Simulated Evolution of Emergent Chiral Structures in Polyalanine

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

Contribution from the Department of Biochemistry and Molecular Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104 and Department of Chemistry, School of Arts and Sciences, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Received June 28, 2004; E-mail: vikas@mail.med.upenn.edu

Abstract: The relationship between monomer chirality and polymer structure has been studied using both theoretical and experimental methods. Atomistic models, such as the ones employed in computational protein folding and design, can be used to study the relationship between monomer chirality and the properties of polypeptides. Using a simulated evolution approach that combines side-chain epimerization with backbone flexibility, we recapitulate the relationship between basic forces that drive secondary structure formation and sequence homochirality. Additionally, we find heterochiral motifs including a C-terminal helix capping interaction and stable helix-reversals that result in bent helix structures. Our studies show that simulated evolution of chirality with backbone flexibility can be a powerful tool in the design of novel heteropolymers with tuned stereochemical properties.

1. Introduction

In any molecular design endeavor, the first priority is to choose a target fold—the three-dimensional backbone topology—for ensuring structural stability and optimally presenting the side chains of functional interest. Advancements in rational and computational methods have made the design of novel backbone topologies possible.[1−3] This has been accomplished either through mathematical parametrization of backbone geometry, or through molecular mechanics minimizations of backbone coordinates while iteratively designing a sequence. A rapidly emerging field in design is the study of foldamers,[4,5] polymers which exhibit the protein-like behavior of naturally assuming stable configurations with secondary and tertiary structure. Synthetic methods allow us to design proteins that are not necessarily limited to the alphabet of the twenty naturally occurring amino acids. The design of novel topologies in polypeptides has been explored by changing the stereochemistry of substitution at the Cα position, resulting in new variations on helices and β-hairpin secondary structures.[6] Herein, we seek

to use computational tools currently applied to protein design to develop novel polypeptide topologies by varying monomer chirality.

The relationship between chirality and helical polypeptide structure was mentioned by Pauling in his seminal description of the α-helix[7] and solidified by pioneers in the field of protein structure such as Ramachandran[8,9] and others. L-amino acids, the ribosomally encoded stereoisomer, preferentially form right-handed helices. D-amino acids produce helices of the opposite sense. The effect of monomer chirality on macromolecular structure has been clearly demonstrated by the chemical synthesis of HIV-protease from D-amino acids, resulting in a molecule with mirror image CD signature and specificity for the stereochemical inverse of the natural substrate.[10] On the other hand, mixtures of L and D monomers can destabilize α-helical polypeptides[11−14] with the $\Delta\Delta G = 0.95$ kcal/mol for L-Ala to D-Ala[15].

These slight energetic effects can have a profound influence on macromolecular structure. In the obligate helix-forming polymers, polyisocyantes, helical screw sense is extremely sensitive to intramolecular or intermolecular chiral perturbations.[16] Surprisingly, even minute asymmetry such as a deu-

---

† Department of Biochemistry and Molecular Biophysics, University of Pennsylvania School of Medicine.

‡ Department of Chemistry, School of Arts and Sciences, University of Pennsylvania.

(1) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. *Science* **2003**, *302*, 1364−1368.
(2) Harbury, P. B.; Plecs, J. J.; Tidor, B.; Alber, T.; Kim, P. S. *Science* **1998**, *282*, 1462−1467.
(3) Walsh, S. T.; Cheng, H.; Bryson, J. W.; Roder, H.; DeGrado, W. F. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 5486-5491.
(4) Hill, D. J.; Mio, M. J.; Prince, R. B.; Hughes, T. S.; Moore, J. S. *Chem. Rev.* **2001**, *101*, 3893−4011.
(5) Nakano, T.; Okamoto, Y. *Chem. Rev.* **2001**, *101*, 4013−4038.
(6) Aravinda, S.; Shamala, N.; Roy, R. S.; Balaram, P. *P. Natl. Acad. Sci. U. S. A.* **2003**, *115*, 373−400.

(7) Pauling, L.; Corey, R. B.; Branson, H. R. *P. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 205−211.
(8) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95−99.
(9) Chandras. R.; Lakshmin. Av; Pandya, U. V.; Ramachan. Gn *Biochim. Biophys. Acta* **1973**, *303*, 14−27.
(10) Milton, R. C.; Milton, S. C.; Kent, S. B. *Science* **1992**, *256*, 1445−1448.
(11) Hermans, J.; Anderson, A. G.; Yun, R. H. *Biochemistry* **1992**, *31*, 5646−5653.
(12) Soares, T. A.; Lins, R. D.; Longo, R.; Garratt, R.; Ferreira, R. *Z. Naturforsch., C. Biosci.* **1997**, *52c*, 89−96.
(13) Wald, G. *Ann. N.Y. Acad. Sci.* **1957**, *69*, 352−368.
(14) Krause, E.; Bienert, M.; Schmieder, P.; Wenschuh, H. *J. Am. Chem. Soc.* **2000**, *122*, 4865−4870.

---

teration-induced stereocenter creates optical activity in otherwise racemic mixtures of helices.[17] A 51−49% mixture of monomer enantiomers result in helical phases with one-third the optical activity of single enantiomer preparations. This cooperative asymmetry is a result of the tight coupling between monomer chirality and secondary structure formation.[18] Theoretical models can explain the sharp cooperative sensitivity to enantiomeric excess, and frequency of helix sense reversals (left to right-handed helix switching in the middle of a chain) in long polymers.[17,19,20] These models draw extensively from polymer statistical mechanics, and are very similar to random field Ising models of helix-coil transitions in proteins.[21] Ising-type models have been effective in the study of polyisocyantes, due to their obligate helicity. These models are extended using atomistic simulations, allowing researchers to study energetics and dynamics of conformational transitions.[22,23]

Atomistic models have been applied to the study of homo-chiral and heterochiral polypeptides. Simulations of dipeptide stereoisomers helped establish the effect of side chain stereochemistry on accessible backbone topology.[8,9,24] Minimizations of D-Ala/L-Ala copolymers reveal many novel backbone topologies with repeating heterochiral sequences including the LD-ribbon (also known as a β-spiral) and the LD-helix (β-helix).[25−27] Some of these postulated structures are found in natural proteins such as the gramicidin β-helix[28−30] and the spider capture-silk β-spiral.[31] These methods are successful in predicting many periodic secondary structures of heterochiral polypeptides.

