J|A|C|S
ARTICLES
Published on Web 12/27/2005

Computational Design of Heterochiral Peptides against a Helical Target

Vikas Nanda*,†,‡ and William F. DeGrado§,⊥

Contribution from the Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Center for Advanced Medicine and Biotechnology, Piscataway, New Jersey 08854, Department of Biochemistry and Molecular Biophysics, School of Medicine, and Department of Chemistry, School of Arts and Sciences, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received July 5, 2005; E-mail: nanda@cabm.rutgers.edu

Abstract: Polypeptides incorporating D-amino acids occasionally occur in nature and are an important class of pharmaceutical molecules. With the use of heterochiral Monte Carlo (HCMC), a method inspired by the de novo design of proteins, we develop peptide scaffolds for interacting with a molecular target, a left-handed α-helix. The HCMC approach concurrently seeks to optimize a peptide sequence, its internal conformation, and its docked conformation with a target surface. Several major classes of interactions are observed: (1) homochiral interactions between two α_L helices, (2) heterochiral interactions between an α_L and an α_R helix, and (3) heterochiral interactions between the α_L target and novel nonhelical structures. We explore the application of HCMC to simulating the preferential enantioselectivity of heterochiral complexes. Implications for biomimetic design in molecular recognition are discussed.

1. Introduction

Creating biomimetic polymers that specifically recognize proteins surfaces is a challenging problem in molecular design. Foldamers are synthetic polymers that recapitulate many of the physical properties of proteins.[1] Heterochiral peptides are one such class of biomimetics with a potential for greater structural diversity than peptides consisting solely of L-amino acids.[2] In this work, we introduce a computational approach for designing heterochiral peptides that recognize a molecular surface.

Ribosomally encoded proteins are composed almost solely of the 20 naturally occurring amino acids. With this small alphabet, proteins are able to achieve a fantastic diversity of structure and function. All of these amino acids, with the exception of glycine, have the *R* stereoisomer of the side chain at the Cα atom. Nature occasionally goes to great lengths to synthesize *heterochiral* peptides, which often function as toxins and antimicrobials. Because the ribosome is already dedicated to synthesizing natural proteins from an RNA template, the cell must exploit alternative routes. Massive protein complexes are required to direct the synthesis of these heterochiral peptides, with each subunit responsible for adding *one* amino acid.[3,4] Although it is resource consuming to make these large multi-domain proteins, combining L- and D-amino acids confers a functional advantage because a greater chemical diversity of backbone topologies becomes available to the cell. Incorporating D-amino acids into peptides also reduces their susceptibility to proteolysis, making them useful for biomedical applications.[5−8] Due to the high cost of creating the nonribosomal synthases, biological heterochiral peptides are generally very short, rarely longer than 10−15 amino acids in length. Through the use of glycine, ribosomally encoded proteins can occasionally access these unique topologies such as the left-handed α-helix of residues 38−56 alanine racemase[9] or the cat-grip motif[10] found in some ion binding sites. These are also relatively short due to the high entropic cost of fixing a glycine in a unique conformation. Such unusual conformations are only found in the context of stabilizing interactions with the remainder of the protein. With synthetic chemistry, we can potentially generate heterochiral peptides of lengths not feasible in nature and which can fold independently.

Atomistic modeling methods have previously been applied to the investigation of heterochiral peptide conformations. Computational and experimental studies focused on repeated heterochiral sequence patterns such as $(DL)_n$ or $(DDLL)_n$.[11−19] Searches for the optimal configurational energy used periodic

† Department of Biochemistry, UMDNJ.
‡ Center for Advanced Medicine and Biotechnology.
§ Department of Biochemistry and Molecular Biophysics, University of Pennsylvania.
⊥ Department of Chemistry, University of Pennsylvania.

(1) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. *Chem. Rev.* **2001**, *101*, 3893−4011.
(2) Aravinda, S.; Shamala, N.; Roy, R. S.; Balaram, P. *Proc. Indian Acad. Sci., Chem. Sci.* **2003**, *115*, 373−400.
(3) Challis, G. L.; Naismith, J. H. *Curr. Opin. Struct. Biol.* **2004**, *14*, 748−756.
(4) Finking, R.; Marahiel, M. A. *Annu. Rev. Microbiol.* **2004**, *58*, 453−488.
(5) Wade, D.; Boman, A.; Wahlin, B.; Drain, C. M.; Andreu, D.; Boman, H. G.; Merrifield, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 4761−4765.
(6) Oren, Z.; Hong, J.; Shai, Y. *J. Biol. Chem.* **1997**, *272*, 14643−14649.
(7) Papo, N.; Shahar, M.; Eisenbach, L.; Shai, Y. *J. Biol. Chem.* **2003**, *278*, 21018−21023.
(8) Van Regenmortel, M. H.; Muller, S. *Curr. Opin. Biotechnol.* **1998**, *9*, 377−382.
(9) Shaw, J. P.; Petsko, G. A.; Ringe, D. *Biochemistry* **1997**, *36*, 1329−1342.
(10) Watson, J. D.; Milner-White, E. J. *J. Mol. Biol.* **2002**, *315*, 183−191.
(11) Chandrasekaran, R.; Ramachandran, G. N. In *2nd American Peptide Symposium*, 1st ed.; Lande, S., Ed.; Gordon and Breach: Cleveland, OH, 1970; Vol. 2, pp 195−215.

10.1021/ja054452t CCC: $33.50

constraints on the backbone conformation, correlating the ϕ and ψ bond rotation angles of even- or odd-numbered residues. Many of the structurally characterized heterochiral secondary structures such as the β-helix of gramicidin and β-spiral of spider silk were also predicted by computer modeling and molecular mechanics simulations.[20−23] Periodic heterochiral sequences yield interesting topologies, generally inaccessible to ribosomally encoded proteins. However, novel conformations can be attained by peptides with aperiodic combinations of L- and D-amino acids as well, such as the "golf club" topology of the antimicrobial, tolaasin.[24,25] The heterochiral Monte Carlo (HCMC) approached developed in our group allows us to go beyond periodic systems and look at peptides with complete variability in chirality.

In our previous work, we used HCMC to study the effect of local backbone−side chain interactions on the coevolution of structure and chirality in polyalanine.[26] Peptides of different lengths were allowed to simultaneously fold and vary side chain chirality. The dominant solutions were left- and right-handed α-helices with a reversal in chirality at the second to last position of the peptide. This change in chirality facilitated a C-capping interaction similar to the Schellman motif mediated by glycine in natural proteins.[27−29] A second class of solutions consisted of peptides where the change in chirality occurred in the middle of the sequence resulting in a bent helix consisting of α_L and α_R components. A number of other solutions were obtained involving aperiodic combinations of L- and D-amino acids. These solutions highlighted the potential for HCMC as a general method for concurrently optimizing backbone flexibility and side chain chirality.

