: (1) It is difficult to incorporate a charge model into such an atom-based approach which is both reasonable and sufficiently fast—thereby limiting the applicability of the program to the design of ligands that bind mostly through hydrophobic interactions; (2) The single-atom buildup approach can result in slow convergence, especially for "spacer" regions that connect tight-binding but spatially separated moieties; (3) One obtains only a single suggestion per cpu-intensive run, and one must often average several runs to obtain a good inhibitor suggestion. In a more global sense, (2) and (3) are both related to inefficiencies of the method, while (1) is a weakness in the model used.

It was our desire to include the benefits of CONCEPTS while avoiding some of the drawbacks that lead us to the method described in this paper: CONCERTS (creation of novel compounds by evaluation of residues at target sites). In essence, we hve retained the basic idea of CONCEPTS: allow the building blocks to continually reorient in the active site using molecular dynamics, and connect them based on geometrical and energetic criteria. But we have changed the elementary building block from an atom to a molecular fragment and have made other changes as well. As long as we utilize only fragments where the change in the charge distribution upon bonding is reasonably small and localized, it is possible to incorporate charges into these

* Corresponding author.
⊗ Abstract published in *Advance ACS Abstracts,* March 15, 1996.

0022-2623/96/1839-1651$12.00/0

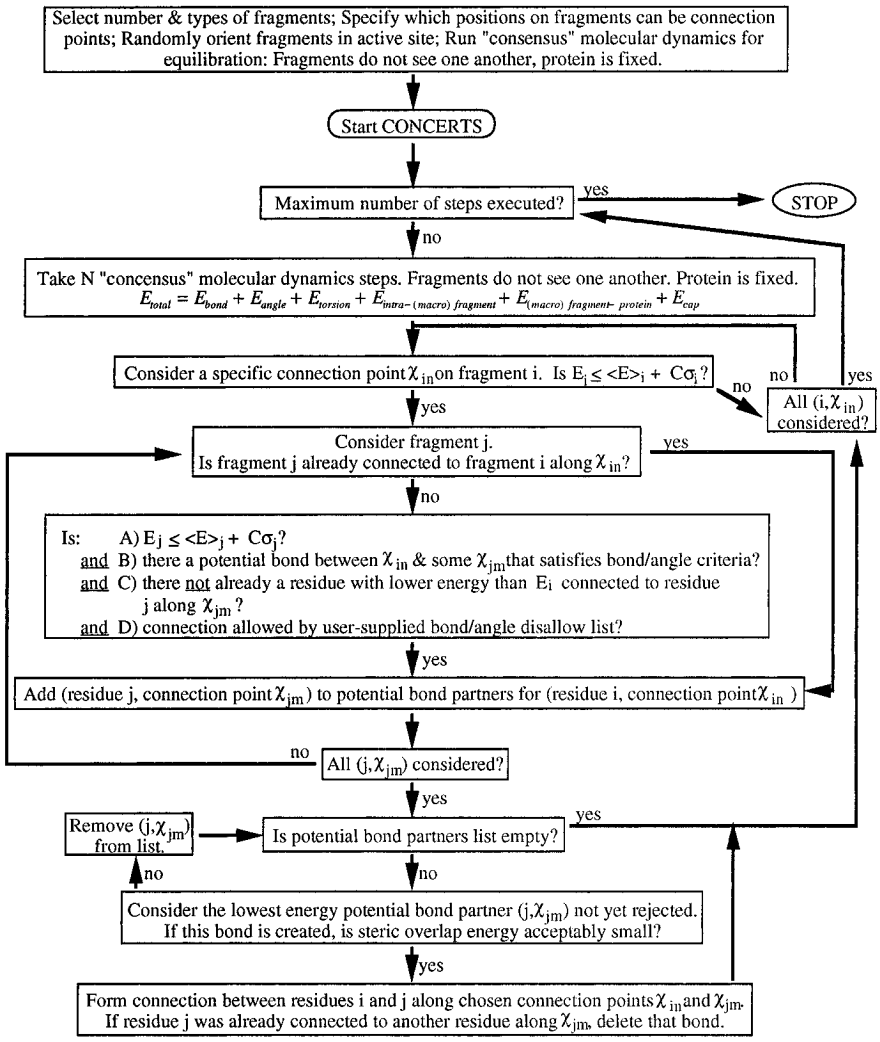

Figure 1. Schematic flowchart illustrating the CONCERTS method. For more details, see the text.

calculations, avoiding CONCEPTS problem (1). While still offering tremendous conformational and topological possibilities, larger fragments can much more quickly span large regions of space, thereby attenuating CONCEPTS problem (2). And in CONCERTS, fragments that do not belong to the same molecule do not see each other. This means that many molecules ("suggestions") can be built in the same binding region during a single run of CONCERTS, avoiding drawback (3). In addition to these benefits, CONCERTS also allows greater control over the types of molecules generated. This should result in a greater percentage of suggestions which are synthetically reasonable.

To evaluate the new method, CONCERTS has been applied to the same test systems as were examined with CONCEPTS: the binding sites of both the 107-residue FK506 binding protein, FKBP-12,[5,6] which is important in a pathway for immunosuppression, and HIV-1 aspartyl protease,[7−9] a 198-residue protein which is necessary for the replication of the AIDS virus. These are good test systems because their structures are well-characterized and because potent inhibitors of both are known.

Methods

CONCERTS has been implemented in a very extensively modified version of the AMBER/SANDER 4.0 minimization/molecular dynamics program.[10,11] An overview of the conceptual flow of the program is presented in Figure 1. A detailed description of the implementation is presented in the Appendix.

In broad terms, the method is as follows. A user-specified set of molecular fragments (small molecules) move about a fixed active site during molecular dynamics. These fragments do not "see" (interact with) each other, so they will tend to arrange themselves in a distribution that reflects the energetics within the active site. At specified intervals, an attempt is made to connect fragments which are favorably oriented relative to one another. Connections can be made between fragments along user-specified bonds to hydrogens. When such a bond is made, the two hydrogens are eliminated. On subsequent steps, the connection between one pair of fragments can be broken in order to form energetically more favorable connections to different fragments. Thus, it is not necessary that the original set of connection choices is ideal; the connectivity can iteratively improve over the course of the simulation. This is one of the major advantages of CONCERTS over other ligand design methods. Many fragments can be joined together in the same "macrofragment" (molecule consisting of several connected fragments). Fragments belonging to the same macrofragment see each other, but do not see any unbound fragments or fragments belonging to other macrofragments. This allows the generation of a large number of ligand suggestions with a single run of CONCERTS.

Simulation and Parameter Choices. Three different sets of fragments were used to generate most of the results we describe here: (A) 1000 copies of the peptide fragment $HCONHCH_3$ (Figure 2); (B) 700 copies of the fragment C_6H_6 (benzene) plus 1000 copies each of the fragments CH_4 (methane), NH_3 (ammonia), HCOH (formaldehyde), and H_2O (water);

(C) 300 copies each of the fragments, NH_3 (ammonia), C_6H_6 (benzene), C_6H_{12} (cyclohexane), HCO_2H (formic acid), C_2H_6 (ethane), C_2H_4 (ethylene), HCOH (formaldehyde), $HCONH_2$ (formamide), CH_4 (methane), CH_3OH (methanol), HSO_2H (sulfinic acid), C_4H_4S (thiophene), H_2O (water). In addition, in one set of simulations, we started from an oligopeptide backbone suggested from a peptide fragment run and attempted to add standard amino acid side chains. In these simulations, 150 copies each of 19 standard amino acid side chains (excluding proline and glycine, and including both δ- and ϵ-protonated histidine side chains as separate fragments) were used. Fragments are not rigid and can undergo conformational fluctuations during MD.

Parameters for all fragments are from the standard Weiner *et al*. force field.[12] Charges for the peptide and amino acid fragments are also from Weiner *et al*. Charges for other fragments were derived from electrostatic potential fitting to the HF/6-31G* wave function using the MK option in Gaussian-92.[13] All fragments are net neutral, except for the amino acid side chains Asp and Glu, which have a net charge of −1, and Lys and Arg, which have a net charge of +1. For all but the peptide and amino acid side chain fragments, all bonds to hydrogens were designated as "connection vectors". For the peptides, all bonds to hydrogens *except* the bond to the hydrogen on the nitrogen were designated "connection vectors". For the amino acid side chain fragments, only bonds to hydrogens connected to C_β were designated "connection vectors". All-atom models were used for all fragments.

CONCERTS simulations were run in the active sites of two proteins: FKBP-12 and HIV-1 aspartyl protease. Crystal structures of both proteins are known.[14,15] Hydrogens were added to the crystal coordinates, and the proteins were minimized *in vacuo*. As this minimization was simply to regularize and remove any bad steric contacts in the crystal structure, an explicit solvent model was not necessary. The resulting coordinates were those used in all subsequent work. In the "FLOAT" phase of initial random fragment positioning, a box with sides of 14−20 Å was used to fully enclose the appropriate binding site. A total of 100 steps of steepest descents + 150 steps of conjugate gradients consensus minimization was carried out on the resulting positions to alleviate any steric overlaps and to prepare the system for subsequent MD. During this minimization phase, the protein was held fixed and the force field given in eq 1 (Appendix) was used (i.e., there were no nonbonded interactions between fragments).