We extend the scope of previous studies relating chirality and secondary structure in polypeptides using simulated evolution,[32] a modeling framework that allows a polymer to concurrently sample multiple chain conformations and vary the chirality of an individual monomer stereocenter. A simple force field which emphasizes sterics and hydrogen bonding drives secondary structure formation. There is no bias on handedness in the force field, so it will not select between enantiomeric configurations. It will however, choose between diastereomeric molecules which have different computed energies. This now allows us to perform polypeptide simulation analogues of experiments done on polyisocynates. Majority-rule experiments can be simulated by weighting the L and D choice during chiral
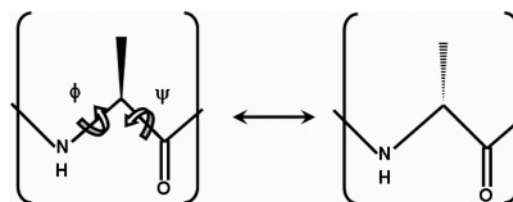
Figure 1. Allowed moves in simulated evolution of polyalanine include inversion of chirality and modification of backbone torsion angles.

mutagenesis, to model the sensitivity of helical screw sense to enantiomeric excess. We successfully model much of the natural physics of helical peptides, including the cooperativity of helix assembly and the generation of helix-capping structural motifs. Additionally, we discover novel super-secondary structures created by heterochiral heteropolymeric backbones, opening the doors for potential new design topologies.

2. Computational Methods

2.1 Simulated Evolution. Monte Carlo methods have been effectively applied to the modeling of protein[33] and heteropolymer conformation[34] and sequence.[35] Simple lattice models effectively reproduced much of the fundamental thermodynamics and kinetics of protein assembly.[36,37] Sequence design with an atomistic implementation, termed simulated evolution, was introduced as an alternative to rational protein design to allow "genetic selection by the computer"[32] for optimal sequences given a target protein backbone. It is a Metropolis-type Monte Carlo combined with Simulated Annealing optimization (MCSA)[38,39] that allows amino acid substitutions in a protein with a fixed backbone. Moves are accepted based on the Metropolis criteria[38] where the probability, a, of accepting a change is

$$a = \min \left(1, \exp(-T^{-1} \cdot (E_i - E_{i-1}))\right) \qquad (1)$$

where T is the selection temperature and E_i is the computed energy of step n as described below. T decreases linearly with iteration number, for a total of N iterations

$$T_n = T_0 - n \cdot \frac{T_0 - T_N}{N} \qquad (2)$$

Simulations begin with an all alanine chain of fixed length in the extended conformation ($\phi, \psi = 180.0\,°$) with each residue assigned an initial chirality at random. For each iteration of MCSA, a randomly chosen position epimerizes and/or alters ϕ, ψ angles (Figure 1). Subsequent Monte Carlo iterations allow for the following: (1) a change in the conformation of a given monomer (all ϕ/ψ's equally weighted), (2) epimerization of a given monomer, (3) both operations on the same monomer. All three options are equally probably for a given iteration. As stable conformations evolve, the stereochemistry of the sequence concomitantly varies to maximize conformational stability.

2.2 Potential Function and Model Set Up. Structures are scored simply using a function that penalizes for steric clashes and rewards for main chain hydrogen bonds. Qualitatively, this simulates peptide behavior in water at low temperature or in a helix promoting solvent.[40] Fitness, E_i, of a given conformation is scored by a simplified energy

(15) Fairman, R.; Anthony-Cahill, S. J.; DeGrado, W. F. *J. Am. Chem. Soc.* **1992**, *114*, 5458−5459.
(16) Green, M. M.; Peterson, N. C.; Sato, T.; Teramoto, A.; Cook, R.; Lifson, S. *Science* **1995**, *268*, 1860−1866.
(17) Lifson, S.; Andreola, C.; Peterson, N. C.; Green, M. M. *J. Am. Chem. Soc.* **1989**, *111*, 8850−8858.
(18) Green, M. M.; Garetz, B. A.; Munoz, B.; Chang, H. P. *J. Am. Chem. Soc.* **1995**, *117*, 4181−4182.
(19) Selinger, J. V.; Selinger, R. L. B. *Phys. Rev. Lett.* **1996**, *76*, 58−61.
(20) Teramoto, A. *Prog. Polym. Science* **2001**, *26*, 667−720.
(21) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526−535.
(22) Young, J. A.; Cook, R. C. *Macromolecules* **2001**, *34*, 3646−3653.
(23) Lifson, S.; Felder, C. E.; Green, M. M. *Macromolecules* **1992**, *25*, 4142−4148.
(24) Rizzo, V.; Lorenzi, G. P. *Macromolecules* **1983**, *16*, 476−482.
(25) Colonnacesari, F.; Premilat, S.; Heitz, F.; Spach, G.; Lotz, B. *Macromolecules* **1977**, *10*, 1284−1288.
(26) Heitz, F.; Detriche, G.; Vovelle, F.; Spach, G. *Macromolecules* **1981**, *14*, 47−50.
(27) Hesselin. Ft; Scheraga, H. A. *Macromolecules* **1972**, *5*, 455−462.
(28) Urry, D. W *P. Natl. Acad. Sci. U. S. A.* **1971**, *68*, 672−76.
(29) Veatch, W. R.; Fossel, E. T.; Blout, E. R. *Biochemistry-US* **1974**, *13*, 5249−5256.
(30) Bamberg, E.; Apell, H. J.; Alpes, H. *P. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 2402−2406.
(31) 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.
(32) Hellinga, H. W.; Richards, F. M. *P. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 5803−5807.

(33) Hansmann, U. H. E.; Okamoto, Y. *Curr. Opin. Struc. Biol.* **1999**, *9*, 177−183.
(34) Shakhnovich, E. I.; Gutin, A. M. *P. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 7195−7199.
(35) Shakhnovich, E. I.; Gutin, A. M. *Protein Eng.* **1993**, *6*, 793−800.
(36) Shakhnovich, E. I.; Gutin, A. M. *Biophys. Chem.* **1989**, *34*, 187−199.
(37) Shakhnovich, E. I.; Gutin, A. M. *Nature* **1990**, *346*, 773−775.
(38) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1951**, *21*, 1087−1092.
(39) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.
(40) Smythe, M. L.; Huston, S. E.; Marshall, G. R. *J. Am. Chem. Soc.* **1995**, *117*, 5445−5452.
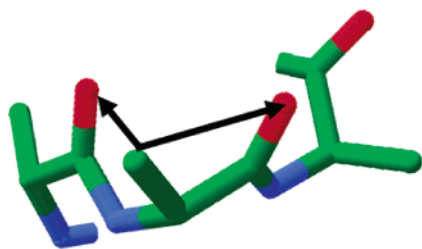
Figure 2. Steric clash of D-alanine Cβ with adjacent carbonyls in an α_R conformation.

function

$$E_i = E_{VDW} + \epsilon_{clash} \cdot N_{clash} + \epsilon_{HB} \cdot N_{HB} \qquad (3)$$

