ConCept: de Novo Design of Synthetic Receptors for Targeted Ligands

Wei Chen and Michael K. Gilson*

Center for Advanced Research in Biotechnology, University of Maryland Biotechnology Institute, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received June 12, 2006

Low-molecular-weight receptors that bind targeted guest molecules have a wide range of potential applications but are difficult to design. This paper describes an evolutionary method for computer-aided design of such receptors that works by linking together chemical components from a user-defined library around a stable conformation of the targeted ligand. The software can operate in three modes: de novo design, in which it builds a wide variety of receptors from small components; macrocycle design, in which it builds homopolymeric macrocycles around the ligand; and elaboration of an existing receptor structure. The top candidates generated by the automatic construction process are further studied with detailed affinity calculations whose validity is supported by prior studies of experimentally characterized host−guest systems. All three modes of operation are illustrated here through the design of novel adenine receptors.

1. INTRODUCTION

Chemical hosts are small receptor molecules that possess a cleft or cavity which enables them to bind another molecule. These small receptors have potential advantages over biomolecular receptors because they can be more chemically and physically stable and allow use of a wider range of chemistries. In addition, their lower molecular weight can make them more efficient in terms of cost and delivery. Small receptors have a number of potential chemical applications in areas such as chemical sensing, separations, and catalysis, and important biomedical applications can also be envisioned. Indeed, cyclodextrins are currently employed to improve the bioavailability and stability of some medications,[1−3] and chemical hosts also have potential for broader use in pharmaceutical formulations and even as pharmaceuticals in their own right, as illustrated by the example of a modified cyclodextrin that scavenges the paralytic agent rocuronium.[4]

However, the development of useful synthetic receptors has not been rapid, perhaps in part because it has been difficult to achieve high affinity, as recently noted.[5] Molecular modeling techniques should be useful for this purpose, but comparatively little effort has been invested in the development of computational tools for receptor design. The situation is in striking contrast with the intense focus on software for designing druglike ligands of targeted proteins, as recently reviewed.[6] The HostDesigner software of Hay and Firman[7,8] appears to be the only integrated software for computer-aided, de novo receptor design, and it appears to be applicable primarily to small metal-containing or anionic ligands with well-defined first-shell geometric preferences. As a consequence, it may not be entirely suitable for the design of new receptors for the larger, flexible ligands that are often of interest in chemical and biomedical applications.

Recently, the second-generation Mining Minima algorithm (M2) has yielded promisingly accurate binding affinities for a range of host−guest systems in both organic[9] and aqueous[10] solvent. It is thus of considerable interest to explore its use as part of an approach to the design of targeted synthetic hosts. The present paper describes a novel evolutionary method for automated de novo design of small targeted receptors, which uses the M2 algorithm as a final step in evaluating the leading candidate designs. The design software, termed ConCept (for Construct reCeptor), provides for fully de novo chemical design, controlled elaboration of an existing receptor framework, or automated construction and evaluation of macrocycles. The present paper describes the algorithm and an illustrative application to the design of aqueous adenine receptors.

2. METHODS

The ConCept design methodology comprises two main steps. First, a simple energy model is used to guide an evolutionary process in which a large number of candidate receptors are automatically constructed around a stable conformation of the targeted ligand. Second, the M2 method of computing affinities is used to evaluate the more promising designed receptors. This section details only the automatic construction procedures, because the M2 method has been described elsewhere.[11−14]

2.1. Automated Receptor Construction. ConCept provides three basic procedures for constructing a receptor. (1) Small, varied chemical components can be used to create a range of completely novel receptor designs, which are built around a low-energy conformation of the targeted ligand. (2) The targeted ligand can be docked to an existing receptor type, such as β-cyclodextrin or molecular tweezers,[15,16] and the receptor can then be extended and elaborated through the addition of small, varied chemical components. (3) Macrocycles can be generated by choosing a single chemical component, modifying it chemically, and linking multiple instances of it to form a homopolymeric macrocycle. The resulting macrocyles can then be used as receptor templates and subjected to further elaboration. All three procedures

* Corresponding author phone: (240)314-6217; fax: (240)314-6255; e-mail: gilson@umbi.umd.edu.

10.1021/ci600233v CCC: $37.00
Published on Web 02/22/2007

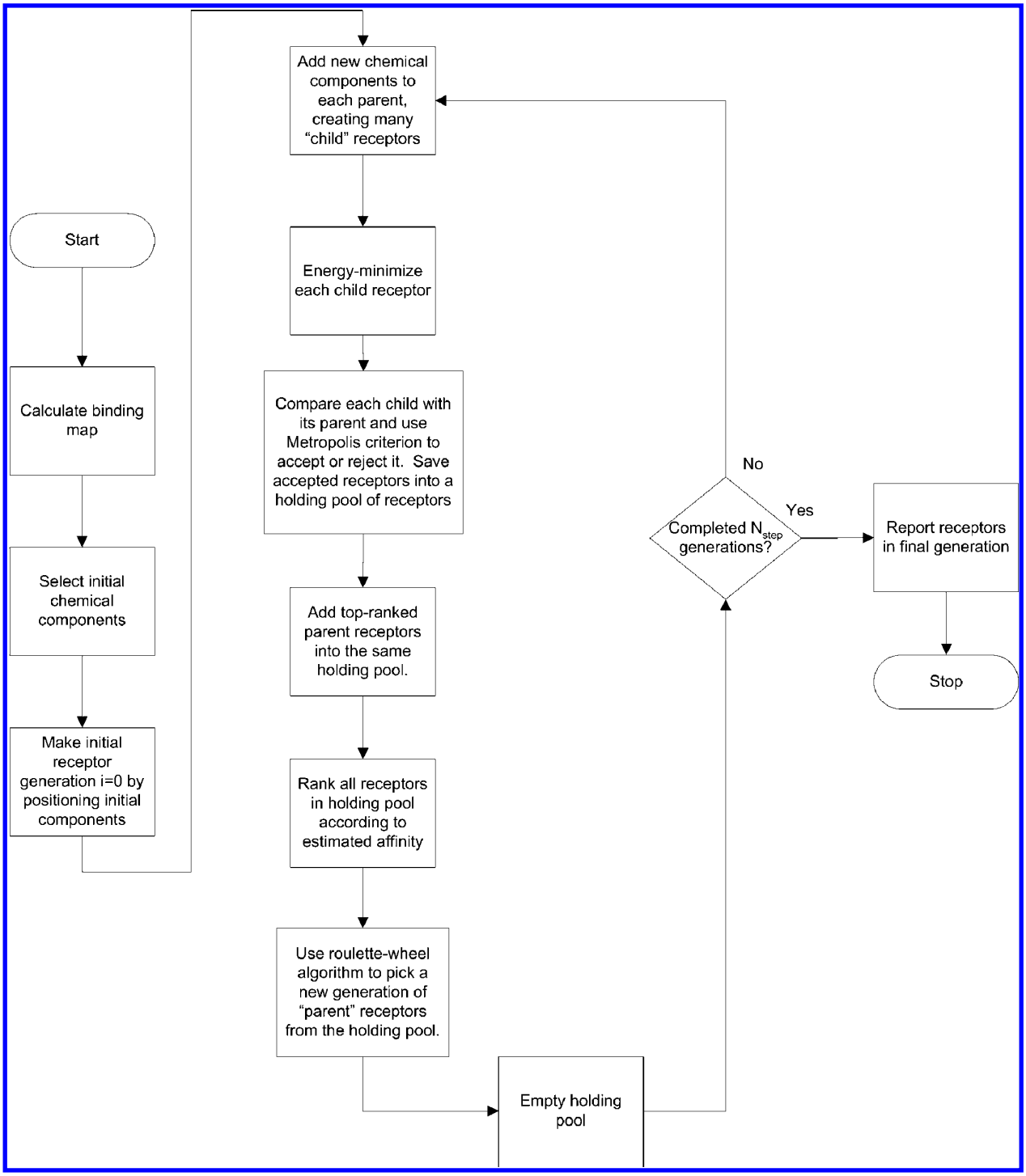

Figure 1. Simplified flow chart of ConCept program for automated design of synthetic receptors.

