J|A|C|S
ARTICLES
Published on Web 11/09/2006

Energetic Determinants of Oligomeric State Specificity in Coiled Coils

Jorge Ramos and Themis Lazaridis*

Contribution from the Department of Chemistry, The City College of CUNY Convent Avenue & 138 Street, New York, New York 10031

Received July 31, 2006; E-mail: tlazaridis@ccny.cuny.edu

Abstract: The coiled coil is one of the simplest and best-studied protein structural motifs, consisting of two to five helices wound around each other. Empirical rules have been established on the tendency of different core sequences to form a certain oligomeric state but the physical forces behind this specificity are unclear. In this work, we model four sequences onto the structures of dimeric, trimeric, tetrameric, and pentameric coiled coils. We first examine the ability of an effective energy function (EEF1.1) to discriminate the correct oligomeric state for a given sequence. We find that inclusion of the translational, rotational, and side-chain conformational entropy is necessary for discriminating the native structures from their misassembled counterparts. The decomposition of the effective energy into residue contributions yields theoretical values for the oligomeric propensity of different residue types at different heptad positions. We find that certain calculated residue propensities are general and consistent with existing rules. For example, leucine at **d** favors dimers, leucine at **a** favors tetramers or pentamers, and isoleucine at **a** favors trimers. Other residue propensities are sequence context dependent. For example, glutamine at **d** favors trimers in one context and pentamers in another. Charged residues at **e** and **g** positions usually destabilize higher oligomers due to higher desolvation. Nonpolar residues at these positions confer pentamer specificity when combined with certain residues at positions **a** and **d**. Specifically, the pair Leu^{a}−Ala$^{g'}$ or the inverse was found to stabilize the pentamer. The small energy gap between the native and misfolded counterparts explains why a few mutations at the core sites are sufficient to induce a change in the oligomeric state of these peptides. A large number of possible experiments are suggested by these results.

Introduction

One of the simplest protein structural motifs is the coiled coil, consisting of two or more helices wound into a superhelix.[1,2] It is characterized by a seven amino acid heptad repeat **a-b-c-d-e-f-g** encoding amphipathic helices (**a** and **d** are hydrophobic; **b**, **c**, **e**, **f**, and **g** are polar). Coiled coils are found in viral fusion proteins, transcription factors, and other types of proteins where rigidity is required. They come in different oligomeric states. The crystal structure of influenza hemagglutinin provided the first high-resolution view of a trimeric coiled coil.[3] Dimeric coiled coils are observed in the "leucine zipper" dimerization elements of the bZIP transcription factors such as the yeast gene regulatory protein GCN4.[4,5] Mutations in the core regions of the coiled coil domain of GCN4 led to trimers and tetramers.[6] A natural, antiparallel tetramer is formed by the heterogeneous nuclear ribonucleoprotein C.[7] Pentameric coiled coils have been observed in thrombospondin (TSP) type 3, 4, and 5,[8] the cartilage oligomeric matrix protein (COMP),[9] the membrane domain of phospholamban (PLB),[10] and an engineered peptide with tryptophan at all **a** and **d** positions.[11]

The packing geometry is different in different oligomeric states.[6] The Cα−Cβ bonds in leucine **d** side chains in the GCN4-p1 dimer are directed into the core and perpendicular to a neighboring Cα−Cα vector sustained by the right and left side of the cavity where the side chain is to be buried. This is referred to as perpendicular packing mode. The β-branched residues, located at the **a** sites, direct their Cα−Cβ bonds away from the core and parallel to the Cα−Cα vector of the cavity where the side chain is to be buried. This is termed parallel packing mode. In trimers, the orientation of the **a** and **d** side chains is intermediate between parallel and perpendicular, termed acute packing mode. In the COMP pentamer, "knobs into holes" type interactions between **b**, **c**, **e**, and **g** side chains

(1) Lupas, A. *Trends Biochem. Sci.* **1996**, *21*, 375−382.
(2) Mason, J. M.; Arndt, K. M. *ChemBioChem* **2004**, *5*, 170−176.
(3) Wilson, I. A.; Skehel, J. J.; Wiley, D. C. *Nature* **1981**, *289*, 366−373.
(4) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. *Science* **1988**, *240*, 1759−1764.
(5) O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. *Science* **1991**, *254*, 539−544.
(6) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. *Science* **1993**, *262*, 1401−1406.
(7) Whitson, S. R.; LeStourgeon W. M.; Krezel, A. M. *J. Mol. Biol.* **2005**, *350*, 319−337.
(8) Qabar, A.; Derick, L.; Lawler, J.; Dixit, V. *J Biol. Chem.* **1995**, *270*, 12725−12729.
(9) Malashkevic, V. N.; Kammerer, R. A.; Efimov, V. P. *Science* **1996**, *274*, 761−765.
(10) Oxenoid K.; Chou, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10870−10875.
(11) Liu, J.; Yong, W.; Deng, Y.; Kallenbach, N. R.; Lu, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16156−16161.

10.1021/ja0655284 CCC: $33.50

become possible, and as a result, the helices have the least solvent-accessible surface area of all oligomers.[8]

The heptad repeat makes identification of coiled coils in protein sequences quite facile, and several bioinformatics prediction programs for detecting this pattern are available, such as COILS,[12] MULTICOIL,[13] SCORER,[14] MATCHER,[15] and MARCOIL.[16] Another program (SOCKET) can be used to detect coiled coils in a given 3D structure.[17] Programs for the prediction of partnering specificity based on favorable interhelical interactions also exist.[18−20] Although detection of coiled coils is relatively easy, predicting the oligomeric state from sequence is much more difficult. Some of the above programs aim to distinguish dimers from trimers, but they are not totally reliable.[21]

Numerous mutagenesis studies have provided insight into the contributions of different residues to stability and oligomeric specificity. Single mutations at the leucine **d** positions were insufficient to cause detectable loss of function.[22] However, interchange of isoleucine and leucine at the core led to a tetramer.[6] Harbury et al. suggested that due to restrictions in packing geometry, leucine side chains prefer structures that direct their Cα−Cβ bond perpendicular to the packing space in the neighboring helix.[6] This tetrameric peptide can switch from parallel to antiparallel arrangement by mutation at an **e** position glutamic acid.[23] A change in oligomeric state from dimer to tetramer can also be induced by inverting the sequence of GCN4-p1.[24] Moitra et al. examined the contribution to stability of different residues at the **d** positions in the C-terminal leucine zipper dimerization domain of vitellogenin binding protein and found leucine to be 2.9 kcal/mol per residue more stabilizing than isoleucine.[25] Because these two residues have the same size and composition, packing interactions must be responsible for this difference. Isoleucines presumably prefer to have their Cα−Cβ bonds parallel to the packing space and, as a result, might provide the most stability at position **a** of dimeric coiled coils.[6] This packing preference extends to other β-branched residues such as valine, which occupies most of the **a** positions in the leucine zipper motif. When both **a** and **d** positions are occupied by isoleucine, the peptide forms a trimer.[6]

Several mutagenesis studies have addressed the function of buried polar residues in the dimerization interface of bZIP proteins. Mutation of the central asparagine to leucine in the peptide pair ACID-p1 and BASE-p1 results in tetramers instead of dimers. The tetramers are more stable than the dimers but lack a unique helix orientation.[26,27] In the original ACID-p1 and BASE-p1 pair, the buried asparagine can interact only when the helices are in an antiparallel orientation. This suggests that hydrophobic interactions can contribute to the stability of the protein, but the requirement to satisfy the hydrogen-bonding potential of the buried asparagine imparts structural uniqueness. This is consistent with the experimental result of asparagine at **a** being insufficient to impart dimer specificity in a membrane bound peptide.[28] An Asn$^{\mathbf{a}}$ → Val$^{\mathbf{a}}$ mutation causes the GCN4-p1 peptide to lose its dimeric specificity and form both dimers and trimers,[7] but the same effect can be induced by other residue substitutions.[29] A definite conclusion cannot be obtained from mutation data since they do not yield absolute specificity.

Shu et al. grafted the core residues of the trimeric HIV gp41 coiled coil onto the GCN4 sequence and found that the resulting peptide (H38-p1) forms a trimer, showing that the core residues of gp41 are the determinants of trimeric specificity. They suggested that polar residues at the core, such as a threonine and a glutamine at **d** might be critical for trimerization.[30] Incorporation of all 20 amino acids in the central **d** position of a model coiled coil showed that threonine, valine, and isoleucine favor the three-stranded state, ionizable residues and tyrosine favor the two-stranded state, and the remaining amino acids, including glutamine, are indifferent.[31] Substitutions at position **a** showed that leucine, tyrosine, glutamine, and histidine favored trimers, while asparagine, lysine, arginine, and tryptophan favored dimers.[32] Also, the GCN4 mutant Asn^{16}→Gln was found to form trimers.[33] Akey et al. inserted single polar residues into the core positions of a GCN4 variant and found that asparagine favors dimers while most serine, threonine, and glutamine substitutions lead to a mixture of dimers and trimers, many of which, however, crystallize as trimers.[34] A Gln$^{27\mathbf{d}}$→Leu$^{27\mathbf{d}}$ mutation in the COMP pentamer increases the temperature of melting up to above 120 °C, but the oligomerization state is not affected.[35,36]

Polar or charged residues outside the core do not seem to provide much stability,[26,37−39] although they can provide specificity for heterodimers versus homodimers[34] and even affect the oligomeric state.[40−42] Kammerer et al.[21] discovered a motif of Arg−Glu salt bridges that seems to stabilize trimers. Mutation of the arginine led to the formation of tetramers. Mutation of an arginine to glutamine in a trimeric coiled coil led to the