For both systems, consensus MD equilibration was carried out at 500 K. An elevated temperature was used to speed the redistribution of fragments into favorable positions. A total of 10 000 steps of MD with a time step of 1 fs were performed. The protein was held fixed, and the same energy function (eq 1) was used as for minimization. Fragments were restrained to a spherical region of radius 10 Å (FKBP-12) or 11 Å (HIV-1) centered on the active site by a "cap" restraining potential with force constant 5.0 kcal/mol (eq 5, Appendix). All simulations were carried out *in vacuo*, with a dielectric constant of $\epsilon = r$. In the case of the aspartyl protease protein, the "flap water"—which is believe to be crucial for inhibitor binding[15]—was included and fixed throughout the simulation. A crude temperature-scaling method was employed, wherein the velocities of the system were scaled to the target temperature any time the actual temperature of the system was more than 100 K away from the target temperature.

A nonbonded cutoff of 7 Å was employed. This nonbonded cutoff is somewhat smaller than the standard cutoff of 8 Å used in much protein modeling and shorter still than some of the extended cutoffs which have been recommended for precise modeling.[16] We chose to use the shorter cutoff because it results in a large cpu savings for a cpu-intensive procedure. Given other approximations in this method, we do not expect the effects of using this cutoff to be significant, and in fact a trial simulation with a longer (8 Å) cutoff did not yield appreciably different predictions (unpublished results).

The actual CONCERTS simulations were run for 6000 steps (MAXSTEP = 6000) using a timestep of 2 fs. The energy function given in eq 7 (Appendix) was used, with a nonbonded cutoff of 7.0 Å and a dielectric constant of $\epsilon = r$. Fragments were restrained to a spherical region centered on the active site, as in equilibration. The protein (and associated flap water for aspartyl protease) was held fixed. To ensure that groups that should be planar did not exceed planarity by too much during high-temperature simulations, the "improper" torsion force constants[12] were scaled up by a factor of 10. The temperature for these runs was changed according to the following schedule:

steps	temperature
0 → 1500	T_{max} → T_{max}
1501 → 3500	T_{max} → 300 K
3501 → 5000	300 K → 300 K
5001 → 6000	300 K → 50 K

This annealing schecule allows the system to rapidly sample conformational space in the early stages and to relax into a relatively low-energy configuration later on. As in equilibration, crude temperature scaling was employed, where velocities were scaled back to the target temperature any time the actual temperature of the system deviated by more than 100 K from the target temperature. For the simulations described herein, a T_{max} of 600 K was used. This peak temperature appeared, qualitatively, to yield more promising predictions than simulations at either 1200 or 300 K. Compared to the runs at 600 K, the 1200 K runs tended to generate many more "odd", higher energy macrofragments, while the 300 K runs tended to yield less diverse and interesting ensembles of structures.

During the CONCERTS simulations, a search for new connections was performed every 5 dynamics steps ($N = 5$; step (8) above). Connections were only attempted to fragments for which $E_i \leq \langle E_i \rangle$, where $\langle E_i \rangle$ is the average energy of all fragments of the same type as i. This corresponds to $C = 0$ in eq 8, Appendix. We found any C roughly in the range $-0.5 \rightarrow 1.0$ seemed to work reasonably well. When determining potential bond partners, bond and angle tolerances of $\delta = 0.3$ Å and $\delta_\theta = 30°$ (eqs 9 and 10, Appendix) were used. $E_{overlap}$, which is the maximum nonbonded interaction energy between the two fragments/macrofragments which will be joined if a bond is accepted, was set to 5000 kcal/mol (disallowing only severe overlaps).

For simulations using the non-peptide fragments, the following atom sequences were disallowed for fragment−fragment bonds, in order to reduce the chances of building chemically unstable or unreasonable compounds: O−O, N−N, N−O, S−O, O−C−O, O−N−O, N−C−N. For the peptide runs, the atom sequences C_α−C_α and C−C_α−C were disallowed, which ensures that peptides are only joined in the appropriate direction and that there is no peptide "branching" at the C_α position. The peptide fragment is defined so that the amide bond does not depend on the relative orientations of two peptides when they are connected (Figure 2). Thus, by starting with peptide fragments in the trans conformation about the amide bond, we effectively ensure that all peptide residues in the final structures will be trans. (In principle, a trans-to-cis isomerization could occur during MD, but the barrier to rotation about this partial double bond is large enough that this was not observed in our calculations.)

When a fragment−fragment bond is created, harmonic bond and angle terms and Fourier torsional terms, corresponding to the bond, angles, and torsions created with the fragment−

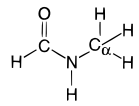

Figure 2. The peptide monomeric fragment used in the CONCERTS simulations described herein. Note that this fragment differs from the typical definition of a peptide fragment, which begins with the amide nitrogen and ends wtih the carbonyl group. Our fragment is defined so that conformation about the amide bond will not be a function of how adjacent peptide fragments are joined together. This avoids unwanted cis−trans racemization.

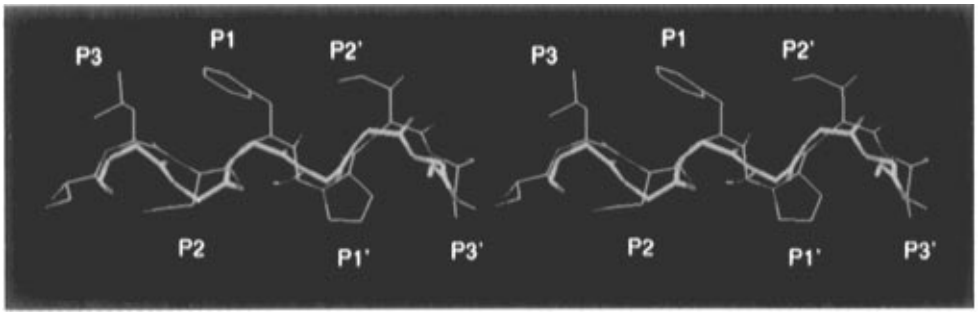

Figure 3. The lowest energy oligopeptide generated by CONCERTS against HIV-1 aspartyl protease (thick lines). This oligopeptide is compared to JG-365, a subnanomolar peptide inhibitor of the protein, which is shown in thin lines. Also shown is the "flap water", which was held fixed during the simulation (see text). The molecules are in the same protein reference frame. Carbon atoms are in green, oxygen atoms are in red, and nitrogen atoms are in blue. Hydrogen atoms are not shown. Subsites in the inhibitor are labeled with their standard P1−P3 and P1′−P3′ designators. The figure is presented as a stereoview. The predicted ligand does a good job of mimicking the backbone conformation of JG-365.

fragment bond, are added to the force field. Standard single bond distance and angle target values are assigned, based on the types of the atoms making the bond and angles (C−C = 1.52 Å; C−N = 1.46 Å; C−O = 1.42 Å; C−S = 1.80 Å; sp^3 angles = 112°; sp^2 angles = 120°). The force constant for fragment−fragment bonds is 300 kcal/mol·Å^2 and the force constant for angles created by a fragment−fragment bond is 30 kcal/mol·rad^2. These values were chosen to be representative of the force constants used in the Weiner *et al.* force fields for bonds and angles.[12] Torsional parameters are assigned using the torsion parameters list from the standard Weiner *et al.* force field.[12] Nonbonded interactions are calculated between all fragments belonging to the same macrofragment (collection of joined fragments), subject to exclusions for 1−2 and 1−3 interactions. The 1−4 interactions both within fragments and between fragments joined by a fragment−fragment bond are reduced by a factor of $^1/_2$.[12]

Once the CONCERTS simulation is complete, the system is consensus minimized for approximately 200 steps, using the same force field as for the MD portion. Again, the protein is held fixed, and the macrofragments do not see one another.

Each simulation results in many macrofragment suggestions comprised of various numbers of fragments. These can be ordered in terms of properties such as potential energy or number of residue and can be examined using a standard visualization program, such as Insight[17] or Quanta.[18] Energies used for comparison are those due to protein−macrofragment nonbonded interactions and can be either total energy, the energy per residue, or $\sum E_{nb(relative)}$, where $E_{nb(relative)}$ is, for each fragment, the nonbonded interaction energy of that fragment minus the average nonbonded interaction energy for all fragments of the same type.

The calculations are relatively compute intensive. Run on a single processor of a SGI R8000 Power Challenge computer, the 6000-step CONCERTS simulation took between $^1/_2$ and $^2/_3$ day for the peptide fragment runs, and between 4 and 7 days for the organic fragment simulations. The CONCERTS code has not been entirely optimized for best cpu performance, and we expect that these run times could be reduced.

Results

HIV-1 Protease. To test the CONCERTS method, we first applied the approach to HIV-1 aspartyl protease, using the set of peptide fragments (A) as our "basis set". This is perhaps the best basic test of the method, since (1) oligopeptide derivatives are known substrates of HIV-1 aspartyl protease[8] and (2) limiting the search to a single fragment type means we can include more copies of that fragment type than in a mixture (given finite computer resources), and that there are considerably fewer fragment configurations to consider.