are iterative and evolutionary. Thus, a population of receptor designs evolves to high affinity through an iterative process of chemical elaboration, evaluation, and selection, which continues for a user-specified number of iterations, N_{step}. The ligand-binding affinities of receptors in the final generation of designs are evaluated with the detailed M2 method,[11−14] which accounts for changes upon binding in the internal energy, solvation energy, and entropy of both the ligand and the receptor. The flowchart in Figure 1 summarizes these procedures, which require on the order of a day on a current commodity computer to generate 10 generations of receptors. Table 1 summarizes the parameters of the method and their default values, which were chosen to balance diversity and quality of the design against computational time.

The following subsections detail the building blocks of the three receptor construction methods: a simplified energy model, a library of chemical components, a grid representation of binding sites around the ligand, and computational approaches to forming bonds between chemical components. Next, the application's use of these methods in the three construction procedures is described. The last two subsections describe ConCept's graphical user interface and a set of illustrative design applications.

2.1.1. Energy Model. The design algorithm seeks to construct receptors that form energetically stable complexes with the targeted ligand. The evolutionary procedures require

Table 1. Parameters of the Automated Design Algorithm[a]

parameter	default value	description	parameter	default value	description
D_{phi}	*30*	*dihedral increment in rotamer sampling*	N_{top1}	5	no. of top poses kept in first step of initial component placement
E_{cutoff}	500.	upper limit of accepted rotamer energies	N_{top2}	5	no. of top poses kept in second step of initial component placement
E_d	*2*	*factor converting distance to a pseudoenergy*	P_{cutoff}	*15*	*cutoff percentile used in generating interaction centers*
f_{elite}	0.20	fraction of parents treated as elite	R_{bump}	*0.6*	*atoms i,j bump if* $r_{ij} < R_{vdw,i} + R_{vdw,j} - R_{bump}$
N_{step}	10	number of generations	$R_{overlap}$	*0.5*	*atoms overlap if* $r_{12} < R_{overlap}$
N_{pop}	200	population size of each generation	R_{angle}	*30*	*an angle is considered reasonable if its value is in the range of* $109° \pm R_{angle}$
N_{child}	3000	maximum no. of children per parent placed in holding pool	$R_{cluster}$	3.0	radius used in counting neighbors of interaction points
N_{link}	*3*	*no. of top linkage points used during receptor modificiation*	$R_{spacing}$	0.5	grid spacing
N_{pose1}	3000	no. of poses used in first step of initial component placement	S_{mut}	on	switch for single-atom mutation
N_{pose2}	1000	no. of poses used in second step of initial component placement	S_{second}	on	switch for formation of secondary links

[a] All except those in italics are adjustable by the user. (Distances in Å, energies in kcal/mol, and angles in degrees.) D_{phi}: Increasing this dihedral step size above the default can markedly reduce CPU time, but some rotamers may be missed. Decreasing its value does not necessarily improve the results. E_d: Functions as a distance tolerance when forming links. A smaller value would allow more structures to pass the distance-based Metropolis Monte Carlo test and might thus reduce the quality of the designed receptors. f_{elite}: Increasing the fraction of elite receptors passed unchanged to the next generation above the default value does not necessarily improve the results, but good candidates will be missed if this value is too small. N_{step}: Increasing the number of generations tends to generate larger receptors and requires more CPU time. N_{pop}: Increasing N_{pop} broadens the number of receptors tested in each generation and has the potential to increase the diversity of the final results, but it also increases the CPU time per generation. N_{child}: The ratio N_{child}/N_{pop} determines the maximum number of children each parent structure can add to the next generation. For a fixed N_{pop}, increasing N_{child} may reduce the diversity of the final results, because the children of some parents may dominate the rankings. N_{link}: Building onto more linkage points for every parent receptor increases CPU time and tends not to improve results because the additional children receptors tend to fail the screening tests for assembling the next generation. Reducing this parameter tends to reduce diversity and make it harder to join unconnected fragments in the Link procedure. N_{pose1}, N_{pose2}: Increasing the values of these two parameters may yield somewhat better-docked initial components but at the cost of somewhat more CPU time. N_{top1}: Significantly affects the CPU time required for fine tuning the placement of initial components, and increasing its value above the default does not necessarily generate better placements. N_{top2}: Determines the number of parents used in the *initial* step of the receptor design procedure. Increasing its value has the potential to increase the diversity of receptors without significantly extending CPU time. P_{cutoff}: Final results are not very sensitive to this parameter, although changing it may change the positions of the interaction centers somewhat. R_{bump}, $R_{overlap}$, R_{angle}: Larger values of these distance and angle tolerances favor the generation of secondary links but may lead to designs with excessive internal strain. $R_{cluster}$: Smaller values result in more interaction centers. $R_{spacing}$: Smaller values of this parameter result in a better binding map but increase CPU time significantly in the map-generating step. On the other hand, if the value is too large, the quality of the binding map may be reduced. S_{mut}, S_{second}: Turning off these two switches in the Grow and Link procedures decreases the diversity of receptors without significantly decreasing CPU time.

many energy evaluations, so a rapid approximation is used in which the energy is estimated as a sum of energy terms:

$$E \approx E_{val} + E_{vdW} + E_{Coul} + E_{GB} + E_{NP} \quad (1)$$

These represent, respectively, the valence energy (bond stretches, angle bends, and dihedral rotations); van der Waals interactions; Coulombic interactions; electrostatic solvation, estimated with a generalized Born model;[17,18] and nonpolar solvation, estimated as proportional to surface area.[10,19] All terms are computed with a full atomic representation of the receptor and the ligand, with bonded and van der Waals parameters drawn from the CHARMM force field, and VC/2004 charges computed with the program Vcharge.[20] Generalized Born cavity radii and the nonpolar solvation coefficient are as previously described.[5,10] The change upon binding is computed with the approximation that both the ligand and the receptor do not change conformation when separated.

2.1.2. Library of Chemical Components. The receptor structures are assembled from prepared libraries of chemical components. These are classified into four types: small, varied organic compounds; receptor templates, such as cyclodextrins; natural and non-natural amino acids, to enable peptide design; and special monomers for the construction of homopolymeric macrocycles. The existing libraries currently contain 90, 12, 59, and 14 components of these respective types, and new components can readily be added to each library. ConCept allows the user to choose which libraries and which component(s) within each library will be available during a given run. Each chemical component is represented as a complete molecule in a low-energy conformation and is associated with its CHARMM force field parameters and a list of its hydrogen bond donors and acceptors, if any.