(12) Lupas, A.; Van Dyke, M.; Strock, J. *Science* **1991**, *252*, 1162−1164.
(13) Wolf, E.; Kim, P.; Berger, B. *Protein Sci.* **1997**, *6*, 1179−1189.
(14) Wolfson, D. N.; Alber, T. *Protein Sci.* **1995**, *4*, 1596−1607.
(15) Fischetti, V. A.; Landau, G. M.; Schmidt, J. P.; Sellers, J. *Inf. Proc. Lett.* **1993**, *45*, 11−18.
(16) Delorenzi M.; Speed T. *Bioinformatics* **2002**, *18*, 617−625.
(17) Walshaw, J.; Woolfson, D. N. *J. Mol. Biol.* **2001**, *307*, 1427−1450.
(18) Fong, J. H.; Keating, A. E.; Singh, M. *Gen. Biol.* **2004**, *5*, R11.
(19) Deppmann, C. D.; Acharya, A.; Rishi,V.; Wobbes, B.; Smeekens, S.; Taparowsky, E. J.; Vinson, V. *Nucleic Acids Res.* **2004**, *32*, 3435−3445.
(20) Mason, J. M.; Scmitz, M. A.; Müller, K. M.; Arndt, K. M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 8989−8994.
(21) Kammerer, R. A.; Schulthess, T.; Landwehr, R.; Lustig, A.; Engel, J.; Aebi, U.; Steinmetz, M. O. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *95*, 13419−13424.
(22) Van Heeckeren, W. J.; Sellers, J. W.; Struhl, K. *Nucleic Acids Res.* **1992**, *20*, 3721−3724.
(23) Yadav, M. K.; Leman, L. J.; Price, D. J.; Brooks, C. L., III, Stout, C. D.; Ghadiri, M. R. *Biochemistry* **2006**, *45*, 4463−4473.
(24) Mittl, P. R.; Deillon, C.; Sargent, D.; Liu, N.; Klauser, S.; Thomas, R. M.; Gutte, B.; Grutter, M. G. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2562−2566.
(25) Moitra, J.; Szilak, L.; Krylov, D.; Vinson, C. *Biochemistry* **1997**, *36*, 12567−12573.
(26) Lumb, K. J.; Kim, P. S. *Biochemistry* **1995**, *34*, 8642−8648.
(27) Oakley, M. G.; Kim, P. S. *Biochemistry* **1998**, *37*, 12603−12610.
(28) Cristian, L.; Nanda, V.; Lear, J. D.; DeGrado, W. F. *Biochemistry* **1998**, *37*, 12603−12610.
(29) Gonzalez, L., Jr.; Woolfson, D. N.; Alber, T. *Nat. Struct. Biol.* **1996**, *3*, 1011−1018.
(30) Shu, W.; Ji, H.; Lu, M. *Biochemistry* **1999**, *38*, 5378−5385.
(31) Tripet, B.; Wagschal, K.; Lavigne, P.; Colin, C. T.; Hodges, R. S. *J. Mol. Biol.* **2000**, *300*, 377−402.
(32) Wagschal, K.; Tripet, B.; Hodges, R. S. *J. Mol. Biol.* **1999**, *285*, 785−803.
(33) Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. *Biochemistry* **1995**, *34* , 11645−11651.
(34) Akey, D. L.; Malashkevich, V. N.; Kim, P. S. *Biochemistry* **2001**, *40*, 6352−6360.
(35) Terskikh, A. V.; Potekhin, S. A.; Melnik, T. N.; Mach, J. P.; Kajava, A. *Lett. Pept. Sci.* **1997**, *4*, 297−304.
(36) Guo, Y.; Kammerer, R. A.; Engel, J. *Biophys. Chem.* **2000**, *85*, 179−186.
(37) Hu, J. C.; Newell, N. E.; Tidor, B.; Sauer, R. T. *Protein Sci.* **1993**, *2*, 1072−1084.
(38) Durr, E.; Jelesarov, I.; Bosshard, H. R. *Biochemistry* **1999**, *38*, 870−880.
(39) Marti, D. N.; Bosshard, H. R. *J. Mol. Biol.* **2003**, *330*, 621−637.
(40) Zheng, X.; Zhu, H.; Lashuel, H. A.; Hu, J. C. *Protein Sci.* **1997**, *6*, 2218−2226.
(41) Krylov, D.; Mikhailenko I.; Vinson C. *EMBO J.* **1994**, *13*, 2849−2861.
(42) Meier, M.; Lustig, A.; Aebi, U.; Burkhard, P. *J. Struct. Biol.* **2002,** *137*, 65−72.

formation of tetramers.[43] Peripheral polar residues are also thought to influence strand orientation by favoring the structure where electrostatic bridges can be formed and disfavoring structures leading to electrostatic clashes between like charges.[44]

The determinants of pentameric coiled coil assembly are more difficult to elucidate due to the smaller amount of available structural data. To date, the only available structures of pentameric coiled coils are the oligomerization domain of COMP studied by X-ray crystallography,[9] and the membrane domain of PLB studied by solution NMR in micelles.[10] However, a substantial amount of information can be extracted by comparing the amino acid sequence of COMP with that of the other members of the TSP gene family. All TSPs contain conserved sets of cysteines at either end of their coiled coil domains; nevertheless, TSP-1 and TSP-2 assemble into trimers whereas TSP-3, TSP-4, and TSP-5 (COMP) form pentamers.[8,45] The amino acid pattern at the **a** and **d** positions of TSP-3, TSP-4, and COMP is very similar to that of the tetrameric GCN4-LI, with a preponderance of leucine at **a** and isoleucine at **d**. The membrane domain of PLB contains leucine at positions 37**a**, 44**a**, and 51**a** and isoleucine at positions 40**d** and 47**d**.[10] Frank et al.[46] synthesized a water-soluble version of PLB by combining the core residues of PLB and the surface residues of COMP. It formed a pentamer, but a variant containing cysteine was mostly a tetramer. DeGrado and co-workers also synthesized a water-soluble analog of phospholamban (WSPLB). It was pentameric, but removal of residues 1−20 shifted the equilibrium toward tetramers.[47] Progressive truncations at the N terminus led to the conclusion that the Leu^{a}−Ile^{d} pattern encodes tetramer specificity while burial of polar amino acids at the segment 21−30 promotes pentamer formation.[48] The crystal structure of the 21−52 variant revealed an antiparallel tetramer. Comparison with a COMP-based pentameric model suggested that hydrogen-bonding interactions at **e** and **g** sites may play a role in determining the topology of the helix bundle.[49]

In summary, the following empirical rules have been established through residue substitutions: (a) leucine favors dimers when at position **d** and destabilizes them when at position **a**;[7] (b) isoleucine favors tetramers when at position **d** but will also favor trimers in the absence of a destabilizing pattern for lower order structures such as leucine at **a**;[6,24] (c) for some patterns of core residues that fit equally well in dimer and trimer structures, the inclusion of polar residues at core positions defines a single oligomeric state; for example, asparagine or arginine at **a** favors dimers and threonine at **d** favors trimers;[32,34] (d) polar or charged residues at **e** and **g** positions influence strand orientation by disfavoring structures where charge clashes will occur;[44] (e) a pattern of Ile^{d}−Leu^{a} favors tetramers, but the inclusion of polar residues, such as asparagine or glutamine, at the core positions might induce pentamer formation.[48]

A significant body of theoretical work has been done on coiled coils. Computer-generated models of leucine zippers together with a semiempirical free energy function found large stabilizing contributions from leucine at **d** positions in coiled coil dimers.[50] They also found stability and specificity contributions from charged residues at **e** and **g** positions. Brunger and co-workers developed a simulated annealing approach to structure prediction of coiled coils.[51−53] Analysis of packing energies of leucine and isoleucine in dimers and tetramers showed the importance of packing in determining oligomeric state.[53] Crick's parametrization for the backbone together with rotamer enumeration was used to predict coiled coil structures of different oligomeric states.[54] Free energy simulations on a dimeric coiled coil yielded 3.4 kcal/mol contribution to stability for a leucine pair at **d** and 0.8 kcal/mol for a valine pair at **a** relative to an alanine pair.[55] It should be noted, however, that these energetics are sequence context dependent. For example, it was found experimentally that the stabilization by leucine in a natural bZIP protein varies by over 2 kcal/mol at two **d** positions examined.[25]

Computational studies on a lattice using a reduced representation of the protein and a knowledge-based potential have been used to predict the oligomeric state of GCN4 and several of its mutants.[56] These studies concluded that the oligomeric state is determined by the balance between packing interactions and entropic factors, with entropy favoring lower order oligomers. Subsequent work used a more sophisticated treatment of the unfolded state[57] and different Monte Carlo methods for calculation of the partition functions.[58,59] This work has confirmed some empirically established oligomeric propensities, such as asparagine at **a** favoring dimers, leucine at **a** favoring trimers and tetramers, and isoleucine at **a** favoring trimers. However, the reduced protein representation and the use of a statistical potential places some limits in the physical interpretation of the results.

The goal of this work is to rationalize and extend the empirical rules derived from experiment using computational methods. We employed the program CHARMM[60] and the effective energy function EEF1.1[61] for the analysis of four coiled coil sequences known to favor different oligomeric states. First, the free energy of each state is estimated as the sum of the average effective energy and the configurational entropy to test whether the correct oligomeric state will be distinguished. Then we use the pairwise decomposability of the energy function to obtain individual residue contributions to the stability and specificity for each structure. We obtain many new insights that could be useful in predicting oligomeric state from sequence.