N_{HB} is the number of amide-to-carbonyl hydrogen bonds in the backbone and N_{clash} is the number of Cβ-to-carbonyl clashes as defined below. ϵ_{clash} and ϵ_{HB} are the energies per clash and hydrogen bond, respectively. ϵ_{clash} is set to 5.0 kcal/mol and ϵ_{HB} is -5.0 kcal/mol. The van der Waals energy, E_{VDW}, is modeled with a 12−6 Lennard-Jones potential and distance cutoff of 4.0 Å. The short cutoff distance is implemented to ensure self-avoiding chains. To facilitate searching sensitivity, atomic radii are scaled down by 10%,[41] and for separations less than the ideal interaction distance, a linear repulsion term with a maximum of 10.0 kcal is introduced.[42]

$$E_{VDW}(i,j) = \epsilon_0 \cdot \left[\left(\frac{\alpha \cdot r_0}{r_{ij}} \right)^{12} - 2 \cdot \left(\frac{\alpha \cdot r_0}{r_{ij}} \right)^6 \right] \text{if } \frac{r_0}{r_{ij}} \leq 1.12 \qquad (4)$$

$$E_{VDW}(i,j) = 10.0 - 11.2 \cdot \left(\frac{r_{ij}}{\alpha \cdot r_0} \right) \text{if } \frac{r_0}{r_{ij}} > 1.12$$

where ϵ_0 is the energy of optimal interaction representing the geometric mean of the well depths for atom types i and j. Optimal interaction radius r_0 is the mean of the atomic radii of atoms types i and j which are scaled by $\alpha = 0.9$. United-atom parameters for atomic radii and vdW association energy are adapted from the AMBER force field.[43]

Other than explicit hydrogen bonds, electrostatic interactions are not included. Hydrogen bonds are satisfied when the backbone nitrogen to backbone carbonyl oxygen distance is less than 3.2 Å and an N−H−O angle greater than 90°. 1,4-atomic interactions are excluded in the vdW potential, so it is necessary to include a repulsive term for backbone torsions. The major effect of the stereocenter on backbone conformation is the Cβ interaction with the carbonyl oxygen of its own, and the previous residue. A clash between Cβ and backbone oxygen of i,i or i,i-1 residues is counted if the atoms were less than 3.0 Å apart (Figure 2). For purposes of post-simulation data analysis, when deciding between hydrogen bond partners, an energy based selection was used with a 12−10 interaction potential and a cosine angular dependence term[44] as appropriate for an sp2-sp2 interaction.

3. Results

3.1 Helix Structure, Cooperativity, and Homochirality.
The structures generated by simulated evolution of polyalanine are predominantly helical, distributed equally between molecules of opposing helical sense and monomer chirality. The helices fall within α_L and α_R areas of backbone torsion space (Figure 3A). Clustering of the structures based on ϕ/ψ torsion RMSD produces two major structure families corresponding to left and

right-handed helices.[45] An analysis of hydrogen bonding patterns shows that most helical residues participate in *i,i+4* hydrogen bond interactions, characteristic of α-helices. Additionally, a significant subpopulation of 3_{10}-helices are found with *i,i+3* hydrogen bonding interactions.[46,47] *i,i+3* interactions are found most frequently on the C-terminal end of the peptide and next most frequently toward the N-terminus of the peptide.[47] For the center of a fifteen-residue chain, the ratio of α-helix to 3_{10}-helix content is approximately 3:2 (Figure 3B). The relative amount of 3_{10}-helix content decreases as the peptide lengthens. The ratio of α-helix to 3_{10}-helix content in the 11-mer is near 1:1 and the 19-mer approxmiately 2:1. These observations are consistent with studies of polyalanine in solution and in high-resolution all-atom molecular dynamics, which show a propensity for 3_{10}-helix in shorter peptides,[40,48−50] and a higher tendency of 3_{10} helix at the C-terminus of a chain.[51]

Validation of the behavior of the minimal atomistic model is the cooperative relationship between helicity and peptide length, which has been clearly demonstrated both theoretically and experimentally.[21,52−57] The entropic barrier of accumulating four successive residues in a left or right α-helical conformation must be overcome before helix propagation can proceed efficiently. One thousand simulations are conducted for peptides of length seven, eleven, fifteen and nineteen amino acids. For peptides of sufficient length, the propensity for contiguous helix is significantly above random expectation (Figure 4A). As secondary structure influences stereochemistry, we also observe significant sequence homochirality of longer peptides (Figure 4B, Tables 1−4). Hydrogen bonding drives helix formation, which then imposes a steric constraint on the sequence.

Seven residue peptides are not of sufficient length to form stable helices, which accounts for their lower observed enhancement of helicity and homochirality. For peptides of length eleven, there are 2048 (2^{11}) possible sequences. However, in one thousand simulations, we only observe 371 unique sequences, again with the population skewed strongly toward homochiral peptides (Table 2). The most frequent sequences D(L)$_8$DD, (L)$_9$DD (L = L-alanine, D = D-alanine) and their enantiomers occur approximately one 100-fold more frequently than randomly expected. For fifteen-mers, the enhancement of D(L)$_{12}$DD and (L)$_{13}$DD and their respective enantiomers occur one thousand times more frequently than random. The effect is most dramatic with homochiral sequences in the 19-residue peptide occurring 7×10^3 more frequently than random expectation. For each additional turn of the helix, there is an approximate order of magnitude enhancement in homochirality.

(41) Dahiyat, B. I.; Mayo, S. L. *P. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10172−10177.
(42) Kuhlman, B.; Baker, D. *P Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10383−10388.
(43) 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.
(44) Gordon, D. B.; Marshall, S. A.; Mayo, S. L. *Curr. Opin. Struc. Biol.* **1999**, *9*, 509−513.
(45) Clustering was performed using *kclust* (http://mmstb.scripps.edu) with a mean angular deviation cutoff of 60° within structure families.
(46) Karle, I. L.; Balaram, P. *Biochemistry-Us* **1990**, *29*, 6747−6756.
(47) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1988**, *201*, 601−619.
(48) Takano, M.; Yamato, T.; Higo, J.; Suyama, A.; Nagayama, K. *J. Am. Chem. Soc.* **1999**, *121*, 605−612.
(49) Miick, S. M.; Martinez, G. V.; Fiori, W. R.; Todd, A. P.; Millhauser, G. L. *Nature* **1992**, *359*, 653−655.
(50) Armen, R.; Alonso, D. O. V.; Daggett, V. *Protein Sci.* **2003**, *12*, 1145−1157.
(51) Karpen, M. E.; Dehaseth, P. L.; Neet, K. E. *Protein Sci.* **1992**, *1*, 1333−1342.
(52) Goodman, M.; Verdini, A. S.; Toniolo, C.; Phillips, W. D.; Bovey, F. A. *P. Natl. Acad. Sci. U. S. A.* **1969**, *64*, 444−450.
(53) Lifson, S.; Roig, A. *J. Chem. Phys.* **1961**, *34*, 1963−1974.
(54) Brooks, C. L. *J. Phys. Chem-Us* **1996**, *100*, 2546−2549.
(55) Young, W. S.; Brooks, C. L. *J. Mol. Biol.* **1996**, *259*, 560−572.
(56) Rohl, C. A.; Scholtz, J. M.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Biochemistry-Us* **1992**, *31*, 1263−1269.
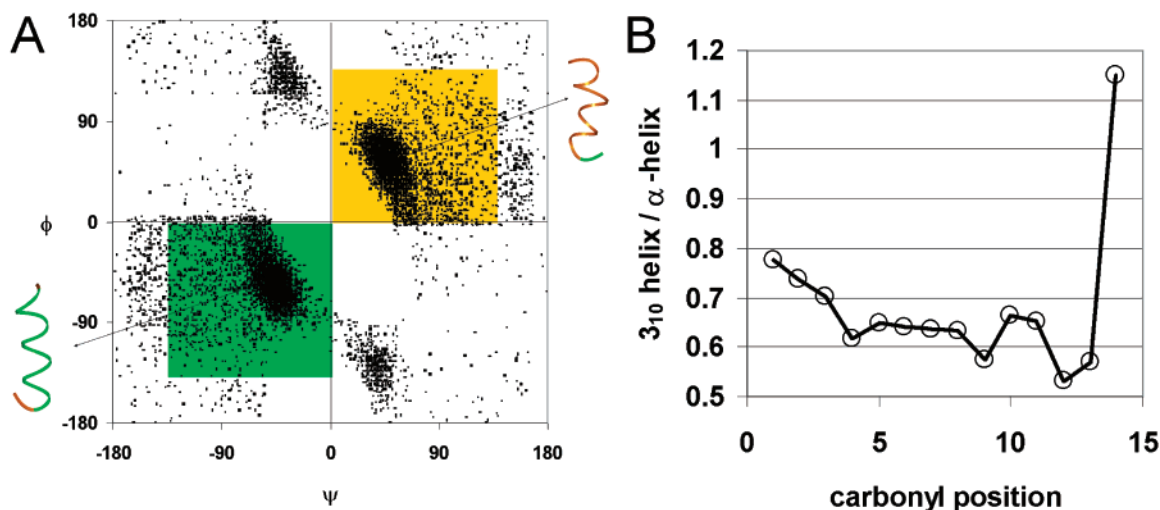(57) Daggett, V.; Kollman, P. A.; Kuntz, I. D. *Biopolymers* **1991**, *31*, 1115−1134.