2.1.3. Ligand Binding Map. A grid of van der Waals and Coulombic potentials around the ligand is used to guide several steps of the design process. A grid with 0.5 Å spacing is centered on the ligand, and any grid point that is outside the steric volume of the ligand is marked as vacant. A grid point i is considered to be outside the ligand if $d_{ij} > R_j + R_H - 0.5$ Å for all N_{lig} ligand atoms j, where d_{ij} is the distance from grid point i to atom j, R_j is the van der Waals radius of atom j, drawn from the CHARMM force field,[21] and R_H is the van der Waals radius of a nonpolar hydrogen atom in the CHARMM force field, 1.33 Å.

All N_{grid} vacant points are then classified as nonpolar interaction points, hydrogen bond interaction points, or noninteracting points, by the following procedure. Each point is probed with a positively charged sp3 nitrogen atom (ammonium cation), representing a hydrogen bond donor; a

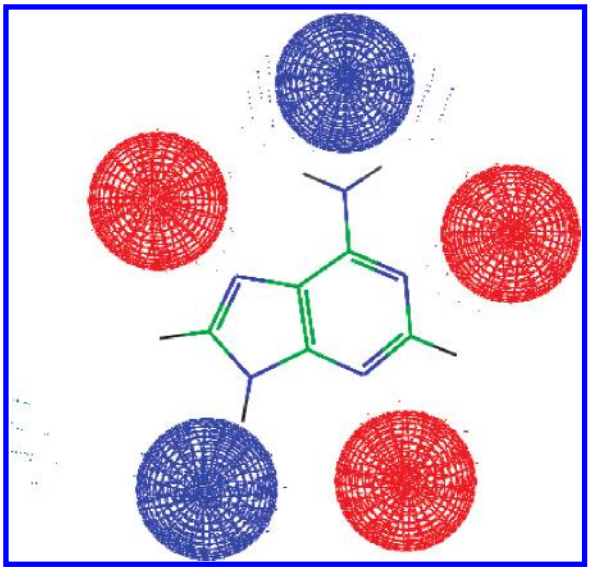

Figure 2. Interaction points (dots) and interaction centers (spheres) computed for adenine. Blue: sites suitable for hydrogen bond acceptors. Red: sites suitable for hydrogen bond donors. Green: sites suitable for nonpolar groups.

negatively charged sp2 oxygen atom (as in a carboxyl group), representing a hydrogen bond acceptor; and an sp3 carbon atom (as in methane), representing a nonpolar group. The nonpolar potential energy of point i is approximated with a Lennard-Jones function $E_i^{NP} = \sum_{j=1}^{N_{lig}} ([A/r_{ij}^{12}] - [B/r_{ij}^{6}])$, where j indexes the N_{lig} ligand atoms and the parameters are those used in the GRID method.[22] Hydrogen-bonding energies are computed as $E_i^{HB} = \sum_j([C/r_{ij}] - [D/r_{ij}])\cos^4\theta$ where r_{ij} is the distance from grid point i to a hydrogen-bond donor or acceptor j and C_j and D_j are again drawn from GRID.[22] All ligand atoms of the relevant type (donor or acceptor) contribute to the respective interaction potential at each grid point. For the nitrogen atom probe, θ is the angle defined by the nitrogen probe, the hydrogen bond acceptor atom of the ligand, and the parent atom of the hydrogen bond acceptor. For the oxygen probe, θ is the angle defined by the oxygen, the polar hydrogen, and the parent atom of the polar hydrogen. This procedure yields a value of E_i^{NP}, a value of $E_i^{HB\ donor}$, and a value of $E_i^{HB\ acceptor}$ at each grid point i. The point is classified as either a potential nonpolar interaction point or a potential hydrogen bonding point, depending upon which energy is more negative, and the associated energy E_i is assigned to the grid point.

The average grid energy is then computed as $\langle E \rangle \equiv (1/N_{grid})\sum_i^{N_{grid}} E_i$, and each grid point whose energy is above $\langle E \rangle$ is marked as a candidate *interaction point* of the respective type. Each candidate interaction point is checked for the number of its neighbors, where a neighbor is another candidate interaction point of the same type within a distance of 3 Å. Points with below-average neighbor counts are discarded; those which remain are classified as interaction points. In addition, candidate interaction points with neighbor counts in the top 15th percentile are clustered, based upon a distance cutoff of 3 Å, and an interaction center of the appropriate type (donor, acceptor, or hydrophobic) is placed at the mean coordinates of each of the resulting clusters, as illustrated in Figure 2.

2.1.4. Chemical Modification: Linking and Mutating. The process of receptor design comprises initial positioning of one or more chemical components around the ligand, followed by multiple iterations of chemical modification and selection. This subsection describes the process of chemical modification.

Each chemical component in the library is tagged with a list of linkage points which are available for connecting two components together, and each type of linkage point is associated with a leaving group and a linkage rule. In the simplest case, the linkage points and leaving groups are hydrogen atoms and the linkage rule is elimination of two hydrogens and formation of a single bond (see Figure 3). More complex rules also are available, such as formation of an amide bond between an amine and a carboxylic acid, with hydrogen and hydroxyl leaving groups. All current rules lead to formation of a single bond, but this can readily be generalized. Every chemical modification is immediately followed by reassignment of atom types and energy parameters (section 2.1.1).

When a new rotatable bond is formed by application of a linkage rule, the newly added component is rotated around the bond in D_{phi} increments to produce 12 different rotamers, and each rotamer is checked for steric clashes. All rotamers whose energies are within E_{cutoff} of the most stable rotamer are kept. These energies typically fall drastically during the energy minimization which is applied to all modified structures (below).

In addition, rotamers that generate steric overlaps can optionally be evaluated for formation of secondary links. A secondary link leads to ring-formation, except when the secondary link is formed between two distinct chemical fragments, as in the Link algorithm (section 2.1.5). Three types of secondary link can be formed. First, when two hydrogen atoms bump but do not overlap (see criteria in Table 1), the colliding hydrogens are deleted and a methylene is added to form a bridge. Second, when two non-hydrogen atoms, each bonded to at least one hydrogen atom, bump each other but do not overlap, an applicable linkage rule is applied to form a new bond, with removal of the appropriate leaving groups. Last, when two non-hydrogen atoms overlap, then one is deleted, its bonds are reconnected to the other, and hydrogen atoms are added or deleted to satisfy valence requirements. In order to avoid generating unstable compounds, the algorithm rejects the resulting receptor design if the new link matches any of a list of undesirable chemistries (Figure 4), or if the newly formed bond angles are unreasonably strained. (See R_{angle} in Table 1.) When different rotamers of an added component produce different secondary links, each of the resulting receptors is kept.

Single-atom mutations allow for further optimization of a receptor immediately after a component has been added by the linking operations described above; this procedure can be turned on or off during a given design run, depending upon whether the user is willing to allow the use of chemistries not present in the library of chemical components. Mutation is based upon a classification of each non-hydrogen atom of the receptor as a hydrophobe, hydrogen-bond donor, or hydrogen-bond acceptor, based upon their CHARMM atom types. Every non-hydrogen atom that lies within 1 grid unit of an interaction point (section 2.1.3) is checked for congruity with the type of the grid site; for example, a donor atom should lie near a donor point. An incongruous atom is mutated to a more suitable

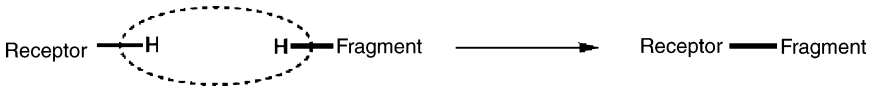

Figure 3. Simplest rule for forming a chemical link between two chemical components.