(43) Beck, K.; Gambee, J. E.; Kamawal A.; Bächinger H. P. *EMBO J.* **1997**, *16*, 3767−3777.
(44) Schnarr, N. A.; Kennan, A. J. *Org. Lett.* **2005**, *7*, 395−398.
(45) Sottile, J.; Selegue, J.; Mosher, D. F. *Biochemistry* **1991**, *30*, 6556−6562.
(46) Frank, S.; Kammerer, A. K.; Hellstern, S.; Pegoraro, S.; Stetefeld, J.; Lustig, A.; Moroder, L.; Engel, J. *Biochemistry* **2000**, *39*, 6825−6831.
(47) Slovic, A. M.; Summa, C. M.; Lear, J. D.; DeGrado, W. F. *Protein Sci.* **2003**, *12*, 337−348.
(48) Slovic, A. M.; Lear, J. D.; DeGrado, W. F. *J. Pept. Res.* **2005**, *65*, 312−321.
(49) Slovic, A. M.; Stayrook, S. E.; North, B.; DeGrado, W. F. *J. Mol. Biol.* **2005**, *348*, 777−787.
(50) Krystek, S. R., Jr.; Bruccoleri, R. E.; Novotny, J. *Int. J. Pept. Protein Res.* **1991**, *38*, 229−236.
(51) Nilges, M.; Brunger, A. T. *Protein Eng.* **1991**, *4*, 649−659.
(52) Nilges, M.; Brunger, A. T. *Proteins: Struct., Funct., Genet.* **1993**, *15*, 133−146.
(53) DeLano, W. L.; Brunger, A. T. *Proteins* **1994**, *20*, 105−123.
(54) Harbury, P. B.; Tidor, B.; Kim, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8408−8412.
(55) Zhang, L.; Hermans, J. *Protein Eng.* **1993**, *16*, 384−392.
(56) Vieth, M.; Kolinski, A.; Brooks, C. L., III, Skolnick, J. *J. Mol. Biol.* **1995**, *251*, 448−467.
(57) Vieth, M.; Kolinski, A.; Skolnick, J. *Biochemistry* **1996**, *35*, 966−967.
(58) Mohanty, D.; Kolinski, A.; Skolnick, J. *Biophys. J.* **1999**, *77*, 54−59.
(59) Viñals, J.; Kolinski, A.; Skolnick, J. *Biophys. J.* **2002**, *83*, 2801−2811.
(60) Brooks B. R.; Bruccoleri R. E.; Olafson B. D.; States D. J.; Swaminathan S.; Karplus M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(61) Lazaridis, T.; Karplus, M. *Proteins* **1999**, *35*, 133−152.

Table 1. Free Energy Estimates for All Sequences Studied[a]

oligomeric order	I	II	III	IV	V
			GCN4-p1		
avg eff. int. (300 000 steps)	−108.08	−130.04	−132.03	−130.16	−136.70
$-T\Delta S^{conf}$	0	6.72	14.46	16.00	18.00
$-T\Delta S^{trans}$	0	2.71	3.91	4.48	5.00
$-T\Delta S^{rot}$	0	4.21	5.53	6.38	6.56
free energy	−108.08 ± 0.19	−116.40 ± 0.26	−108.13 ± 0.26	−103.30 ± 0.36	−107.14 ± 0.25
			H38-p1		
avg eff. int. (300 000 steps)	−106.62	−114.89	−131.44	−131.50	−133.11
$-T\Delta S^{conf}$	0	5.54	10.08	12.33	15.98
$-T\Delta S^{trans}$	0	2.72	3.99	4.55	5.07
$-T\Delta S^{rot}$	0	4.34	5.65	6.41	6.75
free energy	−106.62± 0.19	−102.29± 0.26	−111.72± 0.26	−108.21± 0.36	−105.31± 0.25
			GCN4-LI		
avg eff. int. (300 000 steps)	−98.70	−110.32	−124.75	−134.81	−137.41
$-T\Delta S^{conf}$	0	8.05	15.53	17.29	21.26
$-T\Delta S^{trans}$	0	2.70	3.95	4.58	5.21
$-T\Delta S^{rot}$	0	4.31	5.70	6.60	6.96
free energy	−98.70 ± 0.17	−95.26 ± 0.23	−99.57 ± 0.26	−106.34 ± 0.27	−103.98 ± 0.25
			COMP		
avg eff. int. (300 000 steps)	−121.15	−130.69	−141.42	−144.79	−154.73
$-T\Delta S^{conf}$	0	5.17	5.43	8.84	13.15
$-T\Delta S^{trans}$	0	2.74	3.82	4.61	4.89
$-T\Delta S^{rot}$	0	4.09	5.66	6.67	6.68
free energy	−121.15 ± 0.20	−118.69 ± 0.27	−126.51 ± 0.28	−124.67 ± 0.29	−130.01 ± 0.30

[a] The effective energy and translational and rotational entropy terms are averages over four 0.6 ns (300 frames) Nosé−Hoover MD simulations. The configurational entropy has been evaluated on 15 frames per MD run. All entries are in units of kcal/(mol·helix). Lowest energy values among the oligomers are underlined.

Results

The free energy has been estimated as the average effective energy over the trajectory plus the configurational entropy contributions ($-T[\Delta S^{trans} + \Delta S^{rot} + \Delta S^{conf}]$). The results are shown in Table 1. While the ΔS^{trans} and ΔS^{rot} terms are nearly sequence independent, the $T\Delta S^{conf}$ term is strongly dependent on side chain type. All entropic terms disfavor oligomerization and are essential for the prediction of the native structure because the effective interaction alone favors the structure with the most packing interactions. For example, the average effective interaction over the trajectory yielded a minimum value for the dimer-specific GCN4-p1 and the trimer-specific H38-p1 sequences threaded onto the pentamer backbone (Table 1). Some side chains, which are solvent exposed in lower order oligomers, are substantially buried in the pentamer, resulting in a larger number of side chains getting locked into a single rotameric state. As a result, the conformational entropy loss is the largest.

Because the simulations are done with harmonic constraints and the monomer is a helix, rather than a random coil, the free energies obtained cannot be compared with experimental estimates. For example, the free energy of GCN4-p1 dimer formation under these conditions is about −16 kcal/mol, which is too large compared with the experimental value.[62] However, if simulations are done without constraints, a value closer to experiment is obtained, although still a bit too large (data not shown). This is probably due to the use of a folded helix as a reference state for the monomer.

More detailed analysis is provided by the free energy contribution per residue (Tables 2−5), which combines the effective energy and the side chain entropic conformational free energy (CONF). The effective energy per residue is the intraresidue effective energy plus one-half of the effective interaction energy of the residue with its surroundings; thus the sum of these values is equal to the total effective energy minus the total reference solvation free energy ($\sum_i \Delta G_i^{ref}$). The effective energy is decomposed into van der Waals (VDW), electrostatics (ELEC), solvation (SOLV), and bonded free energy (BOND). Tables of these contributions and of the side chain entropy per residue are given as Supporting Information. We define stability as the energy difference between the monomer and the oligomer in question and specificity as the difference between the oligomer in question and the average energy of all the other oligomers except the monomer.

GCN4-p1 Sequence. The structure of the hydrophobic core interface (**a** and **d** sites) for the dimer of the GCN4-p1 sequence after dynamics is summarized in Figure 1a. Leucine side chains at all **d** positions exhibit the characteristic perpendicular packing mode, while valine side chains at **a** positions exhibit parallel packing except for Val$^{31\mathbf{a}}$, where the deviation from parallel is evident in helix 1. We found that this side chain tends to flip during dynamics. Experimentally, Val$^{31\mathbf{a}}$ amide protons have been found to exchange much faster than their counterparts at other **a** positions.[63] This is most likely the result of fraying at the helix end making the packing geometry more flexible.

We observe that all leucines at **d** sites contribute favorably to dimer specificity (Table 2, underlined entries). However, the

(62) Kenar, K. T.; García-Moreno, B.; Freire, E. *Protein Sci.* **1995**, *4*, 1934−1938.

(63) Goodman, E. M.; Kim, P. S. *Biochemistry* **1991**, *30*, 11615−11620.

Table 2. Average Free Energy ($W - T\Delta S^{conf}$) Per Residue for the GCN4-p1 Sequence during Four 0.6 ns (300 Frames) Nosé−Hoover MD Simulations[a]

residue	I	II	III	IV	V
Ace	−4.04	−3.31	−4.58	−4.60	−4.04
Arg^{2g}	4.04	3.39	6.29	6.93	4.04
Met^{3a}	−1.97	−1.62	−1.18	−0.75	−1.97
Lys^{4b}	0.86	1.26	1.05	0.88	0.86
Gln^{5c}	−8.50	−8.61	−8.49	−8.28	−8.50
Leu^{6d}	−4.21	−6.92	−5.90	−5.49	−4.21
Glu^{7e}	−3.49	−4.53	−4.12	−3.66	−3.49
Asp^{8f}	−10.47	−10.84	−10.64	−10.33	−10.47
Lys^{9g}	−1.17	−0.99	−0.66	−0.56	−1.17
Val^{10a}	−2.53	−4.63	−4.82	−4.30	−2.53
Glu^{11b}	−4.15	−4.79	−4.45	−3.97	−4.15
Glu^{12c}	−4.45	−4.86	−4.44	−4.06	−4.45
Leu^{13d}	−4.68	−7.70	−6.08	−6.78	−4.68
Leu^{14e}	−5.78	−5.81	−5.86	−6.45	−5.78
Ser^{15f}	−8.21	−7.96	−7.81	−8.41	−8.21
Lys^{16g}	0.14	0.71	1.52	1.04	0.14
Asn^{17a}	−11.57	−11.19	−12.22	−10.81	−11.57
Tyr^{18b}	−1.42	−1.57	−1.18	−0.51	−1.42
His^{19c}	−3.32	−3.28	−3.41	−3.70	−3.32
Leu^{20d}	−4.52	−6.60	−6.31	−6.49	−4.52
Glu^{21e}	−4.19	−4.63	−4.36	−4.60	−4.19
Asn^{22f}	−11.63	−11.35	−11.49	−11.14	−11.63
Glu^{23g}	−3.97	−3.74	−4.16	−4.04	−3.97
Val^{24a}	−2.61	−3.68	−4.46	−3.87	−2.61
Ala^{25b} [b]	−5.51	−6.39	−6.21	−5.93	−5.51
Arg^{26c}	1.24	1.87	1.96	1.76	1.24
Leu^{27d}	−4.03	−6.81	−6.45	−6.35	−4.03
Lys^{28e}	1.30	0.59	1.10	1.33	1.30
Lys^{29f}	1.75	1.61	1.85	1.65	1.75
Leu^{30g}	−1.78	−1.70	−1.70	−2.40	−1.78
Val^{31a}	2.20	2.17	1.08	1.32	2.20
Cbx	−1.41	−1.39	−1.44	−1.62	−1.41
$\sum[W - T\Delta S^{conf}]$	−108.08	−123.32	−117.58	−114.15	−108.08

[a] All entries are in units of kcal/(mol·helix). Lowest energy values among the oligomers are underlined. The monomer energies have been included for completeness, but only the multimers are compared. The values reported here do not include the constant $\sum_i \Delta G_i^{ref_{solv}}$ contribution. The error bar on these values is around 0.1 kcal/mol. Ace is the acetyl blocking group and Cbx the methyl amide blocking group. [b] The stabilization observed for Ala^{25b} is an artifact, resulting from the rebuilding of part of helix A of the trimer (see Methods). This rebuilding results in a suboptimal hydrogen bond between Ala^{25b} and Lys^{29f}.

relative stabilization is context-dependent (Leu^{6d}, 1.1; Leu^{13d}, 1.0; Leu^{20d}, 0.4; and Leu^{24d}, 0.3 kcal/mol). The free energy decomposition indicates that their dimer specificities arise from different contributions. Leu^{6d}, Leu^{20d}, and Leu^{27d} favor the dimer in terms of VDW and BOND. Leu^{13d} favors the dimer due to BOND and CONF. Thus the sequence context can have an effect on the energetics of otherwise identical interacting amino acid pairs.