The peptide simulation was performed as described under Methods. After the 6000-step CONCERTS simulation and minimization, the resulting macrofragments were analyzed. A total of 82 macrofragments consisting of four or more peptides were found, of which 35 were tetrapeptides, 27 were pentapeptides, 17 were of hexapeptides, and 3 were heptapeptides. The lowest energy macrofragment generated is a seven-residue oligopeptide, shown in Figure 3. This macrofragment also corresponds to the fifth lowest value of $\sum E_{nb(relative)}$/residue of any oligopeptide of hexapeptide or longer length. As can be seen, the polyglycine peptide does a very good job reproducing the backbone of JG-365,[19] a subnanomolar peptidic inhibitor of HIV-1 protease. (A variety of subnanomolar HIV-1 protease inhibitors have been reported;[9] JG-365 is representative of many of these.) In particular, the peptide mimics nearly all the carbonyl positiions in the known inhibitor. While the five matching carbonyl positions in the peptide macrofragment are offset by 0.9−2.0 Å, the *de novo* lead is clearly suggestive of the experimental binding conformation. It is interesting to note that one carbonyl in the middle of the oligopeptide points in the opposite direction of its inhibitor counterpart. This motif was continually repeated in the set of oligopeptide macrofragments and comes about because in this conformation the carbonyl group can make very favorable hydrogen bonds with the flap water. In the known inhibitor, JG-365, a proline ring and added backbone methyl group disfavor this carbonyl orientation.

The good fit between the lowest energy oligopeptide conformation in this simulation and the known inhibitor prompted us to run a second CONCERTS simulation where we started with the low-energy polyglycine oligopeptide shown in Figure 3 and attempted to add amino acid side chains to it. The simulation was run starting with 10 copies of the polyglycine peptide as a "seed", and amino acid side chains were attached to the C_α positions during the run. A total of 150 fragment copies of each standard amino acid side chain were included in this simulation, as described in the Methods section. Both the oligopeptide and the side chain fragments were allowed to move during the simulation. By the end of the simulation, a peptide side chain had been added to each of the six α carbons in each of the 10 oligopeptides. This two-step approach to peptide ligand design (first design the backbone, then add the side chains) has precedence in earlier ligand design studies.[20]

The structure with the lowest energy (and fifth lowest

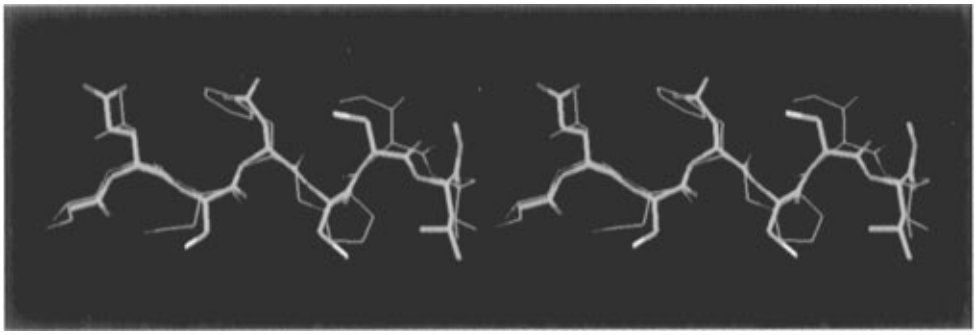

Figure 4. The lowest energy oligopeptide + side chain ligand prediction from a CONCERTS simulation against HIV-1 aspartyl protease (thick lines). The oligopeptide is compared to JG-365 (thin lines). The molecules are in the same protein reference frame. Coloring and presentation are as in Figure 3. Agreement between the entirely *de novo* designed molecule and the known drug is striking.

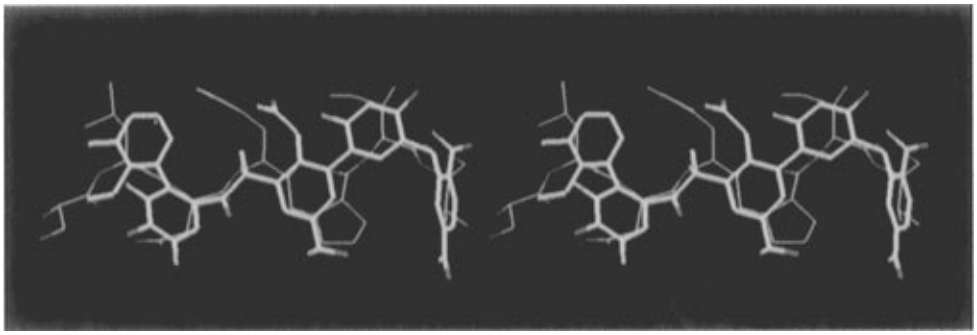

Figure 5. A macrofragment CONCERTS prediction for HIV-1 aspartyl protease, using a limited set of fragments (set "B"; see text). The molecule shown (thick lines) was generated by forming a natural connection between two of the lowest energy suggestions from the CONCERTS run (each of which spanned approximately half of the active site). A three-benzene-ring tail that falls entirely outside the active site, but which was appended to the CONCERTS prediction, has been omitted from this figure for clarity. For comparison, JG-365, in the same protein reference frame, is shown (thin lines). Coloring and presentation are as in Figure 3. The carbon atoms which formed the connection between the two original suggested halves are shown in white. The CONCERTS prediction does a reasonable job of following the backbone of the known drug and also places atoms at or near the locations of all side chains of the drug.

$\sum E_{\mathrm{nb(relative)}}$) resulting from this simulation is shown in Figure 4 and again compared to the known inhibitor. It is clear that with the added side chains, the fit of the backbone of the *de novo* prediction to the known drug has gone from good to outstanding. The exact side chains in the known inhibitor are not replicated, but the orientations and conformations of the side chains which were chosen match very closely. The question of which side chains would, in fact, lead to better binding is probably beyond the ability of the model and parameters we are using. Nonetheless, the *de novo* prediction is obviously acceptable as a starting lead for experimental drug design.

The surprisingly good agreement between the CONCERTS prediction and experiment with the peptide fragments provided us with confirmation that the method can work. We were then encouraged to apply the approach using the more diverse fragment sets described in the Methods section. We first applied the "limited" fragment set (B), consisting of benzene, ammonia, formaldehyde, methane, and water. This set was designed as a compromise between (1) using a small number of groups (thereby reducing the compositional complexity of the space being searched); (2) including groups that can act as both hydrogen bond donors and acceptors; and (3) including groups that can efficiently span conformational space. This set may also help limit the number of chiral centers in any suggested macrofragment, since with these fragments a chiral center can only arise from a tri- or tetrasubstituted methane fragment.

This simulation resulted in a variety of interesting molecules that span the protein active site, and which fill the same regions in space as the known inhibitor. A total of 138 macrofragments (four or more fragments) were generated. Particularly interesting among these were two of the lowest energy macrofragments, each of which filled half of the active site. A natural connection between these macrofragments is suggested which then covers the entire active site. The resulting structure also possesses the lowest $\sum E_{\mathrm{nb(relative)}}$ of any macrofragment suggested. $E_{\mathrm{nb(relative)}}$ for each fragment is the nonbonded interaction energy of that fragment minus the average nonbonded interaction energy for that fragment type. The resulting macrofragment is shown in Figure 5. As can be seen, this macrofragment does an excellent job of following the backbone of the known peptide inhibitor. In addition, atoms are placed at or near all six regions of space that are occupied by side chains in the peptide inhibitor. Particularly appealing, from a synthetic point of view, is that this macrofragment contains only a single chiral center. We note that a three-benzene-ring molecular "tail" generated with this molecule has been omitted from Figure 5. This tail falls outside the active site and only makes a few surface contacts with the protein. It was generated as part of the inhibitor only because we used a liberal set of dimensions when defining the cap on the active site, so as not to spuriously preclude any important structures.