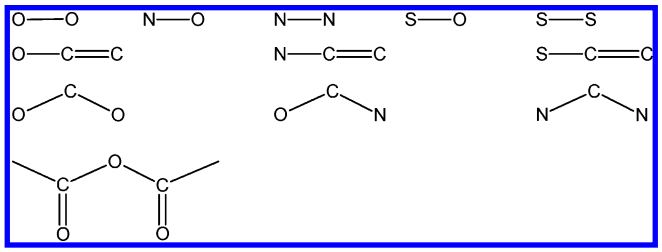

Figure 4. Diagrams of undesirable chemistries that trigger rejection of a new candidate receptor.

atom type, and hydrogen atoms are added or deleted as needed to satisfy valence. Both the mutated and unmutated forms of the receptor are kept.

After addition of a chemical component, creation of any secondary links, and potentially mutation of some atoms, the resulting modified receptors are relaxed by 50 steps of conjugate gradient minimization of the energy (eq 1), with the target ligand held fixed.

2.1.5. Evolution of Receptors. In all three design procedures, receptors evolve through a series of generations, each typically comprising $N_{pop} \approx 200$ candidate receptors, where N_{pop} can be specified by the user at the outset of the procedure. The top $f_{elite} \approx 0.2$ of the receptors in a given generation (elites) are placed in a holding pool of candidates for the next generation. The entire parent generation of receptors is then chemically modified to form a large set of child receptors. Child receptors are compared with their parents, and the best N_{child}/N_{pop} children per parent are added to the holding pool if they pass a Metropolis Monte Carlo accept/reject step. All receptors in the holding pool are then ranked, and roulette wheel selection,[23] with weighting by rank, is used to pick the new generation of N_{pop} receptors. The two stochastic steps allow the design algorithm to escape from local optima in which it might become trapped if a simple greedy optimizer were employed. These procedures are now described in greater detail.

Evolutionary Construction from Small Chemical Components: The Grow and Link Methods. The Grow method (Figure 5, middle) starts with a first generation of N_{top2} (default 5) receptors in which each receptor consists of a single chemical component positioned at one initial interaction center around the ligand (Figure 2), each placement constituting a parent receptor. New components are then linked to the growing receptors with the goal of creating receptors that position suitable additional chemical groups at each other interaction center around the ligand. The Link method (Figure 5, right) starts with a first generation of N_{top2} receptors in which each receptor has a chemical component at each interaction center around the ligand. New components are then attached to connect the initial components into a single receptor molecule.

Both Grow and Link begin by placing one or more user-selected chemical components, drawn from the library, at complementary interaction centers around the ligand. The components are positioned by generating $N_{pose1} \approx 3000$ different poses (position and orientation) of the component at its complementary interaction center, choosing the $N_{top1} \approx 5$ most stable poses according to eq 1, and then fine-tuning these placements by testing another $N_{pose2} \approx 1000$ poses in a $1 \times 1 \times 1$ Å cube centered on each of the 5 poses. The most stable $N_{top2} \approx 5$ poses at each interaction center are kept. The Grow method then builds each parent receptor toward unoccupied interaction centers by computing the distance between each linkage point of the receptor and each unoccupied interaction center and adding new chemical components to the $N_{link} = 3$ linkage points with the shortest distances, according to the methods of section 2.1.4. The Link method builds a single connected receptor from disconnected initial fragments by adding new chemical components to the $N_{link} = 3$ linkage points that are closest to linkage points on the other fragments of the parent receptor.

In both methods, child receptors are formed by adding every available chemical component to each of the selected linkage points on each parent receptor in the current generation. Thus, if 3 linkage points are used and there are N components in the library, then each parent will generate at least $3N$ child receptors, each larger than its respective parent receptor by the addition of one component. The resulting child receptors are then subjected to an extra pruning step based upon how well the added component reaches toward the targeted site or fragment. In the Grow method, this step is based upon the mean distance $\bar{d}$ between any linkage point of the new component and the targeted interaction center. In the Link method, this step is based upon the mean distance between any linkage point of the added component and the nearest linkage point of the targeted fragment. A pseudoenergy is computed as

$$E = E_d(\bar{d} - d_{min})$$

where the value of $E_d = 2$ kcal/mol/Å and d_{min} is the shortest distance in the parent, and the Metropolis sampling criterion with $T = 300$ K is used to accept or reject the child.

Free and Steered Elaboration of an Existing Receptor. ConCept can also be used to elaborate an existing receptor. For example, one might wish to modify β-cyclodextrin to maximize affinity for a drug as part of a formulation project. A simple rigid docking capability allows the targeted ligand to be fitted into the initial receptor, and the receptor is then used as an initial chemical component in either the Grow or Link method, described above. Elaboration of an existing receptor can also be steered by selection of a small set of linkage points on the initial receptor as well as a limited set of chemical components. ConCept was used previously in this way to engineer a cross-link between two user-selected linkage points on a designed receptor.[5]

Evolutionary Construction of Macrocyclic Receptors: The Cyclize Method. ConCept allows evolutionary construction of macrocyclic homopolymer receptors for the targeted ligand. The Cyclize method is associated with a library of suitable monomers; i.e., compounds with at least two linkage points that allow cyclization by a single linkage rule. For

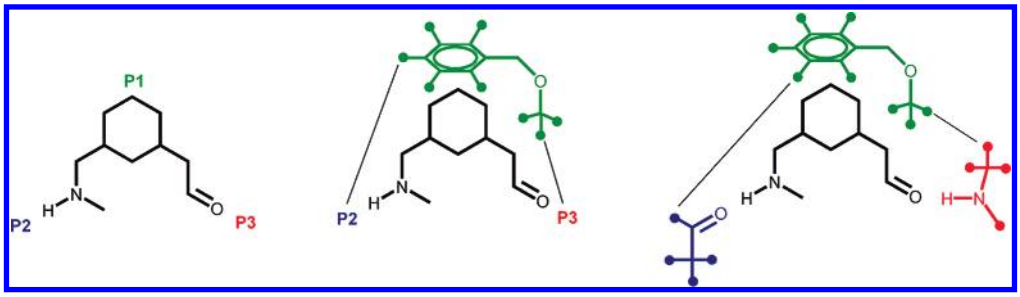

Figure 5. Diagram of Grow and Link algorithms. Left: targeted ligand (black) with 3 interaction centers labeled P1, P2, and P3 and colored according to the convention in Figure 2. Center: Grow algorithm, showing an initial aromatic component at center P1. Eight available linkage points are marked with dots, and thin black lines indicate the targeting of interaction centers P2 and P3 by the closest linkage points. Right: Link algorithm, showing initial components at centers P1, P2, and P3. Sixteen available linkage points are marked with dots, and pairs of closest points are indicated by thin black lines.