Figure 3. (A) Ramachandran plot of the same data set showing strong clustering in the α_L and α_R regions of ϕ/ψ space. Two major structures from clustering of 1000 simulations of the fifteen-mer show α_L (orange) and α_R (green) helices. (B) Using hydrogen bonding pattern as an index of conformation, the ratio of 3_{10} to α-helix is shown as a function of position. Statistics are from 1000 simulations of the nineteen-mer.
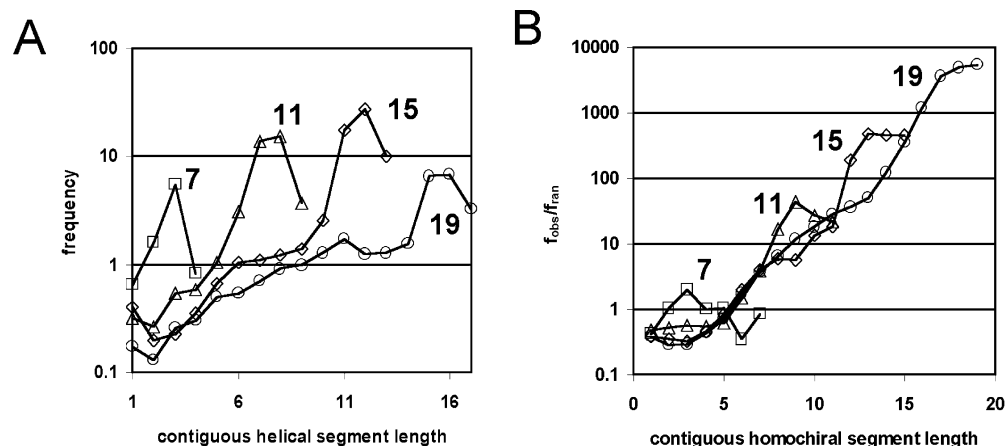
Figure 4. (A) Frequency of contiguous α_L and α_R segments in peptides of varying lengths (square, 7-mer; triangle, 11-mer; diamond, 15-mer; circle, 19-mer) normalized to the probability of independently occurring contiguous segments of the same length where α_R is defined as $0 > \phi > -120$, $0 > \psi > -120$ and α_L is defined as $0 < \phi < 120$, $0 < \psi < 120$. (B) Enhancement in contiguous stretches of homochiral sequence for the same set of simulations normalized to explicit enumeration of all possible sequences for an L-mer (2^L sequences).

3.2 Turn Motifs in Seven-mer Simulations. The seven-residue peptides are of insufficient length to form significant left or right-handed α-helix. The peak of contiguous helix length observed is three residues (Figure 4A), which is not enough to form a hydrogen bonded cycle. Nevertheless, the sequence distribution is not random: with seven positions, there are 128 (2^7) possible sequences, but the actual sequence distribution is quite skewed (Table 1), with LLDDDLL and DDLLLDD accounting for one-quarter of the entire population. This is 15-fold more frequent than randomly expected. These form a stable structural motif: α_R-α_R-α_R-α_L and α_L-α_L-α_L-α_R corresponding to the central residues of DD<u>LLL</u>DD and the LL<u>DDD</u>LL, respectively. The second most populous sequence, D<u>LLL</u>DDD and its enantiomer also primarily adopt the α_R-α_R-α_R-α_L and α_L-α_L-α_L-α_R conformation, with the structural motif shifted back one in sequence. The α_R-α_R-α_R-α_L motif occurs naturally in proteins. It is commonly known as a π-turn and is the most frequent motif found in four residue β-hairpin turns.[58–61] A number of two residue turns are observed and will be discussed later.

3.3 Helix Capping. The most frequently occurring sequences for the longer peptide simulations all contain a sequence reversal at the second to last position from the peptide C-terminus

(Tables 2–4). In all of these peptides, the fully homochiral sequences occur less frequently than peptides with a C-terminal reversal. This is also found in the plots of contiguous homochiral sequence frequency (Figure 4B) which plateau at lengths of 9, 13, and 17 for the 11-, 15-, and 19-mer simulations, respectively. Inspection of these molecules reveals that a change in chirality at the second to last position is accompanied by a reversal of the helix sense. For (L)$_n$DD, the backbone conformation is $(\alpha_R)_n$-α_L-x. This is the equivalent of a canonical Schellman α_L C-capping motif.[62] In natural proteins, this position in an α-helix is often occupied by glycine, which readily assumes the α_L conformation.[63,64] With hydrogen bonding as the primary driving force for structure, this interaction serves to maximize the number of interactions through bifurcated hydrogen bonds at

(58) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989**, *206*, 759–777.
(59) Sibanda, B. L.; Thornton, J. M. *Nature* **1985**, *316*, 170–174.
(60) Rajashankar, K. R.; Ramakumar, S. *Protein Sci.* **1995**, *5*, 932–946.
(61) Gunasekaran, K.; Ramakrishnan, C.; Balaram, P. *Protein Eng.* **1997**, *10*, 1131–1141.
(62) Schellman, C. In *Protein Folding*; Jaenicke, R., Ed.; Elsevier/North-Holland: New York, 1980; pp 53–61.
(63) Presta, L. G.; Rose, G. D. *Science* **1988**, *240*, 1632–1641.
(64) Richardson, J. S.; Richardson, D. C. *Science* **1988**, *240*, 1648–1652.