Rational design methods have been applied to the development of polymers that target protein surfaces.[30] By designing polymers that present functional groups in the same fashion as regular protein secondary structures such as an α-helix or β-strand, it is possible to disrupt protein−protein interactions. Some applications include the rational design of terphenyl-based α-helix mimics targeting the MLCK binding surface of calmodulin[31] and β-peptides that target HIV gp41[32] or disrupt the Bcl-x/Bak interaction.[33,34] D-Amino acids have also been used as antigen mimetics[35] as seen in the case of *retroinverso* peptides consisting of an all-D version of a natural antigen with the amino acids in the reverse order.[36] For a linear peptide lacking any folded structure, this results in an equivalent presentation of side chains to the natural L-amino acid sequence. Because D-peptides are resistant to proteolysis, retroinverso molecules have shown potential for use as vaccines. Given the structural diversity of mixed L- and D-amino acid peptides, it would be interesting to advance design of heterochiral peptides beyond retroinverso applications to problems of molecular recognition where a folded ligand with a defined three-dimensional structure is desirable. Studies on the importance of chirality in α-helix[37−39] and β-sheet[40] interactions highlight the need for a molecular understanding of the underlying molecular basis of enantioselective recognition.

In this work, we present an automated method for designing heterochiral peptides that recognize a target molecule—a left-handed α-helix composed of D-alanine in our case. The conformation of the target is fixed throughout the simulation. A second, variable peptide is then docked against this target. During the HCMC simulation, both sequence and structure are simultaneously optimized, allowing each amino acid in the variable peptide to alternate between L and D chirality, sampling folded conformations of the variable peptide that maximize favorable intramolecular interactions and docking the variable peptide against the fixed target. Several major classes of interactions are found where the variable peptide is (1) a left-handed α-helix composed of D-alanine, (2) a right-handed α-helix of L-alanine, or (3) a heterochiral, nonhelical conformation. An analysis of the peptide complexes shows a recapitulation of helix packing modes found in natural proteins as well as novel binding modes unique to heterochiral systems.

2. Methods

2.1. HCMC. The HCMC approach is based on simulated evolution, a protocol for the sequence redesign of natural proteins.[41] To model molecular recognition of the target molecule and promote folding of the variable polyalanine peptide, the energy calculation is broken into *intra-* and *intermolecular* components. The intramolecular energy of the variable peptide is calculated in the same manner as used in our previous work.[26] Conformational energy is the sum of van der Waals (vdW) interactions, backbone amide to backbone carbonyl hydrogen bonds, and clashes between the methyl carbon (Cβ) of the alanine side chain and the carbonyl oxygen of both *its backbone* and that of the *previous residue*.

$$E_{\text{intra}} = E_{\text{VDW}} + \epsilon_{\text{clash}} N_{\text{clash}} + \epsilon_{\text{hbond}} N_{\text{hbond}} \quad (1)$$

The energy per clash is set to $\epsilon_{\text{clash}} = 5.0$ kcal mol^{-1}, and the energy per hydrogen bond is $\epsilon_{\text{hbond}} = -5.0$ kcal mol^{-1}. The vdW energy is

(12) Colonna-Cesari, F.; Premilat, S.; Heitz, F.; Spach, G.; Lotz, B. *Macromolecules* **1977**, *10*, 1284−1288.
(13) Diblasio, B.; Benedetti, E.; Pavone, V.; Pedone, C.; Gerber, C.; Lorenzi, G. P. *Biopolymers* **1989**, *28*, 203−214.
(14) Heitz, F.; Detriche, G.; Vovelle, F.; Spach, G. *Macromolecules* **1981**, *14*, 47−50.
(15) Hesselink, F. T.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 455−463.
(16) Lorenzi, G. P.; Gerber, C.; Jackle, H. *Macromolecules* **1985**, *18*, 154−159.
(17) Lotan, N.; Hesselin, F. Th.; Benderly, H.; Yan, J. F.; Schechter, I.; Berger, A.; Scheraga, H. A. *Macromolecules* **1973**, *6*, 447−453.
(18) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95 ff.
(19) Rizzo, V.; Lorenzi, G. P. *Macromolecules* **1983**, *16*, 476−482.
(20) Urry, D. W.; Goodall, M. C.; Glickson, J. D.; Mayers, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1907 ff.
(21) Veatch, W. R.; Fossel, E. T.; Blout, E. R. *Biochemistry* **1974**, *13*, 5249−5256.
(22) Bamberg, E.; Apell, H. J.; Alpes, H. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 2402−2406.
(23) Becker, N.; Oroudjev, E.; Mutz, S.; Cleveland, J. P.; Hansma, P. K.; Hayashi, C. Y.; Makarov, D. E.; Hansma, H. G. *Nat. Mater.* **2003**, *2*, 278−283.
(24) Jourdan, F.; Lazzaroni, S.; Mendez, B. L.; Cantore, P. L.; Julio, M.; Amodeo, P.; Iacobellis, N. S.; Evidente, A.; Motta, A. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 534−543.
(25) Nutkins, J. C.; Mortishire-Smith, R. J.; Packman, L. C.; Brodey, C. L.; Rainey, P. B.; Johnstone, K.; Williams, D. H. *J. Am. Chem. Soc.* **1991**, *113*, 2621−2627.
(26) Nanda, V.; DeGrado, W. F. *J. Am. Chem. Soc.* **2004**, *126*, 14459−14467.
(27) Schellman, C. In *Protein Folding*; Jaenicke, R., Ed.; Elsevier/North-Holland: New York, 1980; pp 53−61.
(28) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648−1652.
(29) Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632−1641.
(30) Yin, H.; Hamilton, A. D. *Angew. Chem. Int. Ed.* **2005**, *44*, 4130−4163.
(31) Orner, B. P.; Ernst, J. T.; Hamilton, A. D. *J. Am. Chem. Soc.* **2001**, *123*, 5382−5383.
(32) Stephens, O. M.; Kim, S.; Welch, B. D.; Hodsdon, M. E.; Kay, M. S.; Schepartz, A. *J. Am. Chem. Soc.* **2005**, *127*, 13126−13127.
(33) Gemperli, A. C.; Rutledge, S. E.; Maranda, A.; Schepartz, A. *J. Am. Chem. Soc.* **2005**, *127*, 1596−1597.
(34) Kritzer, J. A.; Stephens, O. M.; Guarracino, D. A.; Reznik, S. K.; Schepartz, A. *Bioorg. Med. Chem.* **2005**, *13*, 11−16.
(35) Benkirane, N.; Friede, M.; Guichard, G.; Briand, J. P.; Van Regenmortel, M. H.; Muller, S. *J. Biol. Chem.* **1993**, *268*, 26279−26285.
(36) Chorev, M.; Goodman, M. *Trends Biotechnol.* **1995**, *13*, 438−445.
(37) Gerber, D.; Shai, Y. *J. Mol. Biol.* **2002**, *322*, 491−495.
(38) Sia, S. K.; Kim, P. S. *Biochemistry* **2001**, *40*, 8981−8989.
(39) Sal-Man, N.; Gerber, D.; Shai, Y. *J. Mol. Biol.* **2004**, *344*, 855−864.
(40) Chung, D. M.; Nowick, J. S. *J. Am. Chem. Soc.* **2004**, *126*, 3062−3063.
(41) Hellinga, H. W.; Richards, F. M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 5803−5807.