Asn^{17a} is highly conserved in bZIP transcription factors and is thought to impart dimeric specificity because the orientation of the asparagine Cα−Cβ bond in the dimer facilitates solvation of its polar component while allowing hydrogen bonding with its counterpart in a neighboring helix.[5] However, these conclusions are based on the crystal structure of the leucine zipper where the central asparagine adopts an asymmetric conformation. Goodman and Kim have reported that amide proton exchange is much faster at the region near Asn^{17a} than for the other parts of the GCN4-p1 dimer indicating motional flexibility on the chemical-shift time scale. The NMR structure of the Jun leucine zipper indicates that this pair of asparagines switches rapidly between two conformations.[63] In agreement with the NMR data of the Jun leucine zipper, the two Asn^{17a}'s of the GCN4-p1 dimer switch between two conformations (hydrogen bond donor/hydrogen bond acceptor) during the course of the MD simulations. However, we found that Asn^{17a} does not favor dimer. Instead it favors the trimer by 1.3 kcal/mol from a combined effect of ELEC and BOND. The VDW term provides stability but does not discriminate between dimer and trimer. SOLV favors the higher oligomers.

Lys^{16g} and Glu^{21e} make little or no contribution to stability but contribute to specificity: they favor dimer due to SOLV, 2.0 and 1.8 kcal/mol, and CONF, 1.8 and 0.3 kcal/mol, respectively. The dimer specificity arises not from salt bridge formation but from higher desolvation of glutamic acid and lysine in the higher oligomers. Higher order structures allow the formation of a Lys^{16g}−Glu$^{21e'}$ salt bridge, but the ELEC gain is outweighed by the SOLV cost of fixing the charges and the entropic cost of restricting the rotation about χ bonds. This result is in agreement with experiment: Lumb and Kim have studied the interhelical electrostatic interactions in the GCN4 leucine zipper and concluded that the interaction of the two charged residues did not contribute to stability. In fact, replacing glutamic acid with glutamine increased the stability of the coiled coil.[64] A similar situation exists for the Lys^{28e}−Glu$^{23g'}$ ion pair, which is observed to form a salt bridge in the crystal structure. SOLV and CONF favor the dimer for both residues, while the other terms favor higher oligomers. Arg^{2g} favors dimer in all three GCN4-based peptides. The dimer specificity is a nearly constant 1.4 kcal/mol independent of the sequence context and is due to SOLV and BOND. In the GCN4-p1 and H38-p1 sequences, the dimer is also favored by ELEC and CONF.

The conclusion is that ionizable residues at **e** and **g** positions favor dimer because they get desolvated in higher oligomers. However, there are some exceptions, like Lys^{9g} in the trimer sequence (Table 3); where the higher oligomers are favored over the dimer (with the trimer and tetramer 1.0 kcal/mol more stable than the dimer). Here SOLV still favors the dimer, but the other terms win out. The CONF term here favors the trimer. The different behavior of Lys^{9g} in the H38-p1 sequence must be attributable to the core substitutions Leu^{13d}→Thr and Leu^{6d}→Gln, but its exact origin is unclear.

Small amounts of stability and specificity for dimer come from solvent-exposed sites, such as Gln^{5c}, Asp^{8f}, Glu^{11b}, Glu^{12c}, and Tyr^{18b}. Their origin has not been analyzed in detail, although it could. For example, Tyr^{18b} favors dimer in GCN4-p1 and GCN4-LI, mainly due to SOLV and secondarily CONF and BOND. It provides a small stabilization to the dimer and destabilizes higher oligomers. This may be due to closer approach to Lys$^{16g'}$ in higher oligomers.

H38-p1 Sequence. Figure 1b shows the core interface details of the H38-p1 trimer structure after dynamics. In the crystal structure, leucine side chains exhibit 3-fold, propeller-like symmetry (acute packing); their Cα−Cβ bonds point toward the solvent while their −Cγ−(Cδ)$_2$ units are directed toward the core. Isoleucines exhibit similar packing symmetry with their −Cγ's pointing toward the solvent and the −Cγ−Cδ units directed toward the hydrophobic core. Even though the trimeric specificity of the H38-p1 peptide was reproduced in our calculations (Table 1), not all of the structural features of the crystal structure were maintained in the simulations. For example, it was found that the hydrogen bond network involving the Gln^{6d} side chains changes after dynamics. Shu et al. have

(64) Lumb, K. J.; Kim, P. S. *Science* **1995**, *268*, 436−439.

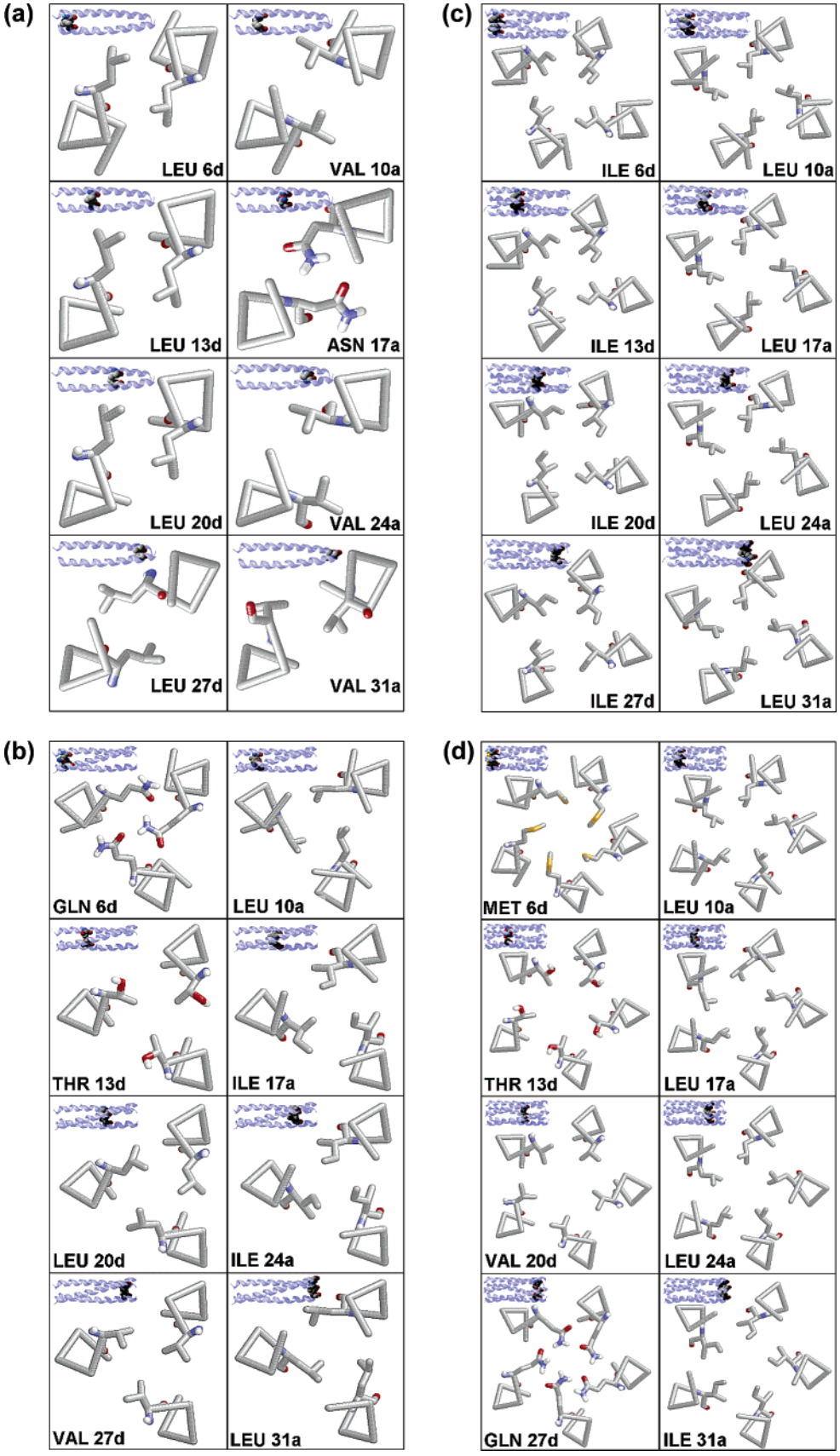

Figure 1. Structural details of the **a** and **d** residues viewed from NH_2 to COOH terminus: (a) GCN4-p1 dimer, (b) H38-p1 trimer, (c) GCN4-LI tetramer, and (d) COMP pentamer. Images at the upper left corners of each panel indicate the location of the side chain within the protein.

suggested that Gln^{6d} imparts trimer specificity due to the formation of a 3-fold symmetric hydrogen bond network between the side chain amide and the backbone carbonyl of the opposite helix.[26] We observed that the hydrogen bond