Table 1. Most Frequent 7-mer Sequences out of 1000 Simulations[a]

LLDDDLL	136	(110)
DLLLDDD	64	(59)
LLDDDLD	42	(32)
LLLLDDD	33	(32)
LDDDLLD	27	(24)
LDDDDLL	27	(26)
LLLLLDD	21	(15)
LDLLLDD	19	(12)
LLDDDDL	18	(12)
LLLDLLL	16	(11)
DDDLLDD	14	(12)
DDLLLLL	14	(10)
DDLLDDD	12	(12)
LDDDDLD	12	(8)
DLDDDLD	10	(6)
LDDLLDD	8	(3)
LLLLLLL	8	(5)
DDLLDDL	7	(4)
DLLDLLL	6	(3)
DLDDDDD	5	(1)

[a] Number in parentheses represents counts of sequence stereoisomers. A total of 91 unique sequences were found.

Table 2. Most Frequent 11-mer Sequence Counts out of 1000 Simulations[a]

DLLLLLLLLDD	71	(63)
LLLLLLLLLDD	70	(51)
LLLLLLLLLDL	22	(23)
DLLLLLLLLDL	22	(15)
DLLLLLLLLLD	18	(8)
LLLLLLLLLLD	8	(18)
DLLLLLLLLLL	14	(14)
LDLLLLLLLDD	5	(14)
LLLLLLLLLLL	8	(13)
DDLLLLLLLLL	2	(7)
DDLLDLLLDDD	7	(4)
LLDDDLDDDLL	5	(1)
DLLLLLLLDDD	5	(5)
DDLLLLLLLDD	5	(5)
DLLLLLLLDDL	4	(5)
LLLLLLDDDDD	4	(3)
DDLLLLLLDDD	2	(4)
LDDDLDDDLLD	4	(2)
LLLLLLLLDDD	4	(4)
DLLLLLLLDLD	1	(4)

[a] Number in parentheses represents counts of sequence stereoisomers. A total of 371 unique sequences were generated.

Table 3. Most Frequent 15-mer Sequences from 1000 Simulations[a]

DLLLLLLLLLLLLDD	49	(41)
LLLLLLLLLLLLLDD	42	(43)
LLLLLLLLLLLLLDL	20	(11)
LLLLLLLLLLLLLLD	15	(14)
LLLLLLLLLLLLLLL	19	(9)
DLLLLLLLLLLLLLL	11	(15)
DLLLLLLLLLLLLLD	14	(12)
DLLLLLLLLLLLLDL	8	(12)
LLLLLLLDDDDDDLL	11	(5)
DLLLLLLLDDDDDLD	5	(5)
LLLLLLLLDDDDDLL	5	(5)
DLLLLLLLDDDDDLL	4	(6)
DLLLLLLDDDDDDLL	6	(3)
DLLLLLDDDDDDDLL	4	(5)
LLLLLLDDDDDDDLL	4	(5)
LLLLLDDDDDDDDLL	7	(1)
DLLLLDDDDDDDDLL	2	(5)
LLLLLLLDDDDDDLD	3	(4)
DLLLLLLLLLLLDDD	4	(2)
DDLLLLLLLLLLLLD	2	(4)
LLLLLLDDDDDDDDL	4	(2)
LLLLLLLLLLLLDLD	2	(4)
LLLLLLLLLLLLDLD	2	(4)
LLLLLLLLLLDDDDD	4	(1)
LLLLLLLLDDDDDLD	4	(1)

[a] Number in parentheses represents counts of sequence stereoisomers. A total of 496 unique sequences were found.

**Table 4.** Most Frequently Occurring 19-mer Sequences Found during 1000 Simulations[a]

LLLLLLLLLLLLLLLLLDD	30	(20)
DLLLLLLLLLLLLLLLLDD	17	(17)
LLLLLLLLLLLLLLLLLLL	14	(10)
LLLLLLLLLLLLLLLLLDL	15	(5)
LLLLLLLLLLLLLLLLLLD	7	(12)
DLLLLLLLLLLLLLLLLLD	9	(6)
DLLLLLLLLLLLLLLLLLL	9	(4)
DLLLLLLLLLLLLLLLLDL	5	(5)
LLLLLLLLLLDDDDDDDLL	4	(5)
DLLLLLLLLLLLLDDDDLL	2	(5)
DLLLLLLLLLLLLDDDDLL	2	(5)
DLLLLLLLLLLDDDDDDLL	3	(4)
LLLLLLLLDDDDDDDDDLL	3	(4)
DLLLLLLDDDDDDDDDDLL	4	(3)
DLLLLLLLDDDDDDDDDLD	1	(5)
LLLLLLLLLLLLLLLLDDD	4	(2)
LLLLLLLLLDDDDDDDDLL	3	(3)
LLLLLLDDDDDDDDDDDLD	0	(5)
DLLLLLLDDDDDDDDDDDD	2	(3)
DLLLLLLLLLDDDDDDDLL	4	(0)
LLLLLLLLLLLLDDDDDLL	0	(4)
LDDDDDLLLLLLLLLLLDL	3	(0)
DLLLLLLDDDDDDDDDDDD	2	(3)

[a] Number in parentheses represents counts of sequence stereoisomers. A total of 690 unique sequences were generated.

the last and second to last position (Figure 5). The π-turn observed in the seven-mer simulations is another example of this motif where the capping interactions are the dominant source of stability in the absence of a stable helix.

We furthermore observe that the end residue preferentially assumes the same chirality as the penultimate position, although this correlation is less stringent. (D)$_n$LD and (L)$_n$DL are also frequently observed. In naturally occurring C-terminal helix capping motifs, this residue would often adopt a β_R conformation.[65] However, in our simulations, the final residue does not have a ψ torsion that affects side chain sterics, making the energetic distinction between β_R and β_L smaller. Likewise, the N-terminal position chirality is not correlated with the rest of

the sequence. Since this position is lacking a ϕ torsion that has steric consequences, we suggest that it is able to accommodate a sequence reversal without energetic penalty by adjusting its backbone conformation.