computed using a 12−6 Lennard-Jones potential with a distance cutoff of 6.5 Å. The vdW term is also radius scaled by a factor of 0.9 to facilitate conformational searching, and the repulsive energy is capped at 10 kcal mol^{-1}.[42,43] Atomic parameters are taken from the united atom set of the AMBER force field.[44] A hydrogen bond is counted if donor and acceptor heavy atoms are within 3.2 Å and the N−H···O angle is greater than 90°. A side chain−backbone clash is counted only if the relevant atoms are separated by less than 3.0 Å.

An intermolecular vdW energy is calculated between the variable and the target chain, with interactions between side chain methyl carbons (Cβ) across the two chains scaled by a factor of 2.0 in order to promote intermolecular contacts. Additionally, a weak restraint is added to improve sampling of closely interacting molecules:

$$E_{\text{restraint}} = \epsilon_{\text{restraint}}(3.0 - r_{\text{centroid}}) \quad (2)$$

where $\epsilon_{\text{restraint}}$ is 1.0 kcal mol^{-1} Å^{-1} and r_{centroid} is the separation (in Å) between the center of mass of the two chains. The total energy is the sum of intramolecular and intermolecular terms:

$$E_{\text{tot}} = E_{\text{intra}} + E_{\text{restraint}} + E_{\text{vdW}}(\text{target, variable}) + E_{\text{vdW}}(\text{target C}\beta, \text{variable C}\beta) \quad (3)$$

Two chains are treated in the current protocol: the *target*, which is a left-handed α-helix composed of D-alanine, and a *variable* peptide, which has a flexible backbone and a variable sequence consisting of L- and D-alanine. The HCMC protocol allows for either a change in the variable peptide sequence and internal conformation or a change in its position relative to the target surface (Figure 1). At the beginning of the simulation, the variable chain starts in an extended conformation ($\phi,\psi = 180.0°$). The initial sequence of L- and D-alanines is randomly specified. At each iteration of HCMC, a random choice is made whether the variable peptide undergoes a change in position relative to the target or, instead, a change in its internal conformation and/or sequence. If a change in position is chosen (Figure 1, right path), the three rotational and translational degrees of freedom are allowed to vary. Each of the Eulerian angles are adjusted by an angle within ±15.0°, and each of the translations (x,y,z) can vary by within ±1.0 Å. All rotations are around the local coordinate frame of the variable peptide with the origin at the center of mass. Upper and lower bounds on r_{centroid} of 12.0 and 3.0 Å are maintained. Any transformation that exceeds these bounds is rescaled by moving the peptide along the r_{centroid} axis to bring it within the distance constraints. If a change in configuration/sequence is chosen instead (Figure 1, left path), then a residue in the sequence of the variable peptide is chosen at random, and one of three permutations is attempted: (1) the chirality of the Cα is inverted (D- to L-alanine or vice versa), (2) the ϕ and ψ angles are changed, or (3) both modifications 1 and 2 simultaneously on the same residue. For the first residue, only ψ is defined, and for the last, only ϕ is defined and thus allowed to vary.

HCMC is run using a Metropolis-type Monte Carlo simulated annealing. Moves are accepted based on the Metropolis criteria,[45] where the probability of accepting the move, a, is

$$a = \min(1, \exp(-T^{-1}(E_i - E_{i-1}))) \quad (4)$$

where T is the selection temperature and E_i is the computed energy. T decreases linearly with iteration number with $T_0 = 10^5$ and $T_{\text{final}} = 1.0$. Each HCMC simulation is run for 10^6 cycles.

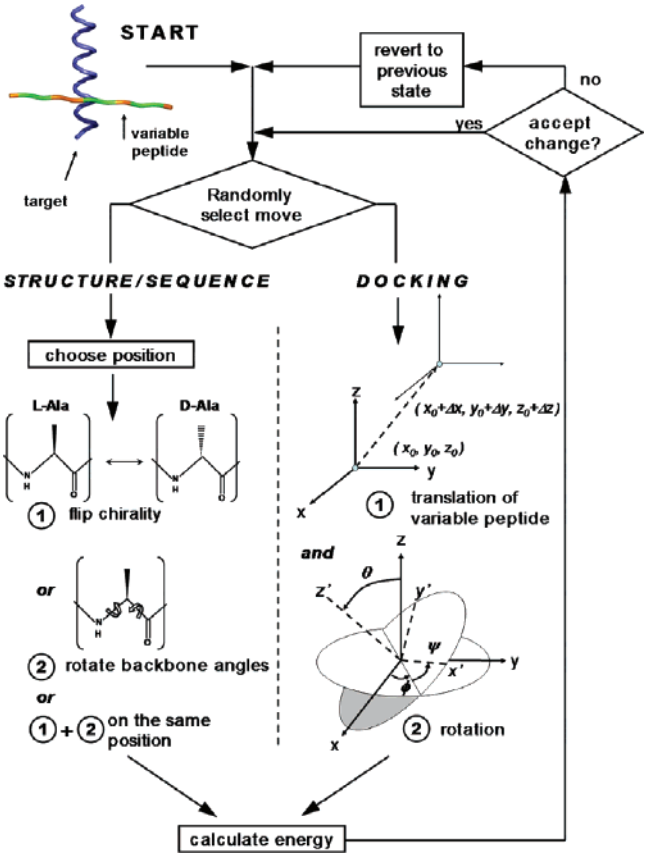

Figure 1. HCMC protocol augmented with docking. The left path of the flowchart is the original HCMC protocol used in our previous study. The right path represents an additional docking move which allows the variable peptide to sample various binding modes on the target surface. Three translational and three rotational degrees of freedom are allowed. All transformations are applied to a coordinate frame which is at the center of mass of the variable peptide.