Table 3. Average Free Energy ($W - T\Delta S^{conf}$) Per Residue for the H38-p1 Sequence during Four 0.6 ns (300 Frames) Nosé−Hoover MD Simulations[a]

residue	I	II	III	IV	V
Ace	−4.67	−3.46	−4.91	−4.80	−4.67
Arg[2g]	3.79	2.10	3.71	3.79	3.79
Ile[3a]	−0.20	−0.98	−1.32	−0.07	−0.20
Lys[4b]	0.58	1.35	0.80	1.03	0.58
Gln[5c]	−8.16	−8.20	−8.43	−7.78	−8.16
Gln[6d]	−9.33	−10.37	−10.86	−10.84	−9.33
Glu[7e]	−4.03	−3.95	−4.10	−3.79	−4.03
Asp[8f]	−10.62	−10.87	−10.63	−10.44	−10.62
Lys[9g]	−2.35	−1.22	−2.45	−2.45	−2.35
Leu[10a]	−4.30	−4.22	−6.24	−6.28	−4.30
Glu[11b]	−4.81	−4.92	−4.74	−4.21	−4.81
Glu[12c]	−4.95	−5.47	−5.13	−4.83	−4.95
Thr[13d]	−4.85	−5.16	−5.35	−5.65	−4.85
Leu[14e]	−5.24	−5.24	−5.53	−5.05	−5.24
Ser[15f]	−8.11	−7.75	−7.78	−8.13	−8.11
Lys[16g]	0.03	0.64	1.18	1.23	0.03
Ile[17a]	−3.11	−3.12	−4.66	−4.02	−3.11
Tyr[18b]	−1.30	−1.19	−1.21	−0.23	−1.30
His[19c]	−3.51	−3.67	−3.45	−3.68	−3.51
Leu[20d]	−4.76	−7.36	−6.93	−7.08	−4.76
Glu[21e]	−3.56	−3.26	−3.49	−4.18	−3.56
Asn[22f]	−11.28	−11.43	−11.40	−11.15	−11.28
Glu[23g]	−4.10	−3.79	−4.29	−5.01	−4.10
Ile[24a]	−1.89	−1.57	−3.77	−3.22	−1.89
Ala[25b]	−5.61	−6.16	−6.13	−6.01	−5.61
Arg[26c]	1.22	1.95	1.17	1.22	1.22
Val[27d]	−2.71	−3.45	−4.74	−4.64	−2.71
Lys[28e]	1.36	0.73	1.07	1.31	1.36
Lys[29f]	1.96	1.94	1.88	1.84	1.96
Leu[30g]	−1.92	−1.94	−2.05	−3.15	−1.92
Leu[31a]	1.21	1.94	−0.03	−1.24	1.21
Cbx	−1.39	−1.26	−1.56	−1.66	−1.39
$\sum[W - T\Delta S^{conf}]$	−106.62	−109.35	−121.36	−119.17	−106.62

[a] All entries are in units of kcal/(mol·helix). Lowest energy values among the oligomers are underlined. The monomer energies have been included for completeness, but only the multimers are compared. The values reported here do not include the constant $\sum_i \Delta G_i^{ref_{solv}}$ contribution. The error bar on these values is around 0.1 kcal/mol. Ace is the acetyl blocking group and Cbx the methyl amide blocking group.

network is not symmetric after dynamics but is formed by the −Nϵ of one helix and the −Oϵ's of the other two helices. Our model, however, does not include a water molecule (W-80) trapped between the three glutamines. As a result, one of the glutamine side chains moves its amide group into the core to fill that space. We found that Gln[6d] has little trimer specificity in terms of free energy, but the ELEC term favors trimer by 1.5 kcal/mol.

All β-branched residues contribute to trimer specificity at both **a** and **d** positions (Table 3, underlined entries). Val[27d] has 0.6 kcal/mol of overall trimer specificity, but the tetramer is equally favored. In agreement with the empirical rules of Harbury et al., valine has a weak preference for the parallel packing exhibited by the **d** sites of the tetramer.[6] Thus the oligomeric preference of valine at **d** can switch between trimer and tetramer; the end result is the combined effect of the four side chains that define the complementary cavity where valine at **d** interacts. For example, Val[20d] in the COMP sequence favors the tetramer.

Ile[3a], Ile[17a], and Ile[24a] favor trimer by 0.7, 1.0, and 1.0 kcal/mol, respectively. Ile[3a] and Ile[24a] derive their specificity from a combined effect of VDW and bonded energy, while the situation is more complex for Ile[17a]. Isoleucines at **a** have a well-defined oligomeric preference because there are two different units attached to the −Cβ. One is a −Cγ, and the other is a −Cγ−Cδ. The latter performs the function of knob in the "knobs into holes" packing pattern. Thus in the trimer sequence and preferred isoleucine rotamer, there is only one position that directs the −Cγ−Cδ unit toward the complementary cavity in the correct orientation, and that is position **a**. Even with this packing constraint, one out of three isoleucines in the H38-p1 tends to flip during some simulations, which suggests that the core in this oligopeptide is poorly packed.

Thr[13d] has been proposed as critical for H38-p1 trimerization because it is part of the HIV-1 gp41 core.[26] A leucine → threonine mutation at position 13**d** in the GCN4-pVL peptide changed the oligomeric distribution from a mixture of dimers and trimers to mostly trimers.[34] Nevertheless, our calculations indicate that Thr[13d] favors the tetramer. It is possible that threonine at **d** is not trimer specific but simply less dimer specific than leucine. Alternatively, threonine may favor trimer by affecting the contributions of neighboring residues. For example, Leu[14e] was tetramer specific in the GCN4-p1 sequence and is trimer specific in the H38-p1 sequence.

Overall, among the eight sequence substitutions going from GCN4-p1 to H38-p1, the most important ones for the switch from dimer to trimer seem to be the three leucines at **d**, which favor the dimer, while their replacements either favor the trimer or disfavor the dimer.

GCN4-LI Sequence. Figure 1c shows the core interface details of the native GCN4-LI tetramer after dynamics. The hydrophobic core remains unchanged after dynamics.

Three out of four isoleucines at **d** sites show the expected tetramer preference: Ile[6d], Ile[13d], and Ile[20d]. Their contributions to tetramer specificity are 0.3, 1.0, and 1.1 kcal/mol respectively. The trimer and pentamer have comparable energies, but the dimer is strongly disfavored. Ile[27d] exhibits a slight preference for pentamer. One of the Ile[27d] residues flips at the Cα−Cβ−Cγ1−Cδ dihedral in three out of four MD simulations. This might be the result of higher flexibility in this region. In all cases, the tetramer is favored by SOLV and BOND.

Leucine at position **a** in the GCN4-LI sequence destabilizes dimers and trimers. Two leucine residues at **a** positions favor the tetramer (Leu[17a] and Leu[31a]), while Leu[10a] and Leu[24a] favor pentamer by a small amount. Thus, in the tetramer sequence, the preferred isoleucine rotamer at position **d** directs tetramer formation in a slightly context dependent manner, while leucine at position **a** acts as a destabilizing factor for lower order oligomers, also in a context dependent manner.

Leu[14e] favors tetramer by 1.3 kcal/mol in terms of VDW. This is likely due to interactions with Ile[13d] from a neighboring chain. Glu[21e] favors the tetramer by 0.6 kcal/mol. It is interesting that in the GCN4-p1 peptide it favors dimer but switches to tetramer in H38-p1 and GCN4-LI. The SOLV and CONF terms still favor dimer, but VDW and ELEC outweigh the desolvation cost in higher oligomers. There are occasional interactions with Lys[16g], His[19c], or both from a neighboring helix. The change in behavior of this residue may be an indirect effect of the replacement of Asn[17a]. Glu[23g] also favors the tetramer, which exhibits low BOND and VDW terms but less desolvation than the pentamer.

COMP Sequence. The pentameric COMP variant exhibits the least disruption of the hydrophobic core upon molecular dynamics. As can be seen from Figure 1d, the side chains at the pentamer core have the highest level of symmetry as compared with other structures. One exception is Gln[27d], whose

amide group flips during the simulations. Gln$^{27\mathbf{d}}$ was built by SCWRL3.0 within a symmetric hydrogen bond network with its amide groups pointing toward the COOH terminus. The crystal structure contains a chloride ion trapped between five glutamines, and the hydrogen bond network is symmetric,[9] but if the ion is replaced by all-trans retinol, four of the glutamine amides point toward the NH_2 terminus.[65] Thus the conformation of the Gln$^{27\mathbf{d}}$ ring depends on the hydrogen-bonding requirements and whether or not a molecule is bound to it.

EEF1.1 predicts the pentameric specificity of Gln$^{27\mathbf{d}}$ to be 1.4 kcal/mol mainly from ELEC. The side chain entropy loss for Gln$^{27\mathbf{d}}$ is the largest in the pentamer. This means that the Gln$^{27\mathbf{d}}$ hydrogen bond network is very stable regardless of the resulting conformation. However, the oligomeric specificity of glutamine at **d** might depend on the sequence context. For example, glutamine at **d** exhibited trimer specificity in H38-p1. Based on sequence analysis, Gln$^{27\mathbf{d}}$ is supposed to be important since it is conserved in COMP from several sources.[9] Nevertheless, the Gln$^{27\mathbf{d}}$→Leu$^{27\mathbf{d}}$ mutation increases the temperature of melting up to above 120 °C.[35,36]

Consistent with the previous observations on GCN4-based peptides, Leu$^{10\mathbf{a}}$, Leu$^{17\mathbf{a}}$, and Leu$^{24\mathbf{a}}$ favor the pentamer by 0.8, 0.5, and 0.7 kcal/mol, respectively, with the tetramer coming second. VDW slightly favors the tetramer, and SOLV favors the pentamer. CONF favors the pentamer over the tetramer, although it is usually lowest for the dimer.

Proline is seldom found in coiled coil sequences probably due to its helix-breaking property. Pro$^{4\mathbf{b}}$ is slightly tetramer specific due to bonded and ELEC energies (VDW favors the pentamer), but it favors lower order oligomers the least (Table 5).

Met$^{6\mathbf{d}}$ slightly favors the pentamer by a combined effect of SOLV, BOND, and ELEC. Met$^{3\mathbf{a}}$ in the GCN4-p1 and GCN4-LI, as well as a similar position in human TSP-4 (data not shown), favors pentamer in terms of BOND, ELEC, and VDW. Thus methionine at the core seems to favor pentamer. The other two residues at **d** positions (Thr$^{13\mathbf{d}}$ and Val$^{20\mathbf{d}}$) do not favor pentamer. Thr$^{13\mathbf{d}}$ favors trimer, and the specificity comes mostly from 0.7 kcal/mol ELEC. This is the effect of interactions of Thr$^{13\mathbf{d}}$ Oγ with the Asn$^{14\mathbf{e}}$ Nδ from a neighboring helix; this residue also favors trimer in terms of ELEC. Val$^{20\mathbf{d}}$ favors tetramer; the source of specificity is not clear, but the BOND term is lowest in the tetramer.