3.4 Helix Reversals Result in Kinked Structures. In the 15 and 19-mer sequence statistics (Tables 3 and 4), there exists a large population of molecules where the sequence reverses midway through the peptide. Conformational characterization of these peptides shows that the conformation of the backbone

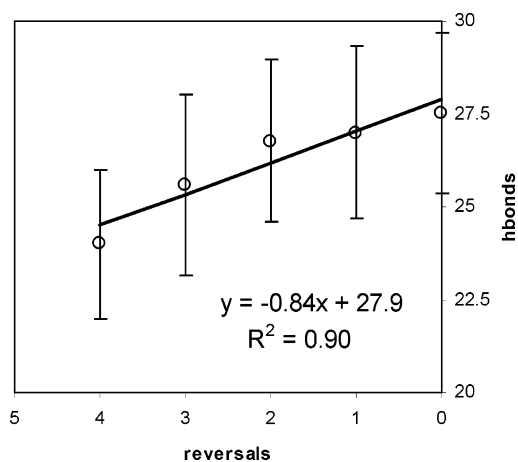(65) Aurora, R.; Rose, G. D. *Protein. Sci.* **1998**, *7*, 21−38.
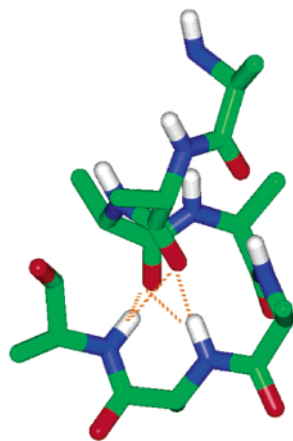
Figure 5. Example of a Schellman C-capping motif, shown here as the last seven residues of a right-handed 15-mer. Positions 14 and 15 are both L-Ala, participating in bivalent hydrogen bonds with the carbonyl acceptors.
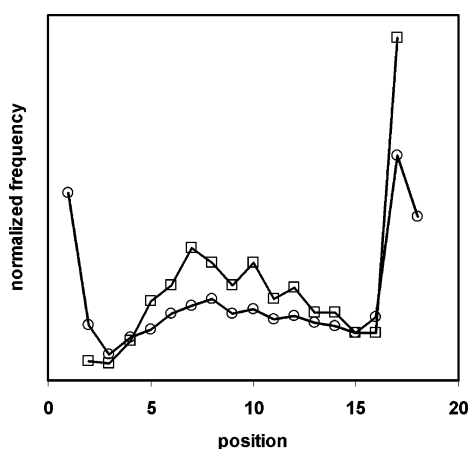
Figure 6. Sequence and conformation reversal histogram for the 19-mer simulations normalized to the total number of reversals. A sequence reversal event (circle) is counted upon a change in chirality between the i and $i+1$ position. A conformation reversal (square) reflects a helix sense change between i and $i+1$ position.


Figure 7. Mean number of hydrogen bonds plotted against the mean number of reversals in the 19-mer simulations (excluding reversals at positions 1, 2, 18, and 19) shows that on average one hydrogen bond is lost per reversal event. No structures accommodated more than four reversals.

changes helical sense at the boundary of sequence chirality change. That is, a peptide with the sequence (L)$_m$(D)$_n$ has the conformation (α_R)$_m$(α_L)$_n$. Figure 6 shows the frequency of sequence and helix reversals in the population of one thousand 19-mer sequences. Reversals occur toward the middle of the molecule (excluding N- and C-terminus reversals which occur for reasons already described). Helix reversals cause a bend between the joined left and right-handed helices of about 120°. An average of one hydrogen bond is lost for each internal sense reversal (Figure 7), suggesting that kinked heterochiral motifs might be stable enough to have structural rigidity.

We suggest two possible origins of these helix reversals during simulated evolution. One is the existence of more than one concurrent helix nucleation event during initial chain collapse from the extended conformation. The chance of two helix nucleation events being the same chirality is 50%. While contiguous helices of one sense are the most stable motif, kinked helices with a sense reversal are favorable enough to be an energetic local minimum trap during minimization. In longer simulations which we have conducted on 30-, 50-, and 100-mer peptides, the potential for multiple simultaneous helix-nucleation events increases the frequency of observed kinked helices (Figure 8A). In five simulations of 50-mer polyalanine,

no homochiral, single helix structures are generated. Another possible origin of helix reversals is nucleation by an internally generated Schellman-like reversal on a growing helix. The propensity for the final two positions on a helix to be opposite in chirality could serve to initiate a new helix of the reverse sense. In an example of a bent helix, we find the first residue of one helix also serves as the C-cap of the previous helix (Figure 8B).

3.5 Simulating Enantiomeric Excess. Earlier, we demonstrated the cooperative relationship between helicity and peptide length and its corresponding effect on sequence homochirality. We can model this cooperativity in the presence of simulated enantiomeric excess of L or D-alanine monomer by biasing the probability of epimerizing D to L vs L to D. This is the similar to "majority rule" experiments performed on polyisocyanates[18,19] without the quenching of sequence space. The nonlinear effect of enantiomeric excess on polyisocyanate helicity reflects the cooperativity of secondary structure formation. Since we have observed homochiral polypeptide helices as the lowest energy state in prior simulations, we will expect a similar dependence on both sequence and structure if the probability of choosing D-Ala versus L-Ala is varied to emulate the effects of enantiomeric excess (Figure 9). The unitless "optical activity", based on helicity of the polymers, is maximal at 10% e.e. of L-Ala or D-Ala. Even at 1% e.e., there is detectable helical excess in the simulated population. The steepness of the simulated nonlinear effect is not as sharp for polypeptides as it is for polyisocyanates, possibly due to the competing effects of nonhelical conformations and the low energy penalty for helix reversals.

3.6 Nonhelical Heterochiral Motifs. The simulations have a strong bias toward generating homochiral, helical structures—presumably due to high count of hydrogen bonding interactions in a helix and the cooperative nature of helix assembly. The overall paucity of sheet structures may also be due to the importance of side chain effects on backbone conformation.[66] However, other regular structures such as turns and sheets have extensive hydrogen bonding networks and cooperative folding reactions. In addition to the numerous variations on helical

(66) Chou, K. C.; Scheraga, H. A. *P. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 7047−7051.
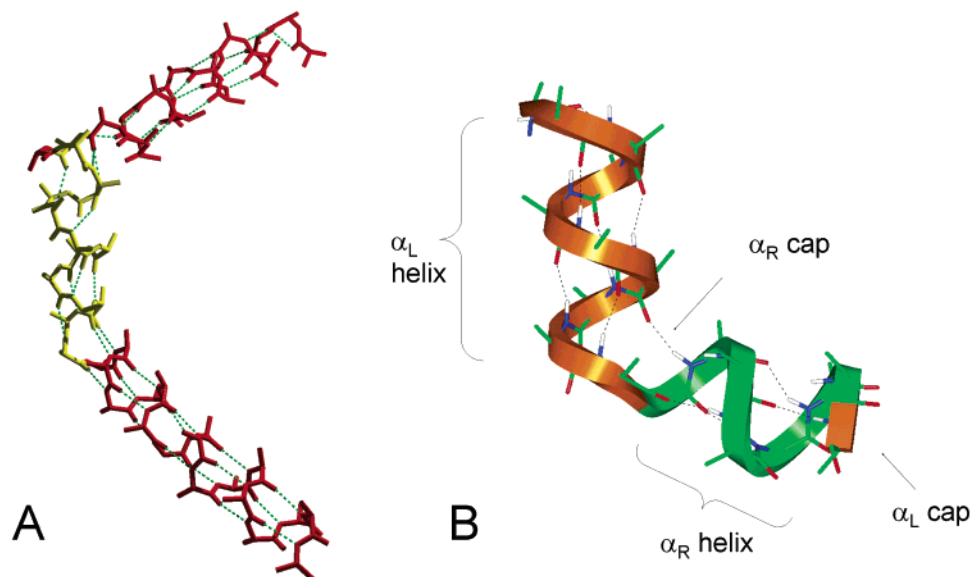
Figure 8. (A) Helix reversals in the 50-mer simulation. Red is D-Ala, yellow is L-Ala. Backbone hydrogen bonds are highlighted in green. (B) Helix reversals include a C-cap at the interface, suggesting a role in nucleating the opposing helix.
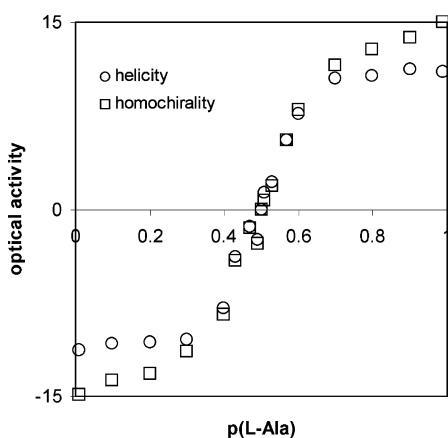
Figure 9. Effect of enantiomeric excess on helicity (circle) and homo-chirality (square) shown by plotting optical activity ($N(\alpha_R) - N(\alpha_L)$) for helicity and $N(L) - N(D)$ for homochirality) as a function of the weight $p(L\text{-Ala})$ of choosing to epimerize L-Ala to D-Ala. Values represent mean over 25 simulations for each value of $p(L\text{-Ala})$.