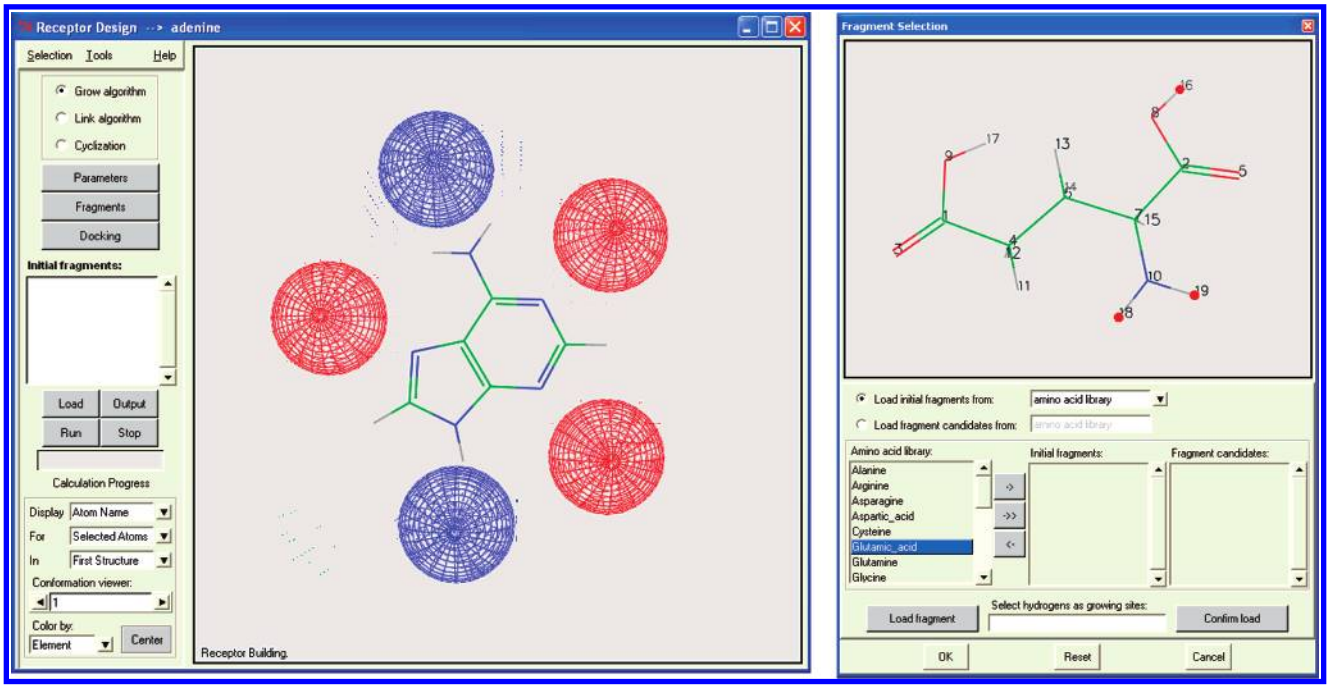

Figure 6. Two windows from the ConCept user-interface. Left: Main window displaying adenine and its binding map. Users can select the Grow, Link, or Cyclize algorithm at the top of the control panel. The Parameter button allows computational parameters to be set. Initial chemical components and component libraries are selected by clicking the Fragments button and then using the Fragment Selection window (right). A text window displays the names of the selected initial fragments. The docking of initial fragments to the target ligand is initiated with the Docking button. Graphical controls are visible at the lower left of the window, and tools for superimposing structures, calculating atom−atom RMSD, etc., are available in the Tools menu (top). Right: Library construction and component selection window. When a library is selected, the names of all of its compounds are shown on the left-most text window; the structure of a compound is displayed when its name is highlighted. The Load button and its associated controls allow the user to load additional components into the current library.

example, amino acids can be used as Cyclize monomers with the amide-formation linkage rule. The method begins with a single user-selected monomer, which is formed into a set of regular polygon homopolymers of various lengths. The ligand is docked to the center of the macrocycle, and the macrocycle is energy-minimized to form the first parent generation. Elite parents are added to the holding pool (above) for the next generation, and new receptors are then formed. This is done by modifying a parent monomer by linking new chemical components to any of N_{link} linkage points other than the linkage points required for the cyclization reaction. The formation of secondary chemical links usually is turned off in order to avoid creating excessively complex ring systems. The modified monomers are then formed into regular polygon homopolymers of various lengths, the ligand is docked to the resulting macrocycles, and the macrocycles are energy-minimized, to form a set of child receptor designs, which are accepted into the holding pool if they pass a Metropolis Monte Carlo accept/reject step based upon their energies relative to the energies of their parents. Roulette wheel selection is applied to the receptors in the holding pool to produce a new generation of receptors. The monomers are then subjected to further rounds of chemical modification and cyclization, and the process iterates until a user-specified number of generations has been completed.

2.2. Graphical User Interface. ConCept is outfitted with a graphical user interface to facilitate setup, execution, and interpretation of the design calculations. Figure 6 provides a sense for the layout by showing the main interface window and a scaled down view of the window used to choose the subset of available chemical components for receptor construction. The interface allows most of the parameters of the algorithm to be changed from their default values (Table 1).

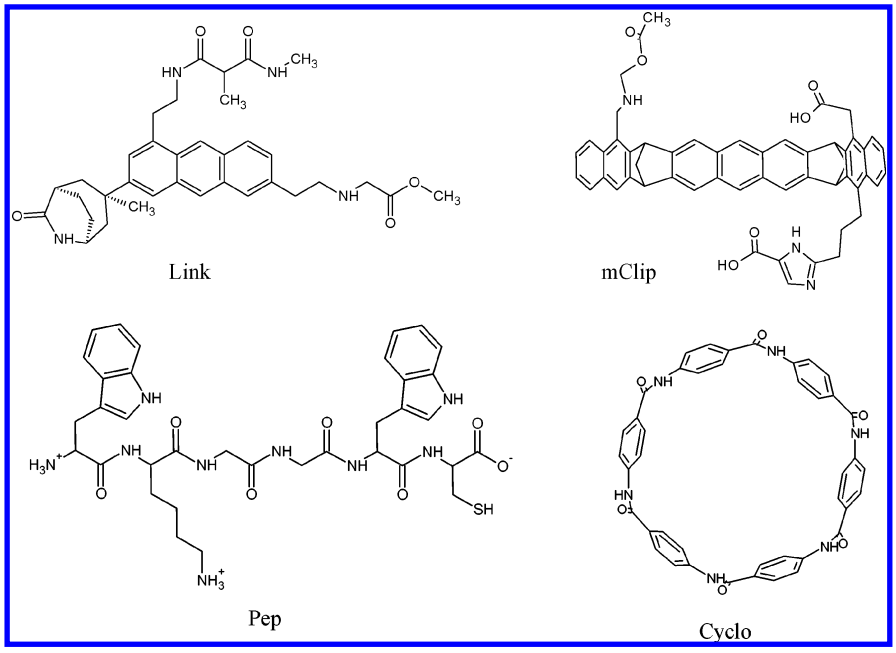

Figure 7. Chemical structures of the designed adenine receptors.

2.3. Illustrative Application: Design of Aqueous Adenine Receptors. The methods provided by ConCept are illustrated here in an application to the design of aqueous adenine receptors. A stable conformation of adenine obtained from a brief conformational search was used as a starting point. Except as otherwise noted below, all parameters of the algorithm were set to their default values (Table 1). Three separate design runs were executed for the Grow and Link methods, two were executed for the Cyclize method, and one was executed for the Receptor Elaboration method. In each case, the highest-ranked 2−3 receptors from each design run, based upon the change in E in eq 1, were further assessed with the M2 algorithm. Thus, M2 runs were carried out for about 15 designed receptors in all. The chemistries used for each design method are as follows:

Grow. Peptides were constructed with glycine as the initial component and a chemical library comprising all 20 biological L-α-amino acids. The design with the highest computed affinity, **Pep**, is described in the Results.