The VDW term of Ala$^{16\mathbf{g}}$ provides most of its 0.7 kcal/mol pentamer specificity, which means that alanine at **g** provides the best fit for a kind of extended knobs-into-holes interaction in the pentamer.[9] Ala$^{16\mathbf{g}}$ interacts with Leu$^{17\mathbf{a}}$ from another helix, which is also pentamer specific. A similar pattern exists for the Ala$^{3\mathbf{a}}$−Leu$^{2\mathbf{g}'}$ pair where the alanine residue shows a relatively high VDW stabilization term in the pentamer. The inverse interaction, observed in Leu$^{23\mathbf{g}}$−Leu$^{24\mathbf{a}'}$, is also pentamer specific. The fact that Leu$^{2\mathbf{g}}$ is not pentamer specific may be due to its position near the end of the helix. We compared the side chain entropy loss of Leu$^{23\mathbf{g}}$ with that of Leu$^{24\mathbf{a}}$ and found the **g** position to have a larger loss of rotamer population indicating that in the pentamer structure the **g** positions are locked as much or even more than the standard core positions.

Three residues at position **e** also exhibit preference for the pentamer structure. This is another extended knobs-into-holes interaction pattern with **e** position residues acting as knobs. One exception is Asn$^{14\mathbf{e}}$, which favors pentamer in terms of VDW but favors the trimer in terms of ELEC. Arg$^{21\mathbf{e}}$ has the advantage of a long nonpolar segment that allows it to serve as a knob while retaining the advantage of electrostatic interactions. Arg$^{21\mathbf{e}}$ favors the pentamer by 1.0 kcal/mol from a combined effect of VDW and ELEC.

Pentamer specific sequences such as TSP-3 and TSP-4 are very similar to COMP in the preponderance of hydrophobic residues at **e** and **g** positions. These include Leu$^{7\mathbf{e}}$, Leu$^{23\mathbf{g}}$, and Val$^{28\mathbf{e}}$, which favor the pentamer by 0.2, 1.9, and 1.6 kcal/mol, respectively. The driving force for all is VDW. Their relative contributions indicate that there might be a preference for β-branched residues at position **e** and leucine at position **g** in the pentamer. These residues are part of an extended core, and their behavior indicates that VDW interactions of the types **a**−**g′** and **d**−**e′** are one of the main determinants of pentamer specificity.

Table 4. Average Free Energy ($W - T\Delta S^{conf}$) Per Residue for the GCN4-LI Sequence during Four 0.6 ns (300 Frames) Nosé−Hoover MD Simulations[a]

residue	I	II	III	IV	V
Ace	−4.78	−3.39	−4.45	−4.56	−4.78
Arg$^{2\mathbf{g}}$	5.27	4.10	5.85	6.49	5.27
Met$^{3\mathbf{a}}$	−2.12	−1.23	−0.54	−1.22	−2.12
Lys$^{4\mathbf{b}}$	0.02	1.55	0.78	0.50	0.02
Gln$^{5\mathbf{c}}$	−9.21	−8.75	−8.43	−8.31	−9.21
Ile$^{6\mathbf{d}}$	−1.68	−3.20	−3.91	−3.92	−1.68
Glu$^{7\mathbf{e}}$	−4.07	−4.10	−4.06	−3.72	−4.07
Asp$^{8\mathbf{f}}$	−10.19	−10.72	−10.52	−10.14	−10.19
Lys$^{9\mathbf{g}}$	−1.40	−1.41	−0.50	−0.86	−1.40
Leu$^{10\mathbf{a}}$	−4.17	−4.87	−5.71	−7.18	−4.17
Glu$^{11\mathbf{b}}$	−4.73	−4.70	−5.01	−4.55	−4.73
Glu$^{12\mathbf{c}}$	−4.32	−4.76	−4.53	−4.18	−4.32
Ile$^{13\mathbf{d}}$	−2.21	−3.04	−4.39	−4.97	−2.21
Leu$^{14\mathbf{e}}$	−5.72	−6.05	−6.14	−6.53	−5.72
Ser$^{15\mathbf{f}}$	−8.06	−7.93	−7.73	−7.95	−8.06
Lys$^{16\mathbf{g}}$	−0.14	−0.32	1.38	0.66	−0.14
Leu$^{17\mathbf{a}}$	−5.56	−6.49	−7.09	−7.71	−5.56
Tyr$^{18\mathbf{b}}$	−1.36	−1.44	−1.33	−0.60	−1.36
His$^{19\mathbf{c}}$	−3.53	−3.47	−3.35	−3.78	−3.53
Ile$^{20\mathbf{d}}$	−2.40	−2.52	−4.23	−4.95	−2.40
Glu$^{21\mathbf{e}}$	−3.38	−3.78	−3.56	−4.21	−3.38
Asn$^{22\mathbf{f}}$	−11.30	−11.56	−11.51	−11.08	−11.30
Glu$^{23\mathbf{g}}$	−3.82	−3.85	−3.90	−4.67	−3.82
Leu$^{24\mathbf{a}}$	−4.15	−4.63	−5.74	−7.31	−4.15
Ala$^{25\mathbf{b}}$	−5.71	−6.29	−6.31	−6.23	−5.71
Arg$^{26\mathbf{c}}$	0.89	1.41	1.63	1.23	0.89
Ile$^{27\mathbf{d}}$	−1.61	−2.54	−4.22	−4.21	−1.61
Lys$^{28\mathbf{e}}$	1.60	0.86	0.65	1.01	1.60
Lys$^{29\mathbf{f}}$	1.80	2.06	1.33	1.82	1.80
Leu$^{30\mathbf{g}}$	−2.09	−2.06	−1.96	−3.09	−2.09
Leu$^{31\mathbf{a}}$	0.77	2.04	−0.14	−1.60	0.77
Cbx	−1.35	−1.20	−1.58	−1.68	−1.35
$\sum[W - T\Delta S^{conf}]$	−98.70	−102.28	−109.22	−117.52	−98.70

[a] All entries are in units of kcal/(mol·helix). Lowest energy values among the oligomers are underlined. The monomer energies have been included for completeness, but only the multimers are compared. The values reported here do not include the constant $\sum_i \Delta G_i^{ref_{solv}}$ contribution. The error bar on these values is around 0.1 kcal/mol. Ace is the acetyl blocking group and Cbx the methyl amide blocking group.

Discussion

Estimates of the free energy of each sequence threaded onto the different oligomeric structures showed that the effective

(65) Guo, Y.; Bozic, D.; Malashkevich, V. N.; Kammerer, R. A.; Schulthess, T.; Engel, J. *EMBO J.* **1998**, *17*, 5265−5272.

Table 5. Average Free Energy ($W - T\Delta S^{conf}$) Per Residue for the COMP Sequence during Four 0.6 ns (300 Frames) Nosé−Hoover MD Simulations[a]

residue	I	II	III	IV	V
Ace	−3.16	−3.33	−3.75	−3.68	−3.16
Leu^{2g}	−3.27	−2.92	−2.71	−2.42	−3.27
Ala^{3a}	0.09	−0.41	−0.48	0.14	0.09
Pro^{4b}	5.56	5.48	5.25	4.84	5.56
Gln^{5c}	−8.04	−8.38	−8.46	−8.46	−8.04
Met^{6d}	−4.51	−3.87	−4.33	−4.09	−4.51
Leu^{7e}	−5.05	−4.76	−5.23	−5.07	−5.05
Arg^{8f}	1.46	1.55	1.44	2.21	1.46
Glu^{9g}	−5.91	−6.04	−6.36	−6.28	−5.91
Leu^{10a}	−4.57	−4.92	−6.29	−6.71	−4.57
Gln^{11b}	−8.65	−8.44	−8.60	−8.49	−8.65
Glu^{12c}	−5.20	−5.42	−5.52	−4.94	−5.20
Thr^{13d}	−4.57	−4.39	−6.06	−5.23	−4.57
Asn^{14e}	−10.59	−10.99	−11.43	−10.56	−10.59
Ala^{15f}	−6.72	−6.81	−6.68	−6.64	−6.72
Ala^{16g}	−6.56	−7.12	−7.20	−7.42	−6.56
Leu^{17a}	−4.15	−4.78	−5.09	−5.55	−4.15
Gln^{18b}	−8.99	−9.20	−9.16	−8.50	−8.99
Asp^{19c}	−10.33	−10.17	−10.63	−10.45	−10.33
Val^{20d}	−3.07	−4.01	−4.35	−5.27	−3.07
Arg^{21e}	2.48	2.36	3.30	2.54	2.48
Glu^{22f}	−4.44	−4.24	−4.44	−4.20	−4.44
Leu^{23g}	−4.95	−5.27	−5.20	−6.27	−4.95
Leu^{24a}	−4.40	−5.37	−6.46	−6.86	−4.40
Arg^{25b}	2.95	2.23	2.44	3.62	2.95
Gln^{26c}	−9.63	−8.74	−9.50	−10.01	−9.63
Gln^{27d}	−8.00	−7.63	−9.67	−9.31	−8.00
Val^{28e}	−1.49	−2.27	−2.09	−2.98	−1.49
Lys^{29f}	1.85	1.90	1.89	1.31	1.85
Glu^{30g}	−1.38	−1.78	−1.60	−1.77	−1.38
Ile^{31a}	3.57	3.53	2.45	2.16	3.57
Cbx	−1.48	−1.29	−1.48	−1.58	−1.48
$\sum[W - T\Delta S^{conf}]$	−121.15	−125.51	−135.98	−135.94	−121.15

[a] All entries are in units of kcal/(mol·helix). Lowest energy values among the oligomers are underlined. The monomer energies have been included for completeness, but only the multimers are compared. The values reported here do not include the constant $\sum_i \Delta G_i^{ref_{solv}}$ contribution. The error bar on these values is around 0.1 kcal/mol. Ace is the acetyl blocking group and Cbx the methyl amide blocking group.

energy alone is not sufficient to discriminate the correct oligomeric state of coiled coils. High order oligomers have the most nonbonded contacts and thus the lowest effective energy. However, inclusion of the entropic contributions allowed the correct oligomeric state to be reproduced for all structures, at least for the four sequences studied here. The validity of this conclusion should be tested by examination of a larger number of coiled coil sequences.