2.2. Analysis of Helix Geometry. Helical solutions from HCMC are analyzed to determine the crossing angle, Ω, and interhelical separation of the variable peptide against the target. The definition of an α_R or an α_L helical solution is satisfied if six contiguous residues are within the backbone angle values defined as α_R or α_L. α_R is defined as $0° > (\phi, \psi) > -100°$ and α_L as $0° < (\phi, \psi) < 100°$. Local helix centers are calculated using all sets of four successive Cα atoms in the helical region by adapting the HELANAL algorithm.[46−48] These are connected with a best-fit line to specify the helical axis. Ω and interhelical separation parameters are calculated at the point of closest approach between the axes of the two helices.

2.3. Geometric Enumeration of Helix Pair Conformations. The goal of this computation is to effectively sample across all possible helix dimer conformations and determine the intermolecular energy of packing as a function of crossing angle, Ω. Ideal 3.6-residue/turn α-helices are generated for both left-handed poly-D-alanine and right-handed poly-L-alanine (Figure 2). Dimer conformations are sampled at all crossing angles, Ω, between −180° and 180° in 10° increments. One helix axis is fixed coincident with z. The second is rotated parallel to the yz-plane. For each Ω, all other degrees of freedom are enumerated to determine the lowest energy configuration. The interhelical separation, r, is sampled between 6.0 and 12.0 Å in 0.5 Å increments. Helical rotations, ϕ_1 and ϕ_2, are sampled from −60° to 60° in 5° increments to allow for all possible facial interactions between the two helices. The

(42) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10383−10388.
(43) Gordon, D. B.; Marshall, S. A.; Mayo, S. L. *Curr. Opin. Struct. Biol.* **1999**, *9*, 509−513.
(44) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S. J.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(45) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1951**, *21*, 1087−1092.
(46) Kumar, S.; Bansal, M. *Biophys. J.* **1998**, *75*, 1935−1944.
(47) Bansal, M.; Kumar, S.; Velavan, R. *J. Biomol. Struct. Dyn.* **2000**, *17*, 811−819.
(48) Lahr, S. J.; Engel, D. E.; Stayrook, S. E.; Maglio, O.; North, B.; Geremia, S.; Lombardi, A.; DeGrado, W. F. *J. Mol. Biol.* **2005**, *346*, 1441−1454.

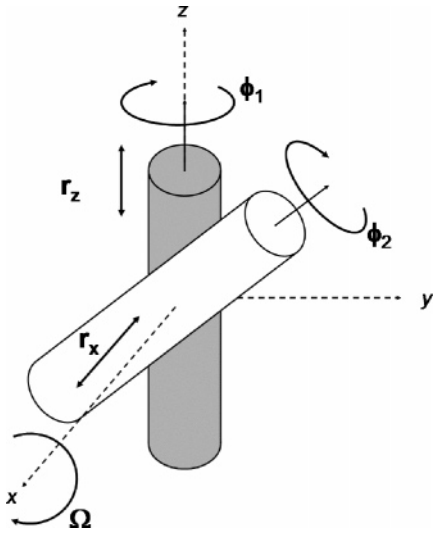

Figure 2. Geometric parameters sampled in mapping out the energy surface of helix pair crossings. For each Ω, a series of r_x, r_z, ϕ_1, and ϕ_2 were sampled (see text), and the lowest energy solution was chosen.

Table 1. Most Frequent Sequences out of 1650 Simulations[a]

DDDDDDDDDLL	21	(9)
DDDDDDDDDLD	13	(15)
LDDDDDDDDLL	10	(15)
LDDDDDDDDDD	14	(14)
LDDDDDDDDLD	13	(8)
DLDDDDDDDLD	8	(13)
DDDDDDDDDDD	13	(12)
DLDDDDDDDLL	8	(12)
DDLLLLLLLDL	11	(9)
DLDDDDDDDDD	11	(3)
LDDDLLLLLDL	11	(5)
LLLLLLDDDDL	10	(0)
LLDDDDDDDLL	10	(7)
LLLLLDDDDDD	10	(5)
DDDDDLLLLDD	10	(0)
LLLLLLLDDDL	9	(5)
DDDDDDDDDDL	5	(9)
DLLLLDDDDDD	8	(5)

[a] The Number in parentheses represents the counts of sequence stereoisomers. A total of 821 unique sequences were found.

helix offset, z_{trans}, is sampled between −2.0 and 2.0 Å in 0.2 Å increments. Energies are calculated using vdW energy alone with a distance cutoff of 6.5 Å. No atomic radius scaling or repulsive energy cutoff is used. Atom potentials are from the AMBER united atom parameter set.[44]

3. Results

3.1. Sequence and Secondary Structure. The HCMC simulations consist of a variable, 11-residue peptide to be docked against a 20-residue, left-handed α-helix composed of D-alanine. Our analysis is based on the final solutions from 1650 independent HCMC simulations. From these simulations, 821 unique sequences are found. This is more diverse than our previous work on an 11-residue polyalanine in the absence of a target, where 371 unique sequences were found in 1000 simulations.[26] The two most frequently occurring classes of sequences contain nine contiguous D- or L-alanines (Table 1). These molecules contain a chirality reversal at the C-terminus which functions as a capping interaction, similar to the capping interactions of glycine in natural proteins and D-amino acids in designed peptides.[27−29,49] Both of these are observed 40-fold more frequently than randomly expected. In addition to solutions with appreciable contiguous homochiral sequence segments, a number of heterochiral sequences are observed containing shorter contiguous segments of L-alanine and D-alanine. These belong to the helix-reversal subclass of structures—which are characterized by a juxtaposition of a left- and a right-handed turn of a helix.[50−52]

An analysis of the secondary structure profiles of this set of simulations corroborates the observations of sequence (Figure 3). The most frequent secondary structures are the left- and right-handed α-helices. There is an even distribution of α_L and α_R content, except notably for $(\alpha_L)_8$−IV, which occurs twice as frequently as $(\alpha_R)_8$−II (II and IV refer to the quadrant of Ramachandran space as denoted in Figure 3). We study the possibility for stereochemical bias introduced by the target in section 3.5.