Given our relative successes with runs using limited libraries of fragments, we proceeded to a third type of simulation which incorporated a wide variety of frag-

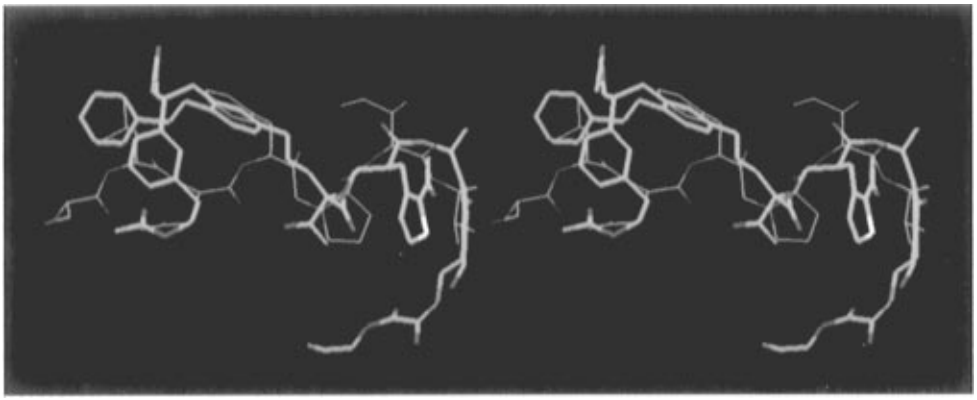

Figure 6. A macrofragment CONCERTS prediction for HIV-1 aspartyl protease, using the "diverse" set of fragments (set "C"; see text). The molecule shown (thick lines), which is among the most interesting, corresponds to the sixth lowest relative energy. JG-365 is shown in the same protein reference frame, for comparison (thin lines). Coloring and presentation are as in Figure 3. The CONCERTS prediction places atoms in all but one of the regions of space occupied by the side chains of JG-365 and takes a novel path to placing a benzene ring in a position nearly identical to that of the phenylalanine side chain in P1.

ment types (set "C" in the methods section). At the onset, we realized that a large fragment library might present so many possibilities that convergence would not occur in a feasibly small number of steps. In practice, we obtained mixed results. As with the limited fragment set (B), we did obtain a number of very interesting macrofragments suggestive of known drugs. But the number of such structures appears to be smaller with fragment set C than with fragment set B. In all, 151 nontrivial macrofragments (four or more fragments) were generated. One of the more interesting predictions is presented in Figure 6. This macrofragment corresponds to the 6th lowest $\sum E_{\text{nb(relative)}}$. Comparison with the prediction in Figure 5 reveals that this macrofragment is not in as good agreement with the known drug. In particular, it does not mimic the peptide backbone of the drug in the nonprime region. But it *does* do a good job of placing atoms in the regions of space occupied by all but one of the side chains of the drug (the space occupied by the serine side chain is missed). And the macrofragment does a notably good job of mimicking the position of the phenylalanine side chain in P1.

FKBP-12. The same fragment sets were next applied to the FKBP-12 protein. Three sets of CONCERTS simulations were carried out, using the peptide fragment set, the "limited" fragment set, and the large "diverse" fragment set.

The predictions obtained by applying the peptide fragment set to FKBP-12 were surprisingly revealing. Specifically, a number of macrofragments were in outstanding agreement with the postulated "binding core" of FK506,[6] a potent nanomolar binder of FKBP-12. One of these macrofragments—corresponding to the 9th lowest $\sum E_{\text{nb(relative)}}$, and the lowest energy 7-mer in a set of 115 macrofragments (four or more fragments)—is shown in Figure 7, along with FK506. It can be seen that the macrofragment overlaps the backbone of the FK506 in the binding core extremely well—in fact, agreement would be nearly perfect with a small translation in the oligopeptide relative to FK506. FKBP-12 is a proline cis−trans rotamase,[21] and it is believed that the pipecolic group in FK506 mimics the binding conformation of the proline ring and amide carbonyl in the bound oligopeptide.[22] The macrofragment suggests how an oligopeptide could be stabilized in the active site by favorable hydrogen-bonding interactions with two carbonyl groups of peptide residues near the proline of interest. Comparison with FK506 suggests why this inhibitor, which is non-peptidic but which places two carbonyl groups in the same positions as the macrofragment, might also bind well to FKBP-12. It is interesting tht the macrofragment places atoms in the regions occupied by the pyranose and cyclohexyl rings in FK506. It has been suggested that the cyclohexyl portion of FK506 is important to recognition by calcineurin.[6]

Application of CONCERTS using the "limited" fragment set (B) yielded a variety of interesting macrofragments. In all, 122 macrofragments consisting of four or more fragments were generated. These structures variously place atoms in the binding core region of the active site, in the pyran region, in the vicinity of the cyclohexyl ring (the "northwest corner"), and in the other regions of space also occupied by the FK506 macrocycle. A large number of the CONCERTS predictions place atoms in many/all of these regions. The downside is that, because the binding site in FKBP-12 is rather wide and accessible, large macrofragments are frequently generated where a significant number of fragments fall at the edges (or outside) of the active site. This side effect could be reduced, of course, by decreasing the volume of the acceptable search space for the CONCERTS run. At any rate, one can very instructively focus on the portions of the macrofragments that fall closest to the active site.

In Figure 8, we present the "limited" fragment set prediction with the third lowest $E_{\text{nb(relative)}}$. For clarity and emphasis, we have removed a branch of the predicted macrofragment which pointed out of the active site. The remaining portion of the molecule is shown, along with FK506. It can be seen that the CONCERTS prediction is in extremely good accord with the known drug. Not only does the macrofragment place oxygens close to four important binding core oxygen positions in FK506 (the oxygens of both carbonyls in the dicarbonyl region, the hydroxyl oxygen off the headgroup of the pyranose ring, and the carbonyl oxygen in the ester group adjacent to the pipecolinyl ring), but it also faithfully follows the general path of the FK506 backbone through most of its circumference. One could imagine building off this scaffold to reproduce most of the functional features of FK506. And, importantly, this

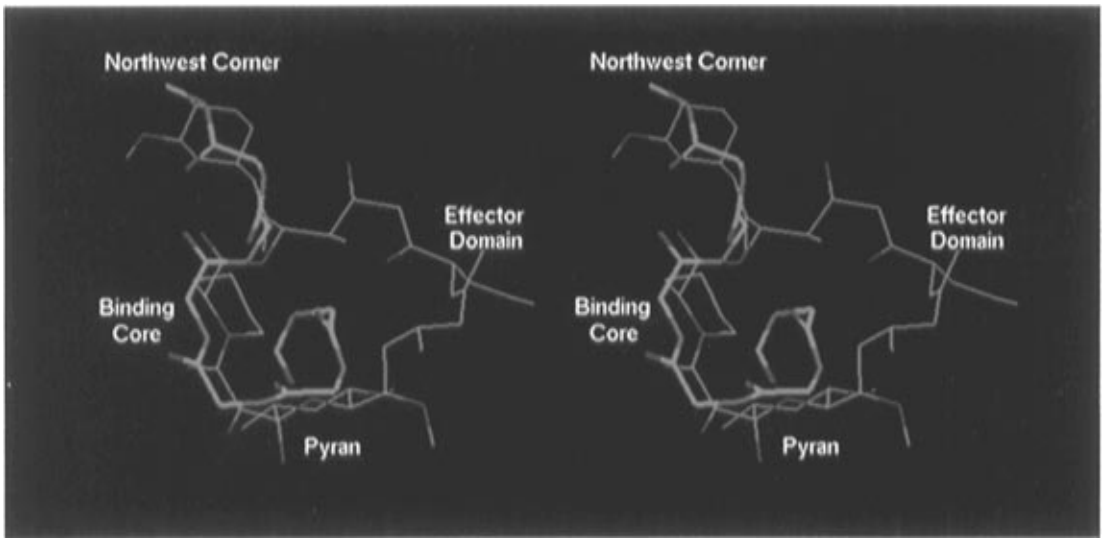

Figure 7. A low-energy oligopeptide generated by CONCERTS against FKBP-12 (thick lines). This oligopeptide is compared to FK506, a nanomolar binder of the protein (thin lines). The molecules are in the same protein reference frame. Carbon atoms are in green, oxygen atoms are in red, and nitrogen atoms are in blue. Hydrogen atoms are not shown. Important domains in FK506 are labeled: the "northwest corner", which includes the cyclohexyl ring and the arm leading to it; the "binding core" which is believed to be responsible for tight binding to FKBP, and which includes the pipecolinyl ring and surrounding carbonyl groups; the pyran ring region; and the "effector domain", which includes atoms that are believed to make contact with calcineurin in a FKBP-12−FK506−calcineurin complex, but that do not make appreciable contact with FKBP-12 itself. The figure is presented as a stereoview. The peptide prediction very closely reproduces the conformation of FK506 in the important "binding core" region, as well as placing atoms in the "northwest corner" and pyran region.

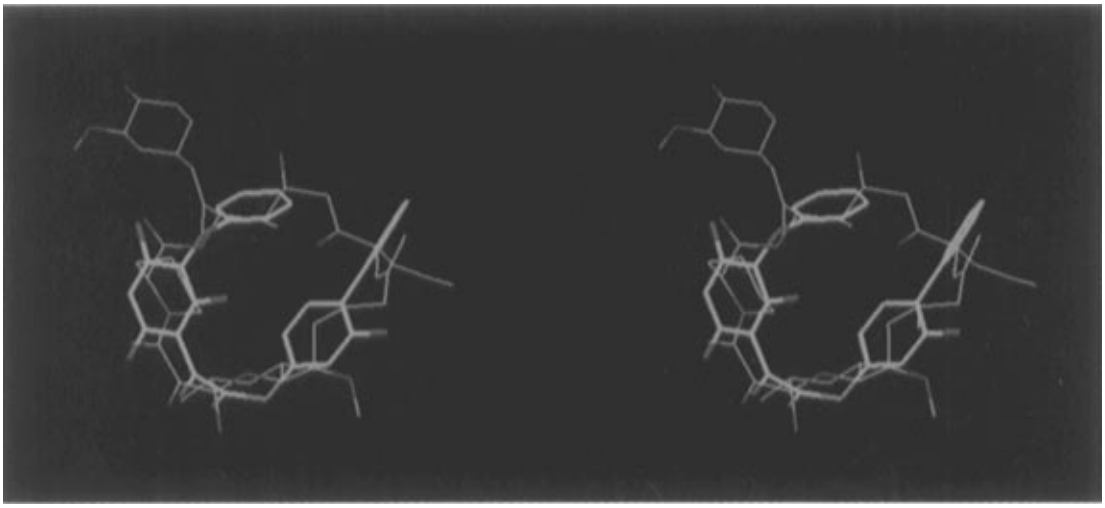

Figure 8. A macrofragment CONCERTS prediction for FKBP-12, using a limited set of fragments (set "B"; see text). This molecule (thick lines) corresponds to the third lowest relative energy of the macrofragments generated. A branch of the predicted ligand that falls entirely outside the active site has been removed from this figure to emphasize the important features. For comparison, FK506, in the same protein reference frame, is shown (thin lines). Coloring and presentation are as in Figure 7. The CONCERTS prediction mimics four important oxygen atoms found in FK506 and also faithfully follows the overall path of the FK506 backbone through most of its circumference.

molecule contains no chiral centers, making it potentially a more synthetically accessible target.