Individual residue contributions to stability and specificity allow us to derive oligomeric propensities of amino acids at different heptad positions. It should be noted that the quantity we calculate is not accessible experimentally. Mutation of one residue to another changes not only the contribution of that residue but also the contribution of the surrounding residues.[66] This needs to be kept in mind when we compare our findings with experiment.

The analysis of residue contributions to oligomeric specificity leads to the following main conclusions:

(1) Leucine at position **d** imparts dimer specificity with the extent of stabilization depending on the sequence context. This is in agreement with the experimental results.[7] The dimer specificity comes mainly from VDW, CONF, and BOND, which is essentially "packing energy".

(2) Isoleucine at **a** confers trimer specificity. Isoleucine at **a** in the trimer structure directs its $-C\gamma-C\delta$ unit toward the protein core, while other structures (notably the dimer) direct the smaller $-C\gamma$ toward the protein core. Thus, the trimer packs the most atoms while retaining a low-energy rotamer about $C\alpha-C\beta$. Again, it is a packing effect.

(3) Leucine at **a** favors tetramers, disfavors lower oligomers, and should be indicative of tetramer or pentamer. There is little difference in packing between tetramer and pentamer, but the trimer and dimer leave the $-C\gamma(C\delta)_2$ unit either solvent exposed or too tightly packed. We noted that for leucine at **a**, only higher order structures will pack the $-C\gamma(C\delta)_2$ unit at the core while retaining the most populated rotamer.

(4) Valine at **a** favors trimer. We do not have much data for valine at **d**. In one case, it favors trimer and in another tetramer. Interestingly, valine is found at **a** as often as at **d** positions in other trimeric coiled coils.[67] With two identical $-C\gamma$ units, the packing modes of valine at **a** and **d** are very similar.

(5) Isoleucine at **d** favors tetramer. From dimer to tetramer, the hole where the isoleucine $-C\gamma-C\delta$ knob packs in a neighboring helix has rotated by 90°. The atoms that would be solvent exposed in the dimer are well packed and shielded from the solvent in the tetramer.

(6) Nonpolar side chains at positions **e** and **g** confer pentamer specificity when combined with certain residues at positions **a** and **d**. It is also expected that side chains that prefer positions **a** and **d** in the dimer will prefer positions **e** and **g** in the pentamer, respectively. Packing in the extended pentameric core (positions **a**, **d**, **e**, and **g**) looks similar to the packing pattern of the dimer core (positions **a** and **d**).

(7) Methionine at **a** or **d** favors pentamer. Part of the specificity comes from BOND, meaning that methionine has the least steric clashes when in the pentameric core. Interestingly, Met^{a} happens to mark the end of the dimerization domain in GCN4-p1.[5] However, a designed peptide with phenylalanine at all core positions forms a pentamer, and when one phenylalanine is changed to methionine, it forms a tetramer.[68] Perhaps there is context dependence in the oligomeric propensity of methionine.

(8) Polar or charged residues at **e** and **g** usually favor the structure that allows for maximum solvation (dimer), but exceptions have been observed. Our observations in the sequences studied suggest that the resulting specificity is likely a tradeoff between [SOLV + BOND + CONF] and [VDW + ELEC]; if the latter terms outweigh the former, the specificity will shift toward higher oligomers.

(9) Bulky side chains at positions **c** and **b** should destabilize higher oligomers for steric reasons. One example is Tyr^{18b} in the GCN4-based sequences.

Overall, our results agree with empirical rules derived from experimental mutagenesis studies. However, there are exceptions. Our calculation results suggest that Asn^{17a} favors trimer in the GCN4-p1 sequence, in contrast to the currently accepted rules. Harbury et al. reported that an Asn^{a} → Val^{a} mutation causes the GCN4-p1 peptide to lose its dimeric specificity and form both dimers and trimers.[6] Mutation to alanine, aminobutyric acid (Abu), or glutamine also had the same effect.[29] However, these mutation data only show that alanine, aminobu-

(66) Lazaridis, T.; Karplus, M. *Biophys. Chem.* **2003**, *100*, 367−395.
(67) Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D. F. *FEBS Lett.* **1996**, *387*, 36−41.
(68) Liu, J.; Zheng, Q.; Deng, Y.; Kallenbach, N. R.; Lu, M. *J. Mol. Biol.* **2006**, *361*, 168−179.

tyric acid, valine, or leucine at **a** are less dimer specific or more trimer specific than asparagine. The absolute specificity is not revealed by mutation experiments. Interestingly, asparagine at **a** was not sufficient to impart dimer specificity in the nonpolar environment of a membrane.[28] This is consistent with our result of ELEC favoring trimers. Perhaps the SOLV contribution is underestimated in our implicit model, and this leads to asparagine favoring trimers also in an aqueous environment.

In this work, we obtained hundreds of free energy contributions and analyzed the physical basis of many of them. In some cases, the physical origin of the small energy differences observed is not obvious. Understanding these subtle effects will require a more detailed decomposition of the energies beyond the residue level used in this work. For example, a difference in solvation energy of a tyrosine residue could come from the backbone, the nonpolar part of the side chain, or the hydroxyl group. Such an analysis could be done in the future for residues of special interest.

Our results expand the empirical rules for the oligomeric propensity of different residue types at different heptad positions beyond the classic **g**−**e**′, **a**−**a**′, and **d**−**d**′ interactions by considering **d**−**a**′, **a**−**d**′, **a**−**g**′, and **d**−**e**′ interactions as well. The magnitude of some energy terms that we obtain indicates that **a**−**g**′ and **d**−**e**′ pairwise interactions play an important role in determining the oligomeric specificity of high order structures. For example, our results indicate that the pentameric state could be encoded into a tetramer specific sequence with mutations at a few **g** position side chains (such as Lys$^{9\mathbf{g}}$, Lys$^{16\mathbf{g}}$, and Glu$^{23\mathbf{g}}$ → Ala in the GCN4-LI peptide). This and many other results from this work are experimentally testable.

One weakness of implicit solvent models is that structured water molecules are ignored. For interactions involving polar residues at the core interface, this can have significant effects on our results. For example, Gln$^{6\mathbf{d}}$ imparts trimer specificity to the H38-p1 peptide, but the energetic advantage of trimer is very small compared with the tetramer. The trimeric specifity might have been underestimated because our calculations do not take into consideration the effect of a water molecule (W-80 in PDB) trapped between the three glutamines. In the absence of a water molecule, the distance between these side chains will change and affect the calculation results of EEF1.1. It should also be kept in mind that this work is based on a pairwise additive effective energy function. This is a simplification whose effects could only be gauged by comparison to more sophisticated methods (polarizable force fields and non-pairwise additive solvation potentials).

The results of this work provide not only qualitative rules for the oligomeric propensities of amino acids at different coiled coil positions but also quantitative estimates of their contributions to stability and oligomeric specificity. These numerical values could be useful in the future for developing algorithms that predict coiled coil oligomeric state from amino acid sequence.

Methods

Energy Function. The probability of a conformation is determined by its effective energy,

$$W = H_{\text{intra}} + \Delta G^{\text{solv}} \qquad (1)$$

where H_{intra} and ΔG^{solv} are the intramolecular energy and the solvation free energy, respectively.[61] In EEF1, the ΔG^{solv} term is approximated as a sum of contributions from all the atoms in the macromolecule. The solvation free energy of each atom is equal to that of the same atom in a small model compound minus the solvation free energy it loses due to solvent exclusion by surrounding atoms:

$$\Delta G^{\text{solv}} = \sum_i \Delta G_i^{\text{solv}} = \sum_i \Delta G_i^{\text{ref}} - \sum_i \sum_{j \neq 1} f_i(\mathbf{r}_{ij}) V_j \qquad (2)$$

where ΔG_i^{solv} is the solvation free energy of group i, ΔG_i^{ref} is the solvation free energy in a fully solvent-exposed model compound, and the summation $\sum_{j \neq 1} f_i(\mathbf{r}_{ij}) V_j$ accounts for the exclusion of solvent around group i due the shielding by groups j. EEF1.1 is an updated parametrization[69] based on potentials of mean force calculated in explicit solvent.[70]

Structures. The structures used in this work were generated beginning with atomic coordinates obtained from the Protein Data Bank: the GCN4-p1 dimer (2ZTA), the H38-p1 trimer (1CE0), the GCN4-LI tetramer (1GCL), and the COMP pentamer (1VDF). The comparison of relative stabilities using EEF1.1 is meaningful only if the structures under investigation have the same number of atoms. To meet this requirement, each monomer unit was truncated to four heptad repeats. It is assumed that the truncation does not affect the oligomeric

Table 6. The Four Sequences Studied, Identified by PDB Code[a]

	II 2ZTA	III 1CE0	IV 1GCL	V 1VDF
1	Ace	Ace	Ace	Ace
2	Arg	Arg	Arg	Leu
3	**Met**	**Ile**	**Met**	**Ala**
4	Lys	Lys	Lys	Pro
5	Gln	Gln	Gln	Gln
6	Leu	Gln	Ile	Met
7	Glu	Glu	Glu	Leu
8	Asp	Asp	Asp	Arg
9	Lys	Lys	Lys	Glu
10	**Val**	**Leu**	**Leu**	**Leu**
11	Glu	Glu	Glu	Gln
12	Glu	Glu	Glu	Glu
13	Leu	Thr	Ile	Thr
14	Leu	Leu	Leu	Asn
15	Ser	Ser	Ser	Ala
16	Lys	Lys	Lys	Ala
17	**Asn**	**Ile**	**Leu**	**Leu**
18	Tyr	Tyr	Tyr	Gln
19	His	His	His	Asp
20	Leu	Leu	Ile	Val
21	Glu	Glu	Glu	Arg
22	Asn	Asn	Asn	Glu
23	Glu	Glu	Glu	Leu
24	**Val**	**Ile**	**Leu**	**Leu**
25	Ala	Ala	Ala	Arg
26	Arg	Arg	Arg	Gln
27	Leu	Val	Ile	Gln
28	Lys	Lys	Lys	Val
29	Lys	Lys	Lys	Lys
30	Leu	Leu	Leu	Glu
31	**Val**	**Leu**	**Leu**	**Ile**
32	Cbx	Cbx	Cbx	Cbx