3.2. Folding Trajectory. Typically, the system folds in two stages during the linear cooling protocol: (1) folding into a compact secondary structure and (2) optimizing interchain contacts (Figure 4). The initial variable peptide collapses from its fully extended state into a subcompact conformation with significant intramolecular hydrogen bonding while maintaining some interactions with the target chain. A folding transition occurs, resulting in a defined secondary structure (a right-handed α-helix in the pictured example, occurring between 300 and 400 K). During the second phase, the variable peptide generally maintains its conformation and predominantly optimizes its geometry while sampling various binding modes against the target. As expected for a helical solution, the energy of the molecule improves with sequence homochirality. During the folding step, the target surface has the potential to influence the sequence and structure of the variable peptide. At 300 K in the pictured example, the variable peptide is in an "ambidextrous" state, with one turn of an α_L helix and one turn of an α_R helix. Preferential stabilizing of either turn by tertiary interactions with the target surface could result in an observed bias of one handedness over multiple simulations.

3.3. Analysis of Helix Pairs. In addition to the sequence and conformation, we are interested in structural details of how the variable peptide docks against the target. As previously indicated, the two most frequent conformations of the variable peptide are an α_L and an α_R helix. These are defined as molecules containing six or more contiguous positions in the corresponding regions of Ramachandran space. Parameters such as crossing angle and interhelical separation are defined from the central axis of each helix at the point of closest approach to the other helix axis. Here we present a separate analysis of α_L–α_L and α_L–α_R pairs.

The first class of solutions, consisting primarily of poly-D-alanine in an α_L conformation, participate in α_L–α_L helical dimers similar to the mirror image of geometries found in natural proteins between two right-handed helices. Helices showed a

(49) Schneider, J. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 2764−2767.
(50) Karle, I. L. *Biopolymers* **2001**, *60*, 351−365.
(51) Aravinda, S.; Shamala, N.; Pramanik, A.; Das, C.; Balaram, P. *Biochem. Biophys. Res. Commun.* **2000**, *273*, 933−936.
(52) Banerjee, A.; Raghothama, S. R.; Karle, I. L.; Balaram, P. *Biopolymers* **1996**, *39*, 279−285.

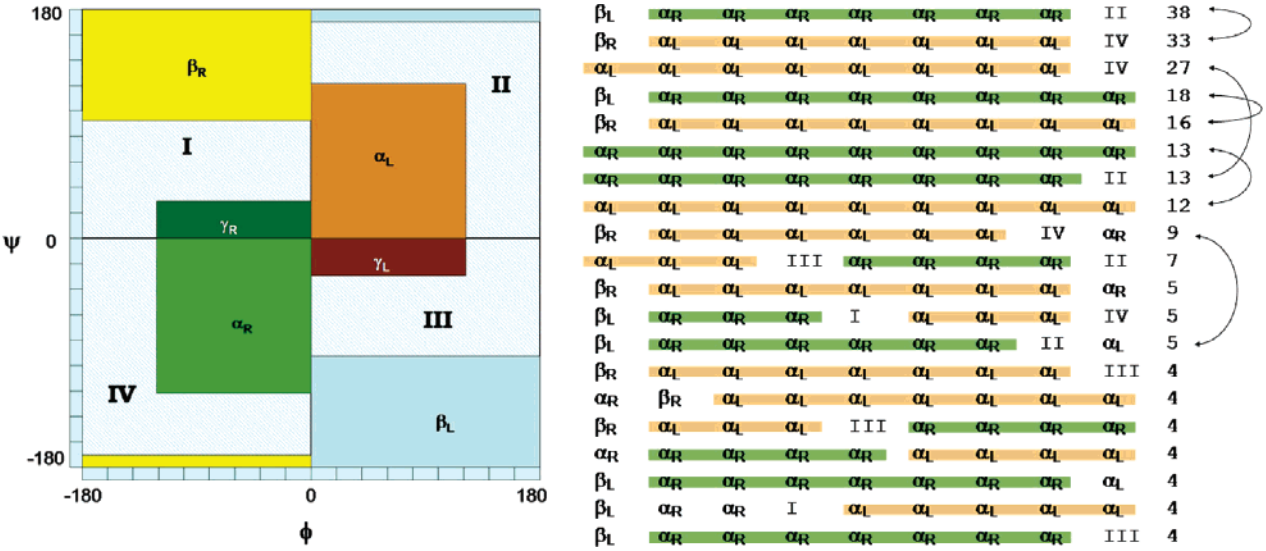

Figure 3. Most frequently occurring conformations from 1650 simulations. Structural enantiomers are linked by arrows on the right. Definitions of conformational states are shown in the Ramachandran key on the left.

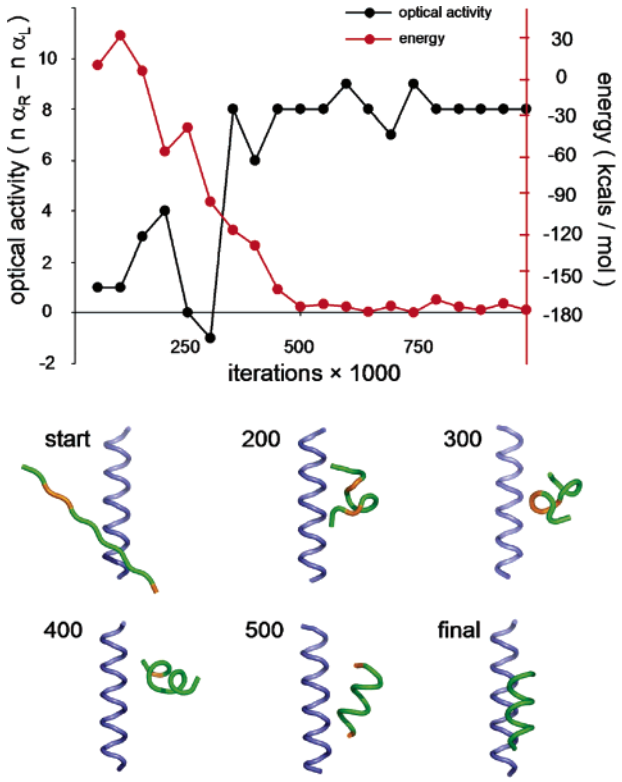

Figure 4. HCMC trajectory of a variable peptide against an α_L helix target, following optical activity ($N(\alpha_R) - N(\alpha_L)$) and the computed energy of the complex. The target helix is in blue, the variable peptide D-alanine in orange, and L-alanine in green.