motifs observed in our simulations, a number of heterochiral, nonhelical structures are found. Each of these structures is infrequent and in most cases, unique. Several examples of nonhelical motifs are shown in Figure 10.

Most of the nonhelical motifs are two-residue β-turns. In the seven-mer simulations, a large fraction of interactions which do not fall into the π-turn category are $i,i+2$ with type I, II, I′ or II′ turns. Figure 10A shows a structure with two turns at opposite ends of the molecule. One is a Type II′ turn (residues 3−6) and the other a Type I′ turn (residues 11−14). Additional stabilization is achieved by propagating β-hairpins out from these turns. A more extensive hairpin is shown in Figure 10B. Alternating D and L-Ala residues form layers of homochiral hydrogen bonding interactions across the two strands. A super-secondary motif consisting of both a β-hairpin and right-handed α-helix is shown in Figure 10C. It satisfies hydrogen bonds by bringing the N and C terminus into close proximity. Other novel secondary structures such as the ones exhibited in Figure 10D and E show the potential for uncovering significantly new topologies. The first is a stack of two turns of opposite chirality.
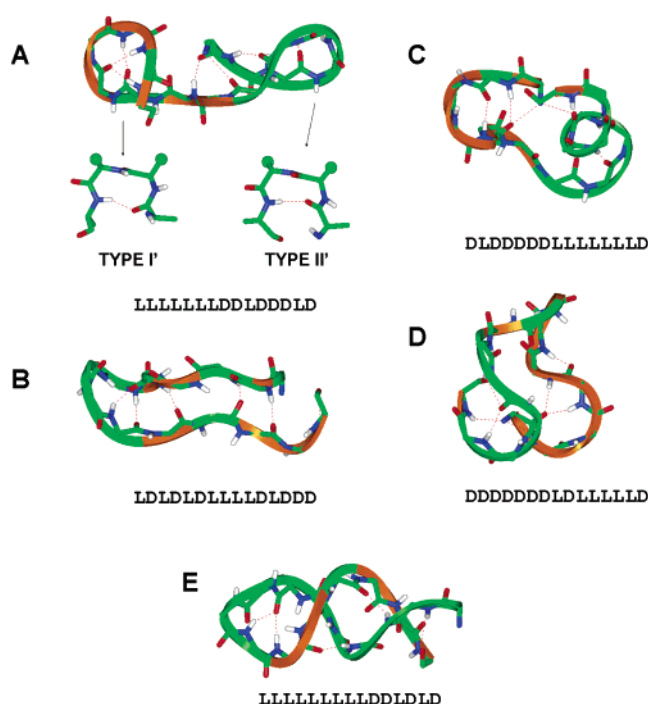
Figure 10. Examples of nonhelical motifs from heterochiral 15-mers. Green and orange ribbons correspond to L- and D-Ala, respectively. Backbone hydrogen bonds are highlighted in red.

The second is a two-stranded helix stabilized by both intra- and interstrand hydrogen bonds. Although these topologies do not represent global minima for the sequences of which they are composed, they do provide templates with extensive hydrogen bonding networks that can be augmented with side chains to achieve specificity and additional stability.

Furthermore, regular patterns of L-Ala and D-Ala can adopt interesting regular conformations.[67] In particular, alternating patterns of D and L-Ala, $(LD)_n$ were shown to theoretically adopt two conformations: the LD-ribbon, which consists of sequential two residue turns where each monomer is involved in an $i,i+2$

(67) Hesselink, F. T.; Scheraga, H. A. *Macromolecules* **1972**, *5*, 455−463.

hydrogen bond, and the LD-helix which is stabilized by $i,i+5$ interactions.[68] Both the LD-ribbon and helix exist in natural proteins, as in the examples mentioned earlier: spider silk (β-spiral) and gramicidin (β-helix). Despite the extensive hydrogen bonding pattern of the LD-ribbon and helix, no examples of pure alternating sequences are found in our simulations of the different length peptides. The lack of LD-ribbon and helix may be due to the choice of alanine as our side chain in these simulations. In gramicidin, a number of the side chains are large or β-branched (such as valine), which promote sheet topologies in many contexts.

4. Discussion

The primary forces that drive secondary structure are steric constraints on backbone flexibility and enthalpic gain of hydrogen bonding.[69] In modeling a flexible polyalanine chain, we focus on these two interactions, with encouraging results. Soft core sterics are implemented to permit a generally self-avoiding chain to pass through high energy intermediates in the search of minima. Combined with a simple distance/angle-cutoff hydrogen bonding potential, this force field gives helical structures that are quite realistic. Our approach is a compromise between full atom, full force field simulations and simple lattice model approaches. Fundamental physical properties of proteins can be modeled by minimal models of structure.[37,70,71] More sophisticated terms could easily be substituted in these studies to improve the accuracy of model generation. The advantage of the simplified force field lies in its reduced computation time, allowing us to quickly sample many states during the Monte Carlo simulation, while still capturing much of the essential physics of protein structure. In design applications, a coarse-grained topology search step would be useful prior to higher resolution backbone optimization and sequence design. For example, iterative refinement of the backbone scaffold is effectively used by the Baker group in the design of a protein with a novel topology.[1] Furthermore, in the design of a non-peptide foldamers and heteropolymers, where an extensive structural database to benchmark force fields is not available, a simple force field can permit the rapid modeling of basic secondary and super-secondary structures with minimal assumptions. A more sophisticated force field that includes long-range interactions could allow us to start evolving helical tertiary structures and β-sheet secondary conformations. Clearly, these are periodic structures where backbone stereochemistry also plays an important role.