Link. Formamide, acetamide and urea were used as initial components, and the full library comprised imidazole, carbazole, *N*-methylacetamide, *N*-methylformamide, acetaldehyde, acetamide, acetone, acetic acid, ammonia, anthracene, benzene, bromine, chlorine, cyclohexane, cyclopentane, dimethylamine, dimethylether, ethane, fluorine, formic acid, formaldehyde, iodine, methane, methanol, methylaetate, methyl formate, methylamine, naphthalene, o-phthalimide, phosphoric acid, propane, pyridine, sulfonic acid, urea, and water. The best resulting receptor is termed **Link**.

Cyclize. The initial monomers were glycine, toluene, 4-(1-methyl-1-phenylethyl)phenol, calixarene, ethane, ethanol, (2*E*)-but-2-ene, methanol, 4-aminobenzoic acid, 4-hydroxybenzoic acid, 3-hydroxybenzoic acid, α-glucose, and propan-1-ol. In order to avoid excessively complex macrocycles, switches $S_{mutation}$ and S_{second} were set to disable single atom mutation and formation of secondary links, and compounds with rings were excluded from the library of components used to modify the monomers. This library comprised *N*-methylacetamide, *N*-methylformamide, acetaldehyde, acetamide, acetone, acetic acid, ammonia, benzene, bromine, chlorine, ethane, fluorine, formic acid, formaldehyde, iodine, methane, methanol, methylacetate, methyl formate, methylamine, phosphoric acid, propane, pyridine, sulfonic acid, urea, and water. The Cyclize calculations yielded the novel macrocyle **Cyclo**, discussed below. The algorithm also happened to generate α- and β-cyclodextrin as potential receptors for adenine, but these well-known receptors are not discussed further.

Receptor Elaboration. A molecular clip[24] was elaborated by the Grow algorithm, using the same chemical library as for the Cyclize runs, above. The S_{second} switch was turned off to avoid excessively complex designs. The best design, termed **mClip**, is presented.

3. RESULTS

Figure 2 illustrates the interaction points (dots) and interaction centers (spheres) computed for adenine and used for the design of new receptors: blue and red indicate points suitable for hydrogen bond acceptors and donors, and green indicates points suitable for nonpolar groups. No nonpolar interaction centers are present because adenine's atoms are significantly polar. Nonetheless, the design algorithm does position nonpolar and/or aromatic groups above and below the adenine plane, as detailed below.

3.1. Grow Method: Receptor Pep. According to the initial design, receptor **Pep** (WKGGWC) sandwiches the adenine between two indole groups, while a salt-bridge between the two chain termini holds the rest of the peptide in what amounts to a cyclic conformation. Polar groups from the peptide furthermore form hydrogen bonds with one edge of the adenine molecule. Full M2 calculations, which include extensive conformational analysis, predict that the most stable conformation of the complex (Figure 8) is similar to that in the design: in particular, the two indole groups are positioned much as in the complex and are thus preorganized to bind adenine. The lysine and cysteine side chains do not interact closely with the adenine guest, so this design could probably

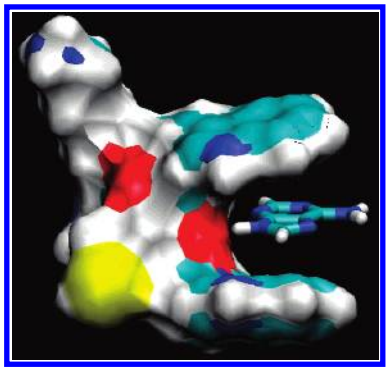

Figure 8. Most stable conformation found in M2 calculations for receptor **Pep** with adenine.

Table 2. Calculated Changes in Boltzmann-Averaged Energy Components, Configurational Entropy, and Standard Free Energy (All in kcal/mol) for Binding of Adenine with Designed Receptors[a]

	Pep	Link	Cyclo	mClip
$\Delta\langle U_{VDW}\rangle$	−10.2	−8.3	−11.5	−11.5
$\Delta\langle U_{Coul}\rangle$	−8.3	−11.6	−8.8	−14.7
$\Delta\langle W_{elec}\rangle$	7.1	10.0	9.8	13.6
$\Delta\langle U_{Coul} + W_{elec}\rangle$	−1.2	−1.6	1.0	−1.2
$\Delta\langle W_{np}\rangle$	−1.8	−1.5	−2.1	−1.9
$\Delta\langle U_{val}\rangle$	−0.3	−1.7	0.9	0.4
$-T\Delta S^{o}_{config}$	11.1	11.2	9.3	11.7
ΔG^{o}_{bind}	−2.4	−1.8	−2.4	−2.5

[a] U_{VDW}: van der Waals energy. U_{Coul}: Coulombic energy. W_{elec}: electrostatic solvation energy. W_{np}: nonpolar solvation energy. U_{val}: sum of bond, angle and torsional energies. $-T\Delta S^{o}_{config}$: free energy contribution from change in configurational entropy. ΔG^{o}_{bind}: computed free energy of binding.

be simplified without loss of affinity by using simpler side chains, as previously done to arrive at a minimal RGD receptor.[5]

The predicted free energy of binding for the **Pep** receptor is −2.4 kcal/mol (Table 2), and the largest favorable energy component is the van der Waals term, not surprisingly given the importance of stacking interactions in the complex. The favorable Coulombic term is largely cancelled by electrostatic desolvation, leaving only −1.2 kcal/mol in net electrostatic energy. As observed previously,[5,9,10] a large loss in configurational entropy, here about 11 kcal/mol, strongly compensates the terms that drive binding.

3.2. Link Method: Receptor Link. According to the initial design, receptor **Link** provides an anthracene group as an aromatic stacking surface for adenine, along with three branches that extend upward to form lateral hydrogen bonds with adenine. This arrangement is similar to that used previously by Rebek and co-workers.[25,26] Importantly, the anthracene moiety, which is employed in the highest affinity Rebek design, was not a user-selected started component here but emerged from the simulated evolution procedure. The most stable conformation predicted by the M2 calculations (Figure 9) deviates somewhat from the design, because the lactam branch turns downward, rather than forming hydrogen bonds with adenine. However, the other two branches form the intended interactions. The predicted affinity is −1.8 kcal/mol (Table 2). Interestingly, the valence term, $\Delta\langle U_{val}\rangle$, which comprises bond-stretch, bond-angle, and torsional energies, is rather favorable, at −1.7 kcal/mol. Examination of the most stable form predicted for the free receptor shows a

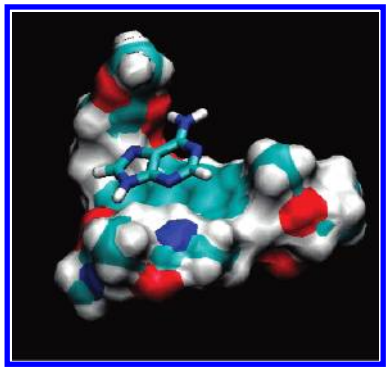

Figure 9. Most stable conformation found in M2 calculations for **Link** with adenine.

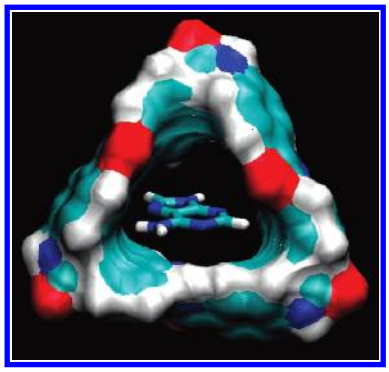

Figure 10. Most stable conformation found in M2 calculations for **Cyclo** with adenine.

number of hydrogen-bonding interactions among the polar branches; it appears that these induce significant valence strain, which is relieved when adenine binds.