mean separation of 6.5 Å between the two helix centers (Figure S1 in the Supporting Information). Of the 170 solutions with at least six contiguous α_L residues, we find that most pack with a crossing angle around $20° < \Omega < 30°$ or $-30° < \Omega < -50°$ (we define $\Omega > 0$ as a right-handed crossing and $\Omega < 0$ as left-handed crossing) (Figure 5A). Crossing angle distributions in natural proteins have similar biases in Ω (with opposite sign given proteins are composed of α_R helices):[53] natural protein helices pack with an angle near $-25°$ as part of the canonical left-handed coiled coil. Geometric analysis of an $\alpha_L-\alpha_L$ pair shows the ridges-in-grooves packing that results in the corresponding right-handed dimer (Figure 6A) (see the Supporting Information methods for ridges-in-grooves analysis of packing). The crossing angle statistics are compared to an energy surface of helix pair interactions. Ω is sampled through 360° in 10° increments. The interhelix packing geometry is optimized and the energy computed as described in the Methods (Figure 5B). The energy surface calculation indicates two major minima near the peaks of the Ω histogram from HCMC. The right-handed crossing angle (25°) is coincident in both HCMC and energetic sampling. However, the left-handed dimer crossing angle differs by 10−20° as estimated by the two methods.

A second class of solutions reveals novel $\alpha_L-\alpha_R$ heterochiral helix packing interfaces which would not occur in natural proteins. Here, 165 solutions were found with six or more contiguous α_R residues. The Ω-distribution is symmetric around $\Omega = 0°$ (Figure 5C). This is due to the enantiomeric relationship between an $\alpha_R-\alpha_L$ helix pair at $\Omega = +x$ and the corresponding $\alpha_L-\alpha_R$ pair at $\Omega = -x$. A frequent packing interface at ±50° is observed. On the basis of the energetic sampling and geometric analysis of heterochiral helix pairs, one would predict a high propensity for a crossing angle of 0° (Figure 6B). This is the basis of the D/L 4-helix bundle designed by Sia and Kim.[38] Instead, there are very few observed examples of this angle. Because packing angles near zero are geometrically unlikely to be sampled, this bias can be adjusted for by multiplying the observed distribution with $\sin^{-1}(\Omega)$.[54] Additionally, the 0° crossing motif requires an ideal helix solution such as the final complex of the HCMC simulation in Figure 4. Most sequences deviate from the α-helix at the N- and C-termini due to end fraying and the presence of capping structures. These end effects disrupt homogeneous helix−lattice packing, but can be accommodated by rotating the helix to the left or right, thus promoting favorable contacts between nonhelical residues and the target (Figure 7).

3.4. Nonhelical Solutions. Clustering solutions based on ϕ/ψ RMSD using *kclust* (http://mmstb.scripps.edu) result in mostly

(53) Chothia, C.; Levitt, M.; Richardson, D. *J. Mol. Biol.* **1981**, *145*, 215−250.

(54) Kroon, J.; Kanters, J. A.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B.; Vliegenthart, J. A. *J. Mol. Struct.* **1975**, *24*, 109−129.

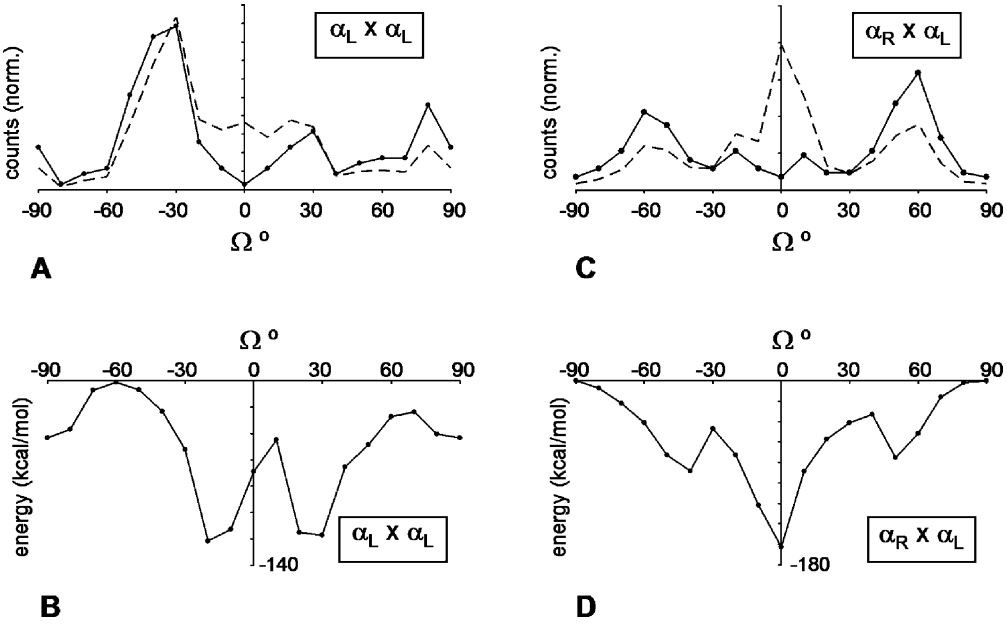

Figure 5. (A) Histogram of the crossing angle, Ω, for α_L helices against the α_L target from 1650 HCMC simulations. The solid line is the raw counts (area normalized), and the dotted line is corrected for orientational sampling. (B) Energy surface for an $\alpha_L-\alpha_L$ helix pair (negative values are more favorable). (C) Histogram of Ω for α_R helices from the HCMC simulations. (D) Energy surface for an $\alpha_R-\alpha_L$ pair.

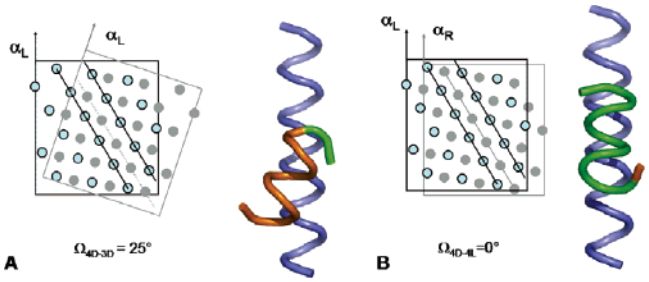

Figure 6. (A) Helix net diagram for a right-handed $\alpha_L-\alpha_L$ dimer. The black helix net with blue circles is facing toward the viewer, and the gray helix net with gray circles is facing away. This is the mirror image of the packing observed in natural protein left-handed coiled coils. (B) An $\alpha_L-\alpha_R$ dimer with a 0° crossing angle. This interface is unique to heterochiral helix dimers.