We followed up the "limited" CONCERTS simulation on FKBP-12 with a second simulation using the "diverse" fragment set (C). As might be expected, the resulting set of macrofragments exhibited great variety. A total of 130 macrofragments (four or more fragments) were generated. While a number of these were promisingly suggestive of features of the FK506 inhibitor, the number that placed atoms in the regions covered by FK506 that are believed to be important for binding was smaller than that observed with the limited fragment set run. Instead, for many of the macrofragments, a majority of the constituent groups were essentially outside or on the edge of the active site. While these groups might make favorable interactions with surface residues of the protein, or with other residues of the macrofragment itself, they are not particularly suggestive of good drugs. It is likely that the increased number of possibilities (both compositional and conformational) presented by the diverse set of fragments means that a relatively small fraction of the macrofragments in this run are sufficiently near a converged low-energy configuration.

Nonetheless, a few promising structures emerge. One of these is shown in Figure 9. The structure shown corresponds to the 18th lowest $E_{\mathrm{nb(relative)}}$, out of a set of 130 macrofragments (four or more fragments) generated from this run. As can be seen, this macrofragment does a good job of following the FK506 backbone in the region

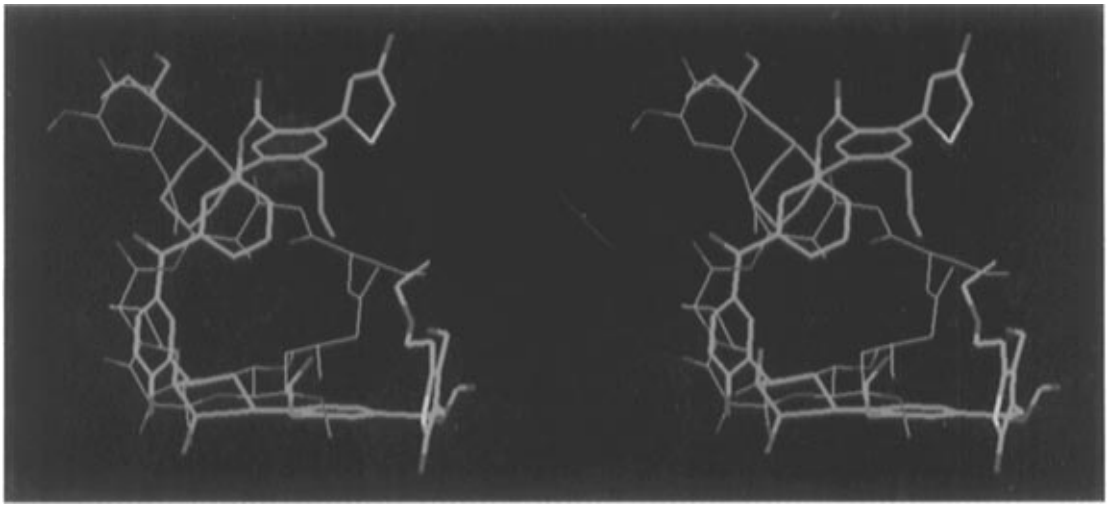

Figure 9. A macrofragment CONCERTS prediction for FKBP-12, using the "diverse" set of fragments (set "C"; see text). The molecule shown (thick lines), which is among the most interesting, corresponds to the 18th lowest relative energy. FK506 is shown in the same protein reference frame, for comparison (thin lines). Coloring and presentation are as in Figure 7. The CONCERTS prediction does a reasonable job of reproducing the features of FK506 in the binding core and also places atoms in the northwest corner and pyranose regions. Note the considerable amount of structure which is superfluous relative to FK506.

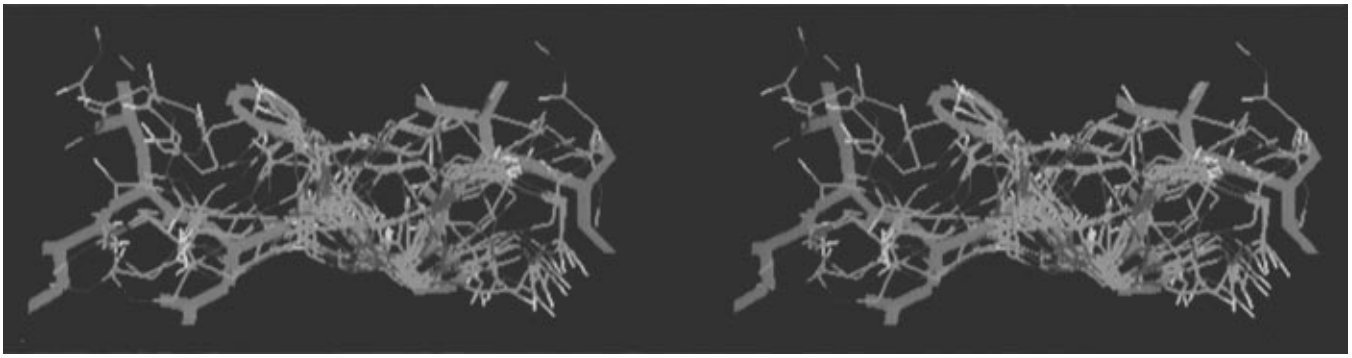

Figure 10. The 20 hexamer and larger macrofragment predictions for the peptide fragment set run in the HIV-1 aspartyl protease active site. For comparison, known inhibitor JG-365 is shown in heavy orange lines. For the ligand predictions, carbon atoms are shown in green, oxygen atoms are in white, and nitrogen atoms are in blue. Hydrogen atoms are not shown. The figure is a stereoview. As can be seen, the oligopeptide ligand predictions effectively sample the active site.

of the binding core. It also places two oxygen atoms in similar locations to those in FK506 (overlapping that of the second carbonyl in the dicarbonyl region, and that of the ester carbonyl). This macrofragment presents a common motif observed among the predicted macrofragments in this run and in the previous run: a modest projection away from the backbone of the FK506 molecule, and out of the active site, which then folds back to place atoms in the "northwest corner" region occupied by the cyclohexyl ring in FK506. Compared with the prediction from the "limited" CONCERTS run, this macrofragment is less concise and, outside of the binding core/pyranose region, does a poorer job of following the circumference of FK506. This molecule is also considerably more complex from a synthetic standpoint: it contains seven chiral centers, compared to none in the molecule presented in Figure 8.

Sampling Issues. Two questions arise in prediction studies like those described here. First, how thorough is the sampling? And second, how predictive is the energy function? The first question can be addressed by examining the space covered, in total, by the predicted ligands. To this end, in Figure 10 we show the 20 hexamer and larger macrofragments generated during the peptide fragment run against HIV-1 aspartyl protease. As can be seen, this ensemble (from which the macrofragment in Figure 3 was extracted) does an excellent job of sampling the active site of the protein.