[a] The Roman numerals indicate the number of helices in the native structure. The **a** positions of the heptad pseudorepeats are bold, and the **d** positions are underlined. Gly^{32} of the 2ZTA peptide has been replaced with a Cbx cap. A tetrapeptide segment (Arg-Leu-Leu-Gln) was removed from the NH_2 terminus of the 1CE0 sequence and an acetyl group was added. The original design of the GCN4-based peptides also included a Gly-Glu-Arg segment at its COOH terminus. Glycine appears in all three pdb files; due to the absence of interpretable electron density, glutamic acid appears only in the B strand of the 1CE0 trimer, and arginine does not appear at all. The dipeptide fragment Met-Asp was removed from the NH_2 terminus of the 1VDF sequence and an acetyl group was added. A 14-residue segment (Thr-Phe-Leu-Lys-Asn-Thr-Val-Met-Glu-Cys-Asp-Ala-Cys-Gly) was removed from the COOH terminus of 1VDF in order to match the length of the GCN4-based peptides.

state. Each of the four sequences was threaded into a monomer, dimer, trimer, tetramer, and pentamer. Side chains were built onto the respective backbone using SCRWL3.0 and its backbone-dependent rotamer library.[71] For the construction of side chains on the helix bundles, SCRWL3.0 was given the atomic coordinates of the neighboring helices (steric boundaries) and sequence files written in uppercase letters; thus all side chains were built by SCRWL3.0, not just the mutated ones. Table 6 lists the PDB codes, sequences, and oligomeric order of the four template structures.

The C-terminal region of helix A in the H38-p1 trimer exhibits some deviation from regular coiled-coil structure. Attempts to build various side chains beyond Val$^{27\mathbf{d}}$ (using our numbering scheme) resulted in unusually high bonded energy terms for the side chains in question. Thus, the peptide segment Lys$^{28\mathbf{e}}$-Lys$^{29\mathbf{f}}$-Leu$^{30\mathbf{g}}$-Leu$^{31\mathbf{a}}$ of helix A was rebuilt using ideal ϕ and ψ values.

Molecular Dynamics Simulations. All structures were subjected to 0.8 ns (400 000 steps) Nosé−Hoover MD simulations. Harmonic constraints with a force constant of 1.0 kcal/Å^2 were applied to the backbone α-carbons to keep the structures close to the desired fold. These constraints are necessary because otherwise many non-native oligomeric states would fall apart upon MD simulation. The constraint affects the translational and rotational entropy loss; thus the entropy values calculated below are approximate. The nonbonded interactions were updated every 20 dynamics steps, and the coordinate frames were saved every 1000 steps.

Free Energy Calculations. The data from each 400 000 step MD simulation from every structure was used to evaluate various components of the standard free energy change,

$$\Delta G = \Delta W - T\Delta S \tag{3}$$

where ΔW is the effective energy change, T is the absolute temperature, and ΔS is the configurational entropy change of the system under investigation. For the purpose of effective energy measurements, each MD simulation was divided into a 100 000 step equilibration phase and a 300 000 step production phase.

The configurational entropy can be divided into translational, rotational, and conformational contributions,

$$T\Delta S = T[\Delta S^{\text{trans}} + \Delta S^{\text{rot}} + \Delta S^{\text{conf}}] \tag{4}$$

The $T\Delta S^{\text{trans}}$ term was evaluated from 400 center of mass coordinates of the configurations saved during the MD simulation. It is assumed that the entropy loss calculated for one component helix within an oligomeric variant applies to any other helix within the same helix bundle. Upon oligomerization, all except one of the helices lose some translational entropy as a result of their transition from the standard state (corresponding to 1 M), where each center of mass moved within a 1660 A^3 volume to a smaller volume defined by the range of x, y, and z coordinates where the helix is restricted to move relative to the other helices.[72] The range of values that define this volume were obtained by subtracting the center of mass coordinates of helix A from the center of mass coordinates of the rest of the protein for each frame. The difference between the largest and the smallest of these values defines the size of a $\Delta x\Delta y\Delta z$ volume element. The entropy loss from the translational restriction is

$$\Delta S^1 = R \ln[\Delta x\Delta y\Delta z/1660\ \text{A}^3] \tag{5}$$

This value would be exact if the center of mass coordinates were equally distributed throughout the $\Delta x\Delta y\Delta z$ volume element. The uneven distribution of data points introduces an additional entropy term per spatial dimension into ΔS^{trans},

$$\begin{aligned}\Delta S^x &= -R[\textstyle\int_{\Delta x} p_i(x) \ln p_i(x)\, dx - \int_{\Delta x} p_i^{\text{flat}}(x) \ln p_i^{\text{flat}}(x)\, dx] \\ &= -R[\textstyle\int_{\Delta x} p_i(x) \ln p_i(x)\, dx + \ln \Delta x]\end{aligned} \tag{6}$$

where $p_i^{\text{flat}}(x) = 1/\Delta x$ and $p_i(x)$ is the probability of finding the center of mass within subdivision i. The total translational entropy with contributions from all three spatial dimensions can then be computed as

$$\Delta S^{\text{trans}} = \Delta S^1 + \Delta S^x + \Delta S^y + \Delta S^z \tag{7}$$

The rotational entropy loss, $T\Delta S^{\text{rot}}$, was calculated from the distribution of orientations of helix A from each oligomer relative to the rest of the protein. To obtain the required data, we aligned the protein with respect to the initial minimized structure using the COOR ORIENT RMS command; then we used COOR ORIENT RMS again to align one of the helices. The transpose of the rotation matrix that CHARMM reports for the second alignment is used to determine the three Euler angles that describe the orientation of one helix with respect to the axis of the bundle. From the distribution of Euler angles, we computed the rotational entropy loss,

$$\begin{aligned}\Delta S^{\text{rot}} = &-R[\textstyle\int_{-\pi}^{\pi} p(\phi) \ln p(\phi)\, d\phi - \int_{-\pi}^{\pi} p_\phi \ln p_\phi\, d\phi] - \\ &R[\textstyle\int_{-1}^{1} p(\theta) \ln p(\theta) \sin\theta\, d\theta - \int_{-1}^{1} p_\theta \ln p_\theta \sin\theta\, d\theta] - \\ &R[\textstyle\int_{-\pi}^{\pi} p(\psi) \ln p(\psi)\, d\psi - \int_{-\pi}^{\pi} p_\psi \ln p_\psi\, d\psi]\end{aligned} \tag{8}$$

where p_ϕ, p_θ, and p_ψ are constants corresponding to flat distributions over the angular displacement range of ϕ, θ, and ψ, respectively. In a freely rotating helix, ϕ is uniformly distributed in the range $(-\pi, \pi)$, $\sin\theta$ in $(-1, 1)$, and ψ in $(-\pi, \pi)$. Thus the constants are obtained from $(\int_{-\pi}^{\pi} p_\phi\, d\phi = \int_{-1}^{1} p_\theta \sin\theta\, d\theta = \int_{-\pi}^{\pi} p_\psi\, d\psi = 1)$.

The ΔS^{trans} and ΔS^{rot} values computed above are for one helix. Thus the entropic cost for an oligomer with N chains is $[(N-1)\Delta S^{\text{trans}}/N + (N-1)\Delta S^{\text{rot}}/N]$. It should be noted that the separation of translational and rotational entropy depends on the coordinate system used and is therefore somewhat arbitrary, although the sum of the two is well defined. The ambiguity in the separation of the two is large for highly flexible molecules.[73] Here, the helices are quite rigid, and the separation obtained should be "reasonable".

The S^{conf} term was computed from the probability distribution of each side chain torsional angle, obtained by rotating each of them independently of the others. Except for proline, alanine, and glycine, all side chains from all residues were rotated about their heavy atom (C, N, O, and S) χ bonds one at a time at 10^0 intervals. From the effective energy profiles, we computed the probability of each conformation,

$$p(\omega) = \frac{\exp[-W(\omega)/(RT)]}{\int \exp[-W(\omega)/(RT)]\, d\omega} \tag{9}$$

where ω denotes one of the sampled side chain conformations and $W(\omega)$ is the associated energy. The conformational entropy of a structure is

$$S^{\text{conf}} = -R\Sigma_{N(\chi)}[\textstyle\int p(\omega) \ln p(\omega)\, d\omega] \tag{10}$$

where the outer summation is over the total number $N(\chi)$ of torsional angles in the protein. The calculation was done on 15 structures obtained from the MD run, and the results were averaged. The conformational

(69) Lazaridis, T. *Proteins: Struct., Funct., Genet.* **2003**, *52*, 176−192.
(70) Masunov, A.; Lazaridis, T. *J. Am. Chem. Soc.* **2003**, *125*, 1722−1730.
(71) Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. *Protein Sci.* **2003**, *12*, 2001−2014.
(72) Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 194−208.
(73) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.

entropy loss, ΔS^{conf}, is the difference between the S^{conf} term of the multimer and that of the monomer, which sets ΔS^{conf} of the monomer equal to zero.

The reproducibility of the results was tested by repeating the simulations with different starting random numbers, and we found the differences to be within the error limits of the simulations. Unless otherwise stated, all energetic and entropic terms reported here correspond to averages from four MD runs.

Acknowledgment. Financial support was provided by the NSF (Grant MCB-0615552). Infrastructure support was provided in part by RCMI Grant RR03060 from NIH.

Supporting Information Available: Tables of free energy and its components. This material is available free of charge via the Internet at http://pubs.acs.org.

JA0655284