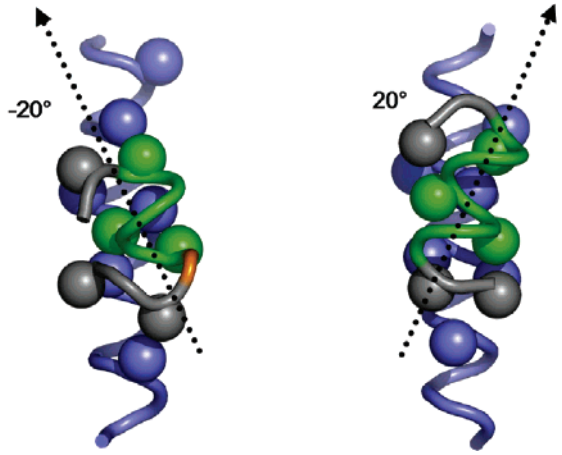

Figure 7. Examples of nonideal heterochiral helix packing with nonhelical interacting side chains in gray. Spheres are shown for Cβs involved in intermolecular contacts.

singletons, other than the previously described helical solutions. However, 45 cases of nonhelical molecules occur which adopt a $\beta_R-(\alpha_L)_3-x_L-(\alpha_R)_3-x_R$, where x_L and x_R are near but do not fall into left or right α-helix. This is effectively the

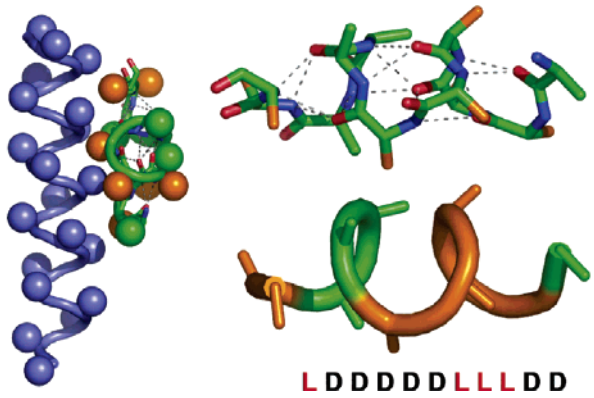

Figure 8. Heterochiral motif consisting of one α_L and one α_R turn docking against the target.

concatenation of a π-turn and its structural stereoisomer, which comprise a common class of 4-residue β-hairpin turns in natural proteins, adopting an $\alpha_R-\alpha_R-\gamma_R-\alpha_L$ conformation.[55−57] In HCMC simulations of longer peptides, we previously found this motif at the helix sense reversal point of ambidextrous helices. These also form a network of local hydrogen bonds promoting the cooperativity of folding. These molecules are able to effectively pack against the target surface, as seen in Figure 8. Between five and seven side chains of the variable peptide participate in intermolecular contacts with the target.

3.5. Enantioselective Enhancement. As suggested earlier, during the initial folding stage of the HCMC simulation, there is the potential for the target to bias the sequence and conformation of the variable peptide. At the 300 K snapshot in Figure 4, the variable peptide is in an ambidextrous state with one α_L and one α_R turn. If either turn has a preferential interaction with the target, it would bias the remaining peptide

(55) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759−777.
(56) Sibanda, B. L.; Thornton, J. M. *Nature* **1985**, *316*, 170−174.
(57) Gunasekaran, K.; Ramakrishnan, C.; Balaram, P. *Protein Eng.* **1997**, *10*, 1131−1141.

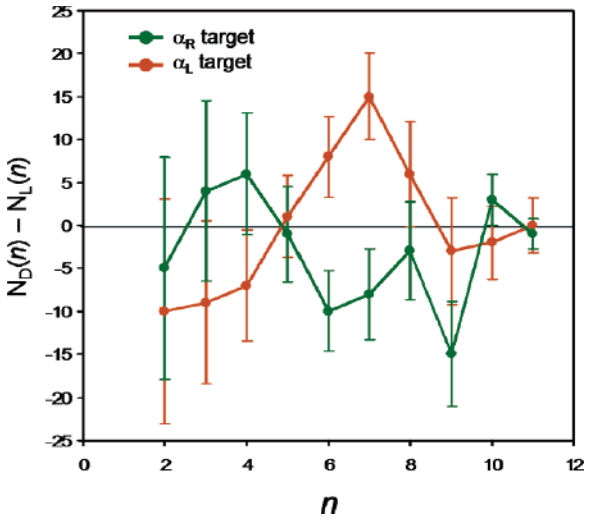

Figure 9. Induced chiral bias from docking a variable peptide against a left-handed (orange) or right-handed (green) helix. *n* is the length of a homochiral segment, defined as ...$a-b_n-c$... where *a* and *c* are either the opposite chirality of *b* or the beginning/end of the peptide. Error bars represent estimated count variance $(N_D(n) + N_L(n))^{0.5}$.

to adopt a similar handedness. To test whether such an intermolecular chiral influence was occurring, we ran 250 simulations of an 11-residue variable peptide against a left-handed 20-residue poly-D-alanine target, as before, and 250 simulations against a right-handed poly-L-alanine target. Looking at total L-alanine and D-alanine or total α_R versus α_L content, there is no clear indication of chiral bias. However, we can represent the chiral bias for contiguous homochiral segments of length *n* as $N_D(n) - N_L(n)$ for the 250 simulations against a particular target (Figure 9). Against a left-handed target, there is a bias toward homochiral D-alanine peptides of lengths six and seven. Conversely, there is a bias toward homochiral L-alanine peptides of the same lengths when presented with an α_R target. With the use of this measure of chiral bias, it is evident that there is some preference for same-chirality interactions. Although the anticorrelation between α_L and α_R continues to longer values of *n*, it is not clear if this is statistically significant due to error from low counts.

4. Discussion

The field of computational biomimetic design against target surfaces is still in its early stages of development. Much of the current research in sequence design is focusing on the heteropolymer statistical mechanics of reduced dimensionality systems.[58−61] When moving from lattice-based methods to atomistic simulations, it is important to retain sufficient simplicity that effective sampling is realistic. The HCMC method concurrently optimizes sequence and structure with a modest move set and a simplified force field. We wish to examine whether HCMC is easily extendable to molecular recognition problems. The method is augmented with a molecular docking move set to evolve a variable peptide against a fixed target surface. This is the first step in the design of novel tertiary folds and the application of computational methods to the direction of heterochiral proteins against specific targets. In addition to concurrent optimization of structure and sequence, this method seeks to maximize favorable interactions with the target.