The second question, how predictive is the energy function, is more difficult to address. But a crude qualitative answer can be obtained by comparing the calculated macrofragment energies to how well the macrofragments address elements of a model of desirable ligand interactions. This model should be derived from an ensemble of tightly binding known inhibitors. For HIV-1 aspartyl protease, we have generated such a model from a set of inhibitors which strongly bind the protease active site.[9] Although hydrophobic interactions can also be significant, for simplicity we have focused on hydrogen binding. The ensemble of known inhibitors suggests that hydrogen bonds between the ligand and residues Asp^{25}, Gly^{27}, Gly^{48}, Asp^{125}, Gly^{127}, Gly^{148}, and between the ligand and the flap water, are generally conserved. In Figure 11, we have plotted the number of appropriate hydrogen bonds formed with these residues versus the E_{nb} for the oligopeptide inhibitors, both normalized by the number of macrofragment residues. As can be seen, there is a distinct correlation between the two quantities, with the number of important hydrogen bonds decreasing with increasing energy. This implies, at least qualitatively, that our scoring function is reasonable and that the best-scoring *de novo*

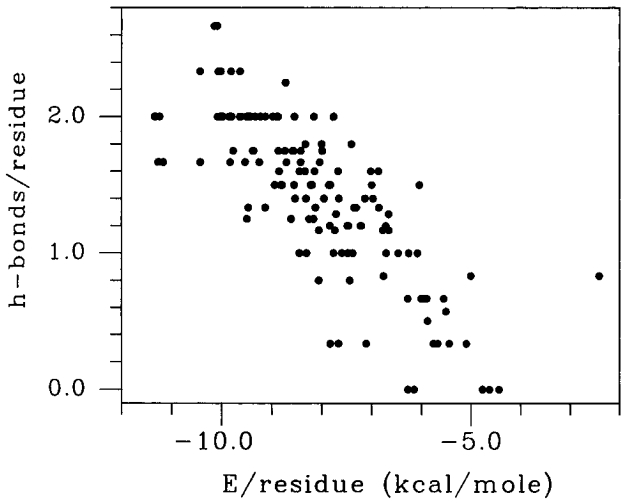

Figure 11. The number of important hydrogen bonds formed per residue versus the macrofragment−protein interaction energy for the peptide fragment/HIV-1 aspartyl protease simulation, both normalized by the number of residues in a macrofragment. Only hydrogen bonds to residues believed to be important for good ligand binding, based on an ensemble of known tightly-binding inhibitors, are counted (see text).

suggestions produced by CONCERTS do a good job of mimicking the specific H-bonding interactions found in known, tight-binding HIV-1 protease inhibitors.

Discussion

We have described and applied a new *de novo* ligand design method, CONCERTS, to the active sites of HIV-1 aspartyl protease and FKBP-12, using three different sets of fragment building blocks. The results of these calculations are quite encouraging. In particular, in each simulation we were able to generate interesting ligands which are suggestive of known inhibitors of these proteins.

It is instructive to compare the types of predictions generated with the different sets of fragments. For example, runs using only peptide fragments appear to produce a greater number of predictions which are in excellent accord with the experimentally known inhibitors. The simulations using a variety of fragments produce structures which are more diverse, and some of the predictions from these runs are also in good agreement with known inhibitors. But others are not, and this tendency appears to increase as the diversity of the basis fragment set is increased. Of course, just because a predicted molecule does not look like a known inhibitor does not mean that the prediction is not valuable. In fact, such molecules may well be the *most* valuable, since they might lead to novel classes of leads.

What is almost certainly true, however, is that the chance of converging on high-quality inhibitor leads is higher for simulatioins using a limited number of fragments. This is apparent comparing the results from simulations using limited sets of organic fragments with those from simulations using more diverse sets of such fragments. If we take a crude qualitative measure of convergence to be the number of resulting macrofragments covering the regions of the active site we believe to be important, the less diverse fragment sets appear to result in greater convergence. This is not surprising, of course. All this says is that for the more diverse fragment sets it may be necessary to run the simulations for a greater number of steps to achieve equivalent convergence. Unfortunately, the rate of convergence should also be improved by including larger numbers of replicates of each fragment type, which will slow down the calculation in an approximately linear fashion. So not only do we need to perform more MD dynamics steps, but each of these steps is also slower. The tradeoff is that with simple fragment sets, one may build in bias (based on the set used) that results in missing interesting possibilities.

Perhaps most provocative of the predictions obtained using CONCERTS was the result from the peptide fragments + side chains simulation for HIV-1 aspartyl protease. In that case, we were able to generate, entirely *de novo*, an oligopeptide inhibitor that is in surprisingly good agreement with a known peptide inhibitor. The result is good enough, in fact, that it would easily serve as a usable scaffold for subsequent drug-discovery work. This serves as clear affirmation of the potential power of the CONCERTS method.

In developing CONCERTS, we set out to preserve some of the features of the CONCERTS approach we published earlier, and to rectify some of the drawbacks we had encountered. In particular, we shifted from an atom-based method to one where the building blocks are fragments, and we expanded the method so that more than one inhibitor suggestion would be made during any single CONCERTS run. Our intent was both to improve the efficiency of the method and to allow a natural way to include charged groups. From the results we have obtained, we believe we have succeeded on both counts. One particular advantage of CONCERTS over CONCEPTS is that we now have a good way of including hydrogen-bonding groups (which include partial charges) in the ligand buildup scheme. The inclusion of such groups is clearly crucial for detailed positioning of inhibitors like the HIV protease oligopeptide, and for appropriate alignment in the "binding core" region of FK506. Unlike CONCEPTS, CONCERTS also allows for the straightforward inclusion of torsional energy terms for macrofragments. Calculating these terms will certainly improve the predictive qualities of the force field used and, ultimately, of the macrofragments generated.

Also of note is the fact that the most "interesting" predicted inhibitors are not always those with the very lowest absolute (or relative) energies. The scoring function, which is based on the Weiner *et al*. force field typically used with AMBER,[12] is reasonably good, but far from perfect. Other elements of the model (fixed protein, no explicit solvent, potential energies rather than free energies, etc.) add to the approximate nature of the energy predictions. For these reasons, it is valuable to examine a range of low-energy structures, not just one or two. With this in mind, the CONCERTS method has been designed so that we obtain a wide range of bound conformations. Of course, without reference to known inhibitors—surely the case in the most potentially valuable applicatioins of a *de novo* approach—we cannot know *a prioro* which of the low-energy predictions are "best". In such a case, one might attempt to discern trends among the predictions or else might try to perform more elaborate modeling on a few promising structures (e.g. while letting the protein move and with explicit solvent included) to attempt to better

rank them before moving ahead. Simple screens are also possible. For example, macrofragments which do not make sufficient close contact with the protein can be eliminated.

A number of improvements and additional experiments present themselves. For example, one could modify the program to facilitate the creation of additional types of structures. By incorporating the ability to perform "fusion" between properly oriented closed ring systems one could increase the number of structural classes that could be generated without increasing the number of fragments in the basis set. One could also envision coupling CONCERTS to methods like CAVEAT[23] or HOOK,[24] which are efficient at connecting disjoint fragments. In such an approach, CONCERTS could be used to generate a series of small tightly binding macrofragments, and these could be fed into the other programs. CONCERTS would then serve a similar purpose to the multiple copy simulataneous search approach, MCSS,[25] and the combination of CONCERTS and CAVEAT is conceptually quite similar to the use of MCSS and HOOK in tandem. This might be more efficient than requiring CONCERTS to also generate reasonable linker regions between the tightly binding subunits. Along these lines, it is worth noting the work of Caflisch *et al.*[20] These authors used simple connection criteria and clustering methods to join up a static set of fragments that had been distributed using MCSS MD followed by MCSS minimization. Of all the published *de novo* design methods, this is probably closest in spirit to the method described here. Their work demonstrated that even such a static approach, one where, unlike CONCERTS, molecular conformation does not continually reflect changes in the topology during the buildup procedure, is capable of generating molecules which are suggestive of known inhibitors. It is also worth noting the recently-published MCDNLG method,[26] which shares similarities to both CONCERTS and the atom-based CONCERTS antecedent, CONCEPTS. In MCDNLG, a randomly chosen collection of atoms is tightly packed into the active site of a receptor, and this set of atoms is then slowly annealed into a chemically stable molecule, using a Monte Carlo search algorithm. MCDNLG has shown promise in trial applications to DHFR, thymidylate synthase, and HIV-1 aspartyl protease.

One can also develop other basis sets of fragments than those used here. The key in this respect will be balancing diversity in the fragment set with decreases in the speed at which the simulation will converge. For example, one could use a full set of amino acid fragments in place of the two-step procedure employed here where side chains were added to a polyglycine chain. In fact, there are likely many situations where the two-step process used here would not work so well, situations where the gross conformation of the backbone is not strongly dictated by hydrogen-bonding interactions with groups lining the binding site. On the other hand, it is far from certain that, if we use a diverse fragment set at the onset, we will efficiently converge on simularly good predictions. The process of adding a large set of side chains onto a well-chosen low-energy oligopeptide scaffold might well converge more quickly than a procedure where one needs to handle all the diversity from the beginning of the simulation.

It is obvious that the CONCERTS approach requires choosing a significant number of constants which define, for example, the relative weights of terms in the force field, the number of replicates of different fragment types, the annealing temperature, and so on. The results we have presented herein are based on a series of simulations carried out using a constant set of control parameters. Only by utilizing a consistent set of simulation conditions, as we have done, can one develop faith that the method wil be generally applicable. We have carried out a limited number of characterization runs (unpublished results) which have led us to believe that the parameter sets used herein are reasonable. However, it is likely that future refinement of these parameters could lead to even better simulation results. This is beyond the scope of the current paper and is left for subsequent studies.

It is interesting to note that CONCERTS parallels the approach of combinatorial chemistry in that both can be applied to determining which sets of functional groups may result in the "best" binding ligand, given a chosen scaffold. However, we are currently limited by cpu and memory considerations to considering only a modest number of fragment types in the active site during any single run. By contrast, in the type of semiexhaustive search for which one frequently employs combinatorial chemistry, it is possible to screen complexes with hundreds or thousands of fragment types. The greatest strength of CONCERTS at present is probably its ability to suggest scaffolds for future ligand design.