3.3. Cyclization Method: Receptor Cyclo. Cyclo is a macrocyclic ring composed of six 4-aminobenzoic acid monomers linked by amide bonds. It happens to be constructed of unmodified initial monomers, because these outperformed the various modified forms tested by the construction software. For example, M2 predicted that a macrocycle composed of six methylbenzenes modified by addition of *para* carboxyl and methoxy groups would bind adenine with an affinity of only about −1 kcal/mol. According to the initial design, the Cyclo macrocycle structure is round, its benzene rings tilt 22 degrees from the central axis, all amide groups are in the Z configuration, and the plane of the bound adenine lies along the axis of the ring. The M2 conformational search discovered a more stable conformation in which the receptor adopts a triangular shape by flipping 3 amides into the E configuration (Figure 10). This is also predicted to be the most stable conformation of the free receptor, so it can be considered preorganized for binding. As for the **Pep** and **Link** receptors, binding appears to be driven primarily by van der Waals interactions, and the affinity is predicted at about −2.4 kcal/mol (Table 2). The net electrostatic energy is unfavorable, due to substantial desolvation of adenine's polar groups without formation of geometrically optimal hydrogen bonds. The loss in configurational entropy is somewhat lower than for the other two receptors, perhaps reflecting the preorganization of the receptor and the relatively roomy binding cavity. This macrocycle may represent a suitable template for further enhancement, such as addition of groups positioned to form tighter hydrogen bonds with adenine.

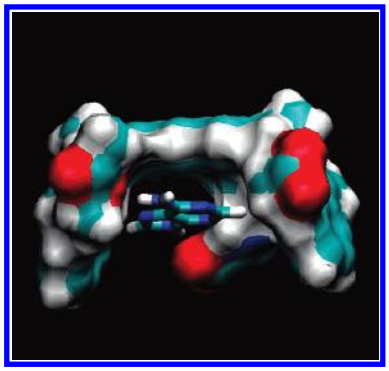

Figure 11. Most stable conformation found in M2 calculations for **mClip** with adenine.

Table 3. Number of Hydrogen Bonds from Each Potential Donor/Acceptor Atom of Adenine (Column Headers) to Each Designed Receptor (First Column), Based upon the Most Stable Predicted Conformation of Each Complex[a]

	N1	C2	N3	N6(cis/trans)	N7	N9
Pep	0	0	1	0/0	0	1
Link	0	0	1	0/1	1	1
Cyclo	0	0	1	1/0	0	0
mClip	0	0	1	0/1	1	1

[a] Geometric criteria for hydrogen bonding as in ref 27.

3.4. Receptor Elaboration Method: Receptor mClip. A molecular clip[24] known to bind small aromatic compounds in organic solvent was selected for elaboration with the Grow algorithm. M2 predicts that the baseline receptor, prior to elaboration, binds adenine with an affinity of +1.2 kcal/mol, suggesting no significant binding in aqueous solution. The modified receptor, **mClip**, allows adenine to stack onto the central anthracene moiety, while polar groups added to the naphthyl arms of the clip provide lateral hydrogen-bonding interactions. The design approach thus is similar to that of **Link**. The added polar groups should also increase the aqueous solubility of the receptor, relative to the highly nonpolar starting clip. M2 analysis yields a very similar global free energy minimum (Figure 11), except that the short carboxylate branch projects into solution, rather than interacting closely with adenine. The predicted binding affinity is −2.5 kcal/mol (Table 2), with balancing energy contributions much as for the receptors discussed above.

3.5. Comparison of Designed Receptors with Adenine-Binding Sites in Proteins. The interactions of the present receptors with adenine resemble those of proteins that bind ligands with adenine moieties such as AMP, ADP, ATP, FAD, NAD, and NADP. A structural analysis of such proteins,[27] representing a variety of folds and families, reveals a tendency for nonpolar groups to be positioned above and below the plane of the adenine moiety, although some proteins employ flat, polar groups such as a histidine imidazole. The present designs are similar in this respect, as they provide flat, nonpolar groups that interact with one or both sides of the adenine plane (Figures 8−11). Table 4 of the structural analysis[27] indicates that about 40% of the proteins analyzed bind adenine without forming hydrogen bonds, while 40% form 1 or 2 hydrogen bonds, and 20% form 3 or 4 hydrogen bonds. (The analysis considers C2 as a hydrogen-bond donor.) Table 3 breaks down the predicted hydrogen bonds of adenine with the designed receptors in the same manner as Table 4 in the protein analysis,[27] using the same geometric criteria as used in the protein analysis. The designed receptors are predicted to form 2−4 hydrogen bonds with adenine, somewhat more than observed for the proteins. The difference can be attributed in part to the unavailability of N9 for hydrogen bonding in the adenine derivatives found in the proteins.

4. DISCUSSION

The present paper demonstrates automated design of small biomimetic receptors by three approaches: de novo design from chemical fragments, formation of homopolymeric macrocycles, and guided elaboration of existing receptors. The adenine receptors generated by the three methods are varied, both structurally and chemically. Their computed affinities lie between −1.8 and −2.5 kcal/mol, the same order as experimentally proven designs of Rebek, which bind adenine with affinities of −0.4 to −2.4 kcal/mol.[25,26] Intriguingly, the new binding sites resemble those of adenine binding proteins, positioning nonpolar groups above and below the adenine ring, while providing lateral hydrogen bonds.

Although we have not assessed the synthetic accessibility of these designs, they appear to be chemically reasonable due, in part, to the use of design rules prohibiting the construction of unsuitable groups, such as peroxides. In general, the synthetic accessibility of the designs generated by ConCept can be maximized by relying on well-characterized component sets and linkage rules, such as amino acids and the peptide linkage. Broader chemistries can be incorporated into peptides by allowing the use of non-natural amino acids, and cyclization can be achieved with a peptide link or with a disulfide bond. The reliable reactions of click chemistry[28,29] may also be valuable in this context. More generally, even if a given design is not synthesizable, it may be possible to devise modifications that will allow a similar compound to be made, as the program CAVEAT[30] was used to guide design of a synthetic glucose receptor.[31]

The ConCept algorithm is similar to the pioneering HostDesigner program[7,8] in constructing receptors via rule-based formation of chemical bonds with chemical components drawn from predefined libraries and in prepositioning chemical components to take advantage of favorable interactions with the ligand. However, whereas HostDesigner is tailored specifically for constructing receptors that bind metal ions[7] or anions[8] via geometrically optimized first-shell interactions, ConCept is designed to take advantage of softer nonbonded interactions with a relatively large polyatomic ligand. In addition, there are significant differences in how the two methods evaluate candidate receptors during and after the construction process.

The use of M2 calculations to evaluate the designs generated by ConCept is an important aspect of the overall design process. The broad conformational search carried out by M2 tests the stability of the bound conformation proposed by the design software, and M2 furthermore provides an estimate of the receptor−ligand binding affinity. Even if the computed binding affinity of a design is not high, the conformational preferences predicted by M2 represent a useful starting point for user-guided elaboration of the receptor. On the other hand, although the validity of the M2 method is supported by the calculation of relatively accurate

affinities for a number of systems in prior studies,[5,9,10] it is not guaranteed to be equally accurate in future applications to receptor design, so trial and error evaluation of ConCept's designs will presumably be required, much as in computer-aided drug design.