The structures generated by simulated evolution show promising correlations with experimental studies done in other laboratories. We find that the amount of 3_{10}-helix versus α-helix character depends both on peptide length and on residue position, as has previously been shown in experimental studies and database surveys.[47,51] Our models are consistent with structural studies on model heterochiral peptides such as those studied by the Balaram group.[6] Polypeptide chain reversals are seen at the C-terminus of model peptides where a D-Ala makes a Schellman-like C-cap on a right-handed helix.[72,73] D-Arg has
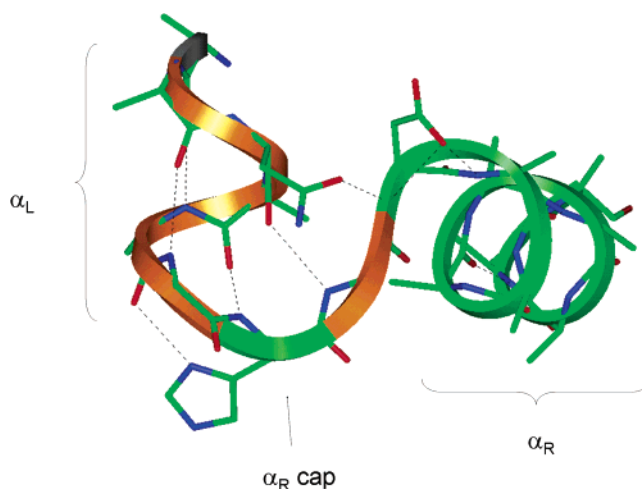
Figure 11. Helix reversal in alanine racemase (1SFT), residues 38–56. Four residues are in a contiguous α_L helix followed by a reverse-Schellman C-cap and finally an α_R helix. α_R residues shown in green and α_L residues in orange.

also been show to confer 1.3 kcal/mol stability when appended to the C-terminus of a right-handed helix.[74] It was also noted in studies of poly-aminoisobutyric acid (Aib) that the addition of a C-terminal L-leucine induces a left-handed sense in the remaining helix.[75] Poly-Aib is normally in thermal equilibrium between left and right-handed forms of a 3_{10}-helix due to the achiral nature of the monomer.[76] This suggests that a Schellman-motif can drive helix sense conformation in obligate helix molecules. However, a chiral residue at the second from N-terminus position of an otherwise achiral polypeptide has no clear influence on helix sense, and the effect varies depending on the solvent used.[77] This is consistent with our observation that chirality of the N-terminal position does not correlate simply with that of the handedness of chirality of the remaining sequence.

The propagation of the C-cap as a helix reversal in longer polyalanine simulations with minimal energetic cost suggests that this may be an exploitable super-secondary motif, a helix kink. The nature of helix reversals involving a change in chirality has been studied in model peptides.[78,79] A small number of examples of helix-reversals are found in natural proteins by searching the Protein Data Bank for contiguous α_L stretches of residues using WHATIF.[80] The structure of alanine racemase[81] contains a mirror-Schellman C-capping interaction which then leads into a right-handed α-helix (Figure 11). These α_L segments are additionally stabilized by extensive networks of side chain-backbone and side chain-side chain hydrogen bonds. This gives some indication of how side chains can be used to stabilize

(68) Chandrasekaran, R.; Ramachandran, G. N. In *2nd American Peptide Symposium*, First ed.; Lande, S., Ed.; Gordon and Breach: Cleveland, OH, 1970; Vol. 2, pp 195–215.
(69) Dill, K. A. *Biochemistry-Us* **1990**, *29*, 7133–7155.
(70) Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *97*, 9412–9426.
(71) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *P. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 12565–12570.

(72) Aravinda, S.; Shamala, N.; Pramanik, A.; Das, C.; Balaram, P. *Biochem. Biophys. Res. Comm.* **2000**, *273*, 933–936.
(73) Aravinda, S.; Shamala, N.; Bandyopadhyay, A.; Balaram, P. *J. Am. Chem. Soc.* **2003**, *125*, 15065–15075.
(74) Schneider, J. P.; DeGrado, W. F. *J. Am. Chem. Soc.* **1998**, *120*, 2764–2767.
(75) Inai, Y.; Ishida, Y.; Tagawa, K.; Takasu, A.; Hirabayashi, T. *J. Am. Chem. Soc.* **2002**, *124*, 2466–2473.
(76) Shamala, N.; Nagaraj, R.; Balaram, P. *J. Chem. Soc. Chem. Comm.* **1978**, 996–997.
(77) Inai, Y.; Kurokuwa, Y.; Kojima, N. *J. Chem. Soc., Perkin Transactions 2: Phys. Org. Chem.* **2002**, 1850–1857.
(78) Karle, I. L. *Biopolymers* **2001**, *60*, 351–365.
(79) Karle, I. L.; Banerjee, A.; Balaram, P. *Folding & Design* **1997**, *2*, 203-210.
(80) Vriend, G. *J. Mol. Graph.* **1990**, *8*, 52–56.
(81) Shaw, J. P.; Petsko, G. A.; Ringe, D. *Biochemistry* **1997**, *36*, 1329–1342.

unusual topologies. While contiguous α_L segments are energetically unfavorable in natural proteins, we can synthesize heterochiral polypeptides where sequence reversals stabilize helical kinks for building novel backbone topologies. Certain bacterial antimicrobial peptides already take advantage of this. The cytolytic peptide, tolaasin contains both L- and D-amino acids, which generate a kinked helix much like the structures we model.[82] By adding more amino acid types to a simulated evolution experiment, it should be possible to stabilize helix kinks and modulate their geometry with side chain − backbone and side chain−side chain interactions.

The strong bias toward helical motifs may be due both to approximations of the force-field, the conformational sampling approach (for example, implementing a local moves protocol may favor long-range interactions by minimizing the perturbation to a local group of residues[83]). Another issue is the choice of side chain. The Cβ atom of alanine is able to enforce some degree of backbone steering. However, other amino acids are more likely to encourage sheet structure.[66] Increasing the library of side chain options may allow us to explore other classes of backbone topologies. Nevertheless, the unusual topologies found in this study are useful as starting points for further design. The algorithm maximizes the number of backbone hydrogen bonds, while minimizing steric clash. These characteristics provide a good foundation for adding stability and specificity through side chain selection. The antiparallel double stranded peptide helix shown in Figure 10E has already been found in alternating D/L norleucine peptides.[84]

5. Conclusions

This study demonstrates the utility of coarse grained backbone topology generating algorithms that are driven by simple force fields in for simulating helix conformation behavior and for generating novel backbone templates. In future work, we intend to apply this model to the design of tertiary structures with heterochiral backbones. An advantage of coarse grained simulations is their application to the rapid modeling of larger systems, such as supramolecular assemblies and chiral product sensors. We hope that simulated evolution protocols such as the one presented here will be very useful in adapting lessons learned from protein design into the larger field of foldamer engineering.

(82) 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.
(83) Elofsson, A.; Legrand, S. M.; Eisenberg, D. *Proteins, Struct. Funct. Genet.* **1995**, *23*, 73−82.
(84) Navarro, E.; Fenude, E.; Celda, B. *Biopolymers* **2002**, *64*, 198−209.