Overall, it appears that CONCERTS is another step forward in progress toward a truly useful *de novo* drug design approach. The underlying key to all these methods is sampling efficiency. The fragment-based approach used by CONCERTS, and the ability of fragment−fragment connections to change (evolve) over time, make the search performed by CONCERTS particularly efficient. It is expected that methods of this type, where a major bottleneck is performing sufficient sampling to obtain converged results, will become ever more valuable as computer resources continue to become cheaper. In the meantime, CONCERTS appears to be efficient enough that it offers, even today, the chance to generate genuinely useful predictions.

Appendix

Detailed Program Description. CONCERTS has been implemented in a modified version of the AMBER/SANDER 4.0 minimization/molecular dynamics program.[10,11] Both CONCERTS and associated program FLOAT are written in Fortran. An overview of the conceptual flow of the program is presented in Figure 1. The implemention is described below in detail.

(1) One selects the number and types of the fragments to be used in the simulation. A molecular fragment can be any residue, functional group, or molecule. In general, one chooses either standard functional groups or polymeric building blocks (e.g., amino acids or nucleotides), but this is not required. Most fragments have a net charge of zero, while some, like amino acid side chains containing a carboxylate group, have a net integral charge. Fragments are defined as in standard AMBER:[11] the topology is specified using a "tree" nomenclature, and each atom is assigned a parameter

type and a charge. Limits on the numbers of various fragments are required to keep the number of nonbonded interactions (and hence cpu time required) tractable and, to a lesser extent, to keep the required computer memory down to a reasonable level.

(2) "Connection vectors" are chosen for each fragment type. A connection vector is a heavy atom−hydrogen bond. During the CONCERTS simulation, bonds between fragments can only be formed along such connection vectors and are created by removing the hydrogens and covalently linking the two associated heavy atoms:

$$(\textit{heavy atom})-(\textit{hydrogen } 1) + (\textit{heavy atom } 2)-(\textit{hydrogen } 2) \rightarrow (\textit{heavy atom } 1)-(\textit{heavy atom } 2)$$

Each fragment type can have as many connection vectors as is desired. Generally, most of the hydrogen−heavy atom bonds in the system are designated as connection vectors. When a fragment−fragment connection is formed, the associated hydrogens effectively "disappear" from the system, and the charges on the hydrogens which have been removed are added into the attached heavy atoms to maintain a constant net charge. (In the following discussion χ_{in} refers to the nth connection vector located on fragment i; likewise, χ_{jm} refers to the mth connection vector on fragment j.)

(3) A protein of interest is chosen whose structure is known. One defines a volume ("binding region") centered on the binding site for this protein. The binding region is specified by user-supplied ranges in x, y, and z coordinates.

(4) The fragments are randomly oriented in the chosen binding region using our program FLOAT, which implements an idea described by Hart and Read.[27] In this program, a fragment is randomly placed (sometimes in the interior of the protein) and then moved using a series of Monte Carlo translations and rotations that bias the fragment to "float" toward the surface of the protein. The bounds defined in step (3) are incorporated to ensure that the fragment floats to the surface in the chosen active site. The float procedure is performed for the every copy of every fragment used.

(5) "Consensus minimization" is performed on the randomly oriented fragments. In consensus minimization, the protein is fixed, and the fragments do not see one another. The fragments themselves are not rigid. Thus, the only nonbonded interactions in the system are *intrafragment* and fragment−protein. Energy terms take the standard forms.[12] The net energy function used is:

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{intrafragment}} + E_{\text{fragment-protein}} + E_{\text{cap}} \quad (1)$$

where

$$E_{\text{bond}} = \sum K_{\text{r}}(r - r_{\text{eq}})^2 \quad (2)$$

$$E_{\text{angle}} = \sum K_{\theta}(r - \theta_{\text{eq}})^2 \quad (3)$$

$$E_{\text{torsion}} = \sum \frac{K_n}{2}[1 + \cos(n\phi - \gamma)] \quad (4)$$

$$E_{\text{cap}} = \sum_{R_{i\rightarrow\text{cent}} > R_{\text{cap}}} K_{\text{cap}}(R_{i\rightarrow\text{cent}} - R_{\text{cap}})^2 \quad (5)$$

and both $E_{\text{intrafragment}}$ and $E_{\text{fragment-protein}}$ are given by the same series of nonbonded terms (E_{nb}), summed over the appropriate subset interactions:

$$E_{\text{nb}} = \sum_{i<j}\left\{\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{\text{ij}}^{6}}\right] + q_i q_j/\epsilon R_{ij}\right\} + \sum_{\text{H-bonds}}\left\{\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}}\right\} \quad (6)$$

E_{total} is the potential energy of the system; K_{r} and r_{eq} are the bond stretching constant and the equilibrium bond distance; K_θ and θ_{eq} are the bond angle stretching constant and the equilibrium bond angle; K_n, n and γ are the torsional force constant, the periodicity of the torsional term, and the phase angle; A_{ij} and B_{ij} are the nonbond (Lennard-Jones) repulsion and attraction coefficients; R_{ij} is the interatomic distance between atoms i and j; q_i and q_j are the atomic partial charges on atoms i and j; and ϵ is the effective dielectric constant. At this stage, the bond, angle, and torsion terms are summed only over intrafragment interactions. E_{cap} is a positional restraint term that ensures that the fragments do not drift more than a specified distance, R_{cap}, from the center of the active site. $R_{i\rightarrow\text{cent}}$ is the distance of atom i from the prespecified center of the "cap". K_{cap} is the cap force constant. Standard (unscaled) forces, as calculated from the first derivative of the above equation, are applied to all fragment−protein interactions. Consensus minimization is required to relax the system in preparation for consensus MD (below). Minimization is performed using a combination of steepest descents and conjugate gradients until a maximum rms gradient of 0.1 kcal/mol·Å is achieved.

(6) "Consensus MD" is performed on the system, starting with the results from consensus minimization. The same energy function is used, and once again the protein is fixed and fragments do not see one another. During consensus MD, the fragments move about the active site independently of one another. At the end of consensus MD, one will have generated an ensemble of the more favorable positions and orientations for each fragment type. The resulting configuration will be used as input into the CONCEPTS program. Note that consensus MD is similar to the MCSS method.[25]

(7) Start the actual CONCERTS simulation. The simulation starts with the equilibrated system generated as described in step (6).

(8) MAXSTEP is the total number of MD steps to be performed. N is the MD step interval between iterations where CONCERTS checks for possible fragment−fragment connections. If the current step number is larger than MAXSTEP − N, the CONCERTS simulation is terminated. Otherwise, take N consensus MD steps. The protein is held fixed. The energy function used during consensus MD in CONCERTS is

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{intra(macro)fragment}} + E_{\text{(macro)fragment-protein}} + E_{\text{cap}} \quad (7)$$

where $E_{\text{intra(macro)fragment}}$ and $E_{\text{(macro)fragment-protein}}$ each represent the sum of three terms, as given by eq 6. This function is very similar to that used during the equili-

bration phase consensus MD (eq 1), and the functional forms of the individual terms are identical. The difference is that, in CONCERTS, fragments can be joined together. Once fragments are joined together (to form a "macrofragment", i.e., any set of two or more bonded fragments), the internal energy contributions (bond, angle, torsion) are calculated not only for the individual fragments, but also for any additional internals that are formed between the joined fragments. Nonbonded interactions are also calculated between any fragments which are part of the same macrofragment (subject to standard 1−2 and 1−3 interaction exclusions[28] across the fragment−fragment bond). Fragments which are not part of the same macrofragment still do not see each other, whether or not they belong to other macrofragments.

(9) Start at a randomly chosen connection vector χ_{in}, the nth connection vector on fragment i. This and the following steps will be repeated until every connection vector χ_{in} in the system has been considered. When all connection vectors have been considered for this cycle, return to step (8) and carry out another N steps of consensus MD.

(10) Compare the nonbonded energy of fragment i, E_i, to the average energy for all fragments of the same type. If

$$E_i \leq \langle E \rangle_i + C\sigma = \Sigma_i \tag{8}$$

where $\langle E \rangle_i$ is the average value of E_i for all fragments of the same thpe as i, σ is the standard deviation in the values of E_i for all fragments of the same type as fragment i, and C is a user defined constant, then proceed to step (11). If eq 8 is not satisfied, do not look for new bonding partners for fragment i (although we keep any fragment−fragment bonds already made to j), and return to step (9). Limiting the search to fragments whose energy is more favorable than an average-based threshold has precedence in the GroupBuild method.[29]

(11) Build up a list of all "potential bonding partners" along χ_{in}. This is a list of all connection vectors χ_{jm} which meet the geometric, energetic and atom sequence criteria as specified in the following steps. To create this list, consider a connection vector χ_{jm} on residue j ($j \neq i$) which has not yet been considered for χ_{in}. At this stage, every connection vector χ_{jm} ($j \neq i$) is potentially a connection partner for χ_{in}, even a χ_{jm} currently used in a connection to another fragment. If all potential connection vectors χ_{jm} have been considered, skip to step (17). Otherwise, if connection vectors χ_{in} and χ_{jm} are already being used to bond fragments i and j to each other, store χ_{jm} in the potential bonding partners list, and consider the next X_{jm}. If not, subject χ_{jm} to the following tests.