With the use of a geometric analysis that treats helices as ridges and grooves (see methods in the Supporting Information), it is possible to demonstrate some of the observed crossing angles including the 25° right-handed $\alpha_L-\alpha_L$ pair and the 0° $\alpha_L-\alpha_R$ pair. This approach has been instrumental in understanding fundamental features of helix packing in proteins. Similarly, using rigid body docking of ideal helices, it is possible to determine favorable modes of association. Such methods have been used to study conformations of helical oligomers, particularly in the membrane.[62,63] However, the observed deviations from ideal helical geometry in this study demonstrate the utility of using a modeling approach with full backbone flexibility. We are able to build upon insight from geometrical modeling and energy-based docking toward the design of more complex ligands directed against a target surface. Capping interactions which stabilize the helix termini also have the potential to contribute side chain interactions to the target surface (as seen in Figure 7). To accommodate these additional contacts, deviations from ideal packing geometry are needed. Also, conformations only accessible to heterochiral molecules, such as the helix reversal consisting of two α-turns of opposing handedness, are potential ligand conformations for this helical target (Figure 8). This conformation maintains a network of hydrogen bonds, making it likely to fold cooperatively and remain stable.

Because folding and sequence design are coupled with molecular recognition, the most common solutions not only seek to optimize intermolecular contacts, but they also sample *designable* conformations.[64−66] The HCMC protocol incorporates both energetic and entropic considerations into an optimal design. The more frequently sampled dimer conformations will be able to accommodate multiple sequences—a favorable feature for design. Designable structures are fertile starting points for additional sequence patterning.

The transfer of chiral information from the target to the variable peptide may be used to probe fundamental questions of homochirality. Studies done by the Ghadiri group have shown that in a self-templating peptide system, homochiral interactions are favored over heterochiral ones.[67] However, examples of engineered heterochiral helical bundles exist.[38,39] Likewise, different groups have found homochiral or heterochiral preferences for β-amyloid formation in model peptides.[68,69] Computational methods such as HCMC may be used to design peptides that favor either heterochiral or homochiral complexes. Understanding what is required to achieve enantioselectivity is one way to explore the molecular origin of protein homochirality in our biosphere. Concurrent docking and folding protocols

(58) Golumbfskie, A. J.; Pande, V. S.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 11707−11712.
(59) Kriksin, Y. A.; Khalatur, P. G.; Khokhlov, A. R. *J. Chem. Phys.* **2005**, *122*, 114703.
(60) Castells, V.; Yang, S. X.; Van Tassel, P. R. *Phys. Rev. E* **2002**, *65*, 031912.
(61) Jayaraman, A.; Hall, C. K.; Genzer, J. *Phys. Rev. Lett.* **2005**, *94*, 078103.
(62) Adams, P. D.; Arkin, I. T.; Engelman, D. M.; Brunger, A. T. *Nat. Struct. Biol.* **1995**, *2*, 154−162.
(63) Torres, J.; Kukol, A.; Arkin, I. T. *Biophys. J.* **2001**, *81*, 2681−2692.
(64) Li, H.; Tang, C.; Wingreen, N. S. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 4987−4990.
(65) Emberly, E. G.; Wingreen, N. S.; Tang, C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11163−11168.
(66) Li, H.; Helling, R.; Tang, C.; Wingreen, N. *Science* **1996**, *273*, 666−669.
(67) Saghatelian, A.; Yokobayashi, Y.; Soltani, K.; Ghadiri, M. R. *Nature* **2001**, *409*, 797−801.
(68) Esler, W. P.; Stimson, E. R.; Fishman, J. B.; Ghilardi, J. R.; Vinters, H. V.; Mantyh, P. W.; Maggio, J. E. *Biopolymers* **1999**, *49*, 505−514.
(69) Chalifour, R. J.; McLaughlin, R. W.; Lavoie, L.; Morissette, C.; Tremblay, N.; Boule, M.; Sarazin, P.; Stea, D.; Lacombe, D.; Tremblay, P.; Gervais, F. *J. Biol. Chem.* **2003**, *278*, 34874−34881.

could be adapted to the study of chiral domino systems where a chiral molecule can induce gross conformational change in an otherwise achiral system.[70,71]

In computational de novo protein design, one of the most important challenges is the concurrent optimization of target *conformation* and *sequence*. Both components on their own present formidable computational barriers to evaluation, and in combination, the problem becomes prohibitively large to solve directly. At the outset of this project, it was not clear whether it would be possible to concurrently search through full backbone flexibility, side chain chirality (sequence), and intermolecular conformation (docking) using a Monte Carlo simulated annealing approach. By focusing on the primary interactions that determine a polypeptide fold (hydrogen bonding and steric repulsion), it was possible to create realistic intermolecular complexes. This observation is a significant advance in computational protein and foldamer design, suggesting that MCSA techniques can be effectively applied to concurrent structure optimization and sequence patterning of a bimolecular complex using reasonable computational resources. Previous design efforts have focused on reduced dimensionality models of polymer-surface complexes. Our work is an important step beyond these studies to atomistic level simulations. By focusing on design of a polyalanine scaffold, it is possible to rapidly explore a large variety of sequences and conformations. We believe this approach will be particularly suited to the design of foldamers, where the available fold-space is not known from experimental structures.

This study is intended as a first step in a design approach where real surfaces can be targeted using HCMC. By expanding the side chain alphabet from alanine to a larger library of residues, we can provide increased driving force for association using larger hydrophobic groups and improved specificity of recognition using polar residues. The backbones generated from HCMC of polyalanine provide some clues as to what limitations of heterochiral geometries are possible—making the inclusion of greater sequence diversity computationally practical. More generally, by expanding computational design methods to heterochiral molecules, we can ascertain whether our accumulated knowledge of protein design is idiosyncratic to proteins or whether the tools developed for protein design and structure prediction will have utility in the larger problem of foldamer engineering.

Acknowledgment. This research was supported by the National Institutes of Health Grants (HL07971-0 and GM54616) and by the Nano/Bio Interface Center, National Science Foundation DMR 0425780.

Supporting Information Available: Additional simulation methods and results. This material is available free of charge via the Internet at http://pubs.acs.org.

JA054452T

(70) Inai, Y.; Ousaka, N.; Okabe, T. *J. Am. Chem. Soc.* **2003**, *125*, 8151−8162.
(71) Inai, Y.; Ishida, Y.; Tagawa, K.; Takasu, A.; Hirabayashi, T. *J. Am. Chem. Soc.* **2002**, *124*, 2466−2473.