One limitation of this first version of ConCept arguably is that the design process occasionally incorporates groups that are predicted to project into solution, like the lysine and cysteine side chains of the **Pep** receptor. One approach to strengthening this aspect of the program would be to adjust the figure of merit used in the evolutionary algorithm so that it would penalize receptors by size or by the number of atoms lacking strong interactions with the ligand. An alternative approach would be to add a specific "pruning" step to the algorithm. Another direction for future development may be to improve the treatment of energy and conformation during the design process. Currently, the evolutionary part of the algorithm relies upon energy minimization with a simplified energy model, while the final evalution with M2 is more time-consuming. It is worth investigating whether the yield of high quality receptors for a given investment of computer time would be improved by incorporating a more detailed treatment during the design process.

It seems reasonable to expect that ConCept and other computational methods for constructing and screening targeted receptors will facilitate the discovery of useful molecules. The breadth of chemical exploration and depth of conformational and energetic analysis of such methods should be a valuable complement to less automated approaches that are based upon knowledge and intuition. It is hoped that the present software will prove useful to chemists seeking to create small receptors with biomedical and industrial applications.

ACKNOWLEDGMENT

This publication was made possible by Grant Number GM61300 from the National Institute of General Medical Sciences (NIGMS) of the NIH. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS. A release version of ConCept is under development.

REFERENCES AND NOTES

(1) Yang, G. F.; Wang, H. B.; Yangt, W. C.; Gao, D. Q.; Zhan, C. G. Bioactive Permethrin/Beta-Cyclodextrin Inclusion Complex. *J. Phy. Chem. B* **2006**, *110*, 7044−7048.

(2) Gibaud, S.; Ben, Z. S.; Mutzenhardt. P.; Fries, I.; Astier, A. Melarsoprol-Cyclodextrins Inclusion Complexes. *Int. J. Pharm.* **2005**, *306*, 107−121.

(3) Challa, R.; Ahuja, A.; Ali, J.; Khar, R. K. Cyclodextrins in Drug Delivery: An Updated Review. *AAPS Pharmscitech* **2005**, *6*, E329−E357.

(4) de Boer, H. D.; van Egmond, J.; van de Poll, F.; Bom, A.; Booij, L. H. D. J. Chemical Encapsulation of Rocuronium by Synthetic Cyclodextrin Derivatives: Reversal of Neuromuscular Block in Anaesthetized Rhesus Monkeys. *Br. J. Anaesth.* **2006**, *96*, 201−206.

(5) Chen, W.; Chang, C.-E.; Gilson, M. K. Concepts in Receptor Design: Targeting the RGD Peptide. *J. Am. Chem. Soc.* **2006**, *128*, 4675−4684.

(6) Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discovery* **2005**, *4* (8), 649−663.

(7) Hay, B. P.; Firman, T. K. HostDesigner: A Program for the de novo Structure-Based Design of Molecular Receptors with Binding Sites That Complement Metal Ion Guests. *Inorg. Chem.* **2002**, *41*, 5502−5512.

(8) Bryantsev, V. S.; Hay, B. P. De novo Structure-Based Design of Bisurea Hosts for Tetrahedral Oxoanion Guests. *J. Am. Chem. Soc.* **2006**, *128*, 2035−2042.

(9) Chang, C.-E.; Gilson, M. K. Free Energy, Entropy, and Induced Fit in Host-Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.

(10) Chen, W.; Chang, C.-E.; Gilson, M. K. Calculation of Cyclodextrin Binding Affinities: Energy, Entropy, and Implications for Drug Design. *Biophys. J.* **2004**, *87*, 3035−3049.

(11) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047−1069.

(12) Chang, C.-E.; Gilson, M. K. Tork: Conformational Analysis Method for Molecules and Complexes. *J. Comput. Chem.* **2003**, *24*, 1987−1998.

(13) Chen, W.; Huang, J.; Gilson, M. K. Identification of Symmetries in Molecules and Complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1301−1313.

(14) Chang, C.-E.; Potter, M. J.; Gilson, M. K. Calculation of Molecular Configuration Integrals. *J. Phys. Chem. B* **2003**, *107*, 1048−1055.

(15) Ruloff, R.; Seelbach, U. P.; Merbach, A. E.; Klarner, F.-G. Molecular Tweezers as Synthetic Receptors: The Effect of Pressure and Temperature on the Formation of Host-Guest Complexes. *J. Phys. Org. Chem.* **2002**, *15*, 189−196.

(16) Klarner, F.-G.; Lobert, M.; Naatz, U.; Bandmann, H.; Boese, R. Synthesis and Supramolecular Properties of Trimethylene-Bridged Clips. *Chem.- Eur. J.* **2003**, *9*, 5036−5047.

(17) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. *J. Phys. Chem.* **1997**, *101*, 3005−3014.

(18) Gilson, M. K.; Honig, B. The Inclusion of Electrostatic Hydration Energies in Molecular Mechanics Calculations. *J. Comput.-Aided. Mol. Des.* **1991**, *5*, 5−20.

(19) Friedman, R. A.; Honig, B. A Free-Energy Analysis of Nucleic-Acid Base Stacking in Aqueous-Solution. *Biophys. J.* **1995**, *69*, 1528−1535.

(20) Gilson, M. K.; Gilson, H. S. R.; Potter, M. J. Fast Assignment of Accurate Partial Atomic Charges: An Electronegativity Equalization Method That Accounts for Alternate Resonance Forms. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1982−1997.

(21) 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.

(22) Goodford, P. J. A Computational-Procedure for Determining Energetically Favorable Binding-Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(23) Leach, A. R. *Molecular Modelling: Principles and Applications*, 2nd ed.; Prentice Hall: 2001; p 480.

(24) Klarner, F.-G.; Panizky, J.; Blaser, D.; Boese, R. Synthesis and Supramolecular Structures of Molecular Clips. *Tetrahedron* **2001**, *57*, 3673−3687.

(25) Rotello, V. M.; Viani, E. A.; Deslongchamps, G.; Murray, B. A.; Rebek, J., Jr. Molecular Recognition in Water - New Receptors for Adenine-Derivatives. *J. Am. Chem. Soc.* **1993**, *115*, 797−798.

(26) Kato, Y.; Conn, M. M.; Rebek, J., Jr. Hydrogen-Bonding in Water Using Synthetic Receptors. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1208−1212.

(27) Cappello, V.; Tramontano, A.; Koch, U. Classification of Proteins Based on the Properties of the Ligand-Binding Site: The Case of Adenine-Binding Proteins. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 106−115.

(28) Kolb, H. C.; Finn, M. G.; Sharpless, K. B. Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew. Chem., Int. Ed.* **2001**, *40*, 2004.

(29) Lewis, W. G.; Green, L. G.; Grynszpan, F.; Sharpless, K. B. Click Chemistry in situ: Acetylcholinesterase as A Reaction Vessel for the Selective Assembly of A Femtomolar Inhibitor from An Array of Building Blocks. *Angew. Chem., Int. Ed.* **2002**, *41*, 1053.

(30) Lauri, G.; Bartlett, P. A. CAVEAT - A Program to Facilitate the Design of Organic-Molecules. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 51−66.

(31) Yang, W.; He, H.; Drueckhammer, D. G. Computer-Guided Design in Molecular Recognition: Design and Synthesis of A Glucopyranose Receptor. *Angew. Chem., Int. Ed.* **2001**, *40*, 1714−1718.

CI600233V