(12) Is $E_j \leq \Sigma_j$? If not, go back to step (11) and consider another connection vector χ_{jm}.

(13) Check to see if the potential fragment−fragment connection between residues i and j is allowed. The user can specify sequences of atoms (atom−atom and atom−atom−atom) which are disallowed based on atom name, atom type, or parameter type. If the proposed connection is disallowed, go back to step (11) and consider another connection vector χ_{jm}.

(14) Determine the fragment−fragment connection bond distance r and all valence angles θ that would result if a fragment−fragment bond was made along $\chi_{in} - \chi_{jm}$. Compare these to the target "ideal" values r_0 (bonds) and θ_0 (angles). Are

$$r - r_0 \leq \delta_r \tag{9}$$

$$\theta - \theta_0 \leq \delta_\theta \tag{10}$$

where δ_r and δ_θ are specified by the user? If not, go back to step (11) and consider another connection vector χ_{jm}.

(15) If the connection vector χ_{jm} is already being used to make a fragment−fragment bond with some fragment k, compare the relative nonbonded energy of that fragment, $E_{k(\text{relative})}$, to the nonbonded energy of fragment i. If $E_{k(\text{relative})} \leq E_{i(\text{relative})}$, go back to step (11) and consider another connection vector χ_{jm}. $E_{i(\text{relative})}$ is the nonbonded energy of fragment i minus the average nonbonded energy for all fragments of the same type.

(16) The proposed connection satisfies all necessary criteria. Add χ_{jm} to the "potential bonding partners" list and return to step (11) to consider another connection vector χ_{jm}.

(17) At this point, all possible connection partner/vectors χ_{jm} to connection partner/vector χ_{in} have been considered, and the "potential bonding partners" list of those meeting the appropriate criteria has been formed. If the list is empty, no partners are available for χ_{in}. In this case, return to step (9) and consider another χ_{in}. Otherwise, consider the potential partner on the list with the lowest fragment−protein nonbonded energy.

(18) Calculate the nonbonded + electrostatic interaction *between* the macrofragments containing the fragments i and j (excluding 1−2 and 1−3 interactions that would result from creating the bond along χ_{in} and χ_{jm}). If this interaction energy is greater than a user specified limit, E_{overlap}, remove χ_{jm} from the list of potential partners and return to step (17).

(19) Accept the macrofragment connection $\chi_{in} - \chi_{jm}$. Make the changes in topology to form a bond between fragments i and j, along χ_{in} and χ_{jm}. If either connection vector was formerly used to make a bond to a different fragment, first delete that bond and free up the associated connection vectors. When a connection vector is freed up, the hydrogen atom at the end of the bond (which was removed when the fragment−fragment bond was created) reappears, and the appropriate charge is transferred from the attached heavy atom back to the hydrogen itself. Return to step (9) to consider another χ_{in}.

When the CONCERTS procedure described above terminates (after the specified number of steps of MD have been run; see step 8), one obtains a set of many macrofragments. Each of these can be considered a ligand "suggestion". A small amount of consensus minimization (using the energy function given in eq 7, i.e., where the macrofragments do not interact with one another) is subsequently performed on the system to reduce internal strain in each macrofragment. The resulting macrofragments can be visualized and analyzed. The number of macrofragments resulting from any run depends on the number and types of fragments used in the run and on the various user-set parameters governing the run.

References

(1) Desai, M. C.; Zuckerman, R. N.; Moos, W. H. Recent Advances in the Generation of Chemical Diversity Libraries. *Drug Devel. Res.* **1994**, *33*, 174−188.

(2) Murcko, M. A. In Recent Advances in Ligand Design Methods. Lipkowitz, K. B., Boyd, D. B., Eds.; Reviews in Computational Chemistry; VCH Publishers: New York, in press.
(3) Pearlman, D. A.; Murcko, M. A. CONCEPTS: New Dynamic Algorithm for *de novo* Drug Suggestion. *J. Comput. Chem.* **1993**, *14*, 1184−1193.
(4) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: New York, 1987.
(5) Rosen, M. K.; Schreiber, S. L. Natural Products as Probes of Cellular Function: Studies of Immunophilins. *Ang. Chem., Int. Ed. Engl.* **1992**, *31*, 384−400.
(6) Goulet, M. T.; Rupprecht, K. M.; Sinclari, P. J.; Wyvratt, M. J.; Parsons, W. H. The Medicinal Chemistry of FK-506. *Perspect. Drug. Disc. Des.* **1994**, *2*, 145−162.
(7) Debouck, C.; Metcalf, B. W. Human Immunodeficiency Virus Protease: A Target for AIDS Therapy. *Drug. Dev. Res.* **1990**, *21*, 1−17.
(8) Huff, J. R. HIV Protease: A Novel Chemotherapeutic Target for AIDS. *J. Med. Chem.* **1991**, *34*, 2305−2314.
(9) De Clercq, E. Towards Improved Anti-HIV Chemotherapy: Therapeutic Strategies for Intervention with HIV Infections. *J. Med. Chem.* **1995**, *38*, 2491−2517.
(10) Peralman, D. A.; Case, D. A.; Caldwell, J. C.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 4.0. University of California, San Francisco, 1991.
(11) Pearlman, D. A.; Case, D. A.; Caldwell, J. C.; Ross, W. S.; Cheatham, T. E. III; DeBolt, S.; Ferguson, D. M.; Seibel, G. L.; Kollman, P. A. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.
(12) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230−252.
(13) Frisch, M. J.; Trucks, G. W.; Head-Gordon, M.; Gill, P. M. W.; Wong, M. W.; Foresman, J. B.; Johnson, B. G.; Schlegel, H. B.; Robb, M. A.; Replogle, E. S.; Gomperts, R.; Andres, J. L.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Martin, R. L.; Fox, D. J.; Defrees, D. J.; Baker, J.; Stewart, J. J. P.; Pople, J. A. *Gaussian 92*; Gaussian Inc.: Pittsburgh, PA, 1992.
(14) Van Duyne, G. D.; Standaert, R. F.; Karplus, P. A.; Schreiber, S. L.; Clardy, J. Atomic Structure of FKBP-FK506, an Immunophilin-Immunosuppressant Complex. *Science* **1991**, *252*, 839−842.
(15) Wlodawer, A.; Erickson, J. W. Structure-Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543−585.
(16) Loncharich, R. J.; Brooks, B. R. The effects of truncating long-range forces on protein dynamics. *Proteins Struct. Funct. Gen.* **1989**, *6*, 32−45.
(17) Biosym Technologies, InsightII, version 2.3, San Diego, CA, 1995.
(18) Molecular Simulations Inc., Quanta, version 4.1, Burlington, MA, 1995.
(19) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate- based hydroxyethylamine inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805−8809.
(20) Caflisch, A.; Miranker, A.; Karplus, M. Multiple Copy Simultaneous Search and Construction of Ligands in Binding Sites: Application to Inhibitors of HIV-1 Aspartic Proteinase. *J. Med. Chem.* **1993**, *36*, 2142−2167.
(21) Harding, M. W.; Galat, A.; Uehling, D. E.; Schreiber, S. L. A Receptor for the Immunosuppressant FK506 is a Cis-Trans Peptidyl-Prolyl Isomerase. *Nature* **1989**, *341*, 758−760.
(22) Orozco, W.; Tirado-Rives, J.; Jorgensen, W. L. Mechanism for the rotomase activity of FK506 binding protein from molecular dynamics simulations. *Biochemistry* **1993**, *32*, 12864−12874.
(23) Lauri, G.; Bartlett, P. A. CAVEAT: A program to facilitate the design of organic molecules. *J. Comput.−Aided Mol. Des.* **1994**, *8*, 51−66.
(24) Eisen, W. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecular binding site. *Proteins: Struct., Funct., Genet.* **1994**, *19*, 199−221.
(25) Miranker, A.; Karplus, M. Functionality Maps of Binding Sites: A Multiple Copy Simultaneous Search Method. *Proteins: Struct., Funct., Genet.* **1991**, *11*, 29−34.
(26) Gehlhaar, D. K.; Moerder, K. E.; Zichi, D.; Sherman, C. J.; Ogden, R. C.; Freer, S. T. De novo design of enzyme inhibitors by Monte Carlo ligand generation. *J. Med. Chem.* **1995**, *38*, 466−472.
(27) Hart, T. N.; Read, R. J. A multiple-start Monte Carlo docking method. *Proteins: Struct., Funct., Gen.* **1992**, *13*, 206−222.
(28) Burkert, U.; Allinger, N. L. *Molecular Mechanics*; ACS Monographs No. 177; American Chemical Society: Washington, DC, 1982.
(29) Rotstein, S. H.; Murcko, M. A. Groupbuild--A fragment-based method for de novo drug design. *J. Med. Chem.* **1993**, *36*, 1700−1710.

JM950792L