ENERGETIC BASIS OF COILED COIL TOPOLOGY

AND OLIGOMERIC STATE SPECIFICITY

by

JORGE RAMOS

A dissertation submitted to the Graduate Faculty in Chemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2009

This manuscript has been read and accepted for the
Graduate Faculty in Chemistry in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

	Themis Lazaridis
Date	Chair of Examining Committee
	Mahesh K. Lakshman
Date	Executive Officer

Themis Lazaridis

Dixie Goss

Manfred Philipp
Supervision Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

ENERGETIC BASIS OF COILED COIL TOPOLOGY

AND OLIGOMERIC STATE SPECIFICITY

by

Jorge Ramos

Adviser: Professor Themis Lazaridis

The coiled-coil protein oligomerization motif consists of two or more α-helices oriented parallel or antiparallel, which wrap around each other in a slight left-handed superhelical twist. The typical sequence of a coiled coil is characterized by a heptad repeat commonly denoted by the letters ***abcdefg***, where residues in positions ***a*** and ***d*** are predominantly hydrophobic, while those in positions ***b***, ***c***, ***e***, ***f***, and ***g*** are usually charged or polar. Empirical rules have been established on the tendency of different core sequences to form a certain topology and oligomeric state but the physical forces behind this specificity are unclear. In this thesis we examine the ability of an effective energy function (EEF1.1) to discriminate the correct topology and oligomeric state for a given sequence using a molecular dynamics approach. We find that inclusion of entropic terms 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, while other residue propensities are sequence context dependent. A variety of features contribute to the topological specificity of the motif, including electrostatics, side chain entropy change, steric matching, and the desolvation of hydrophobic side chains. Our results establish that the oligomeric state is dictated by similar rules in both parallel

and antiparallel conformations but alignment of α-helices requires a broader set of both lateral and vertical interaction patterns. We found that the antiparallel topology can be directed by ***a/e'*** electrostatic attractions in the dimer, with ***e/e'*** and ***g/g'*** making minimal contributions. The antiparallel trimer topology is mainly the result of steric matching ***a/e'*** and ***d/g'*** side chain pairs in two antiparallel faces. The antiparallel tetramer is stabilized by similar interactions as the trimer in addition to ***b/e'*** electrostatics, which are only available in this oligomeric state. This work provides useful methodology and rules for designing coiled coils with a well defined and predictable three-dimensional structure.

ACKNOWLEDGMENTS

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I want to thank the Department of Chemistry of the City College of New York for giving me part of the financial support needed during my stay at this institution.

I am deeply indebted to my supervisor Prof. Themis Lazaridis from the City College of New York whose help and suggestions helped me in all the time of research and writing of this thesis. I have furthermore to thank the members of my thesis supervision committee: Prof. Dixie Goss from Hunter College and Prof. Manfred Philipp from Lehman College for their stimulating support. Without them, this research would not have taken a turn for the most strenuous path.

My former colleagues from the Lazaridis Laboratory supported me in my research work. I want to thank them for all their help, support, interest and valuable hints.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

Fig. #

1 INTRODUCTION

In this chapter, the structural properties of coiled coils are introduced. Also the current knowledge regarding coiled coils is reviewed and an overview of this thesis is given.

1.1 Coiled coils in outline

The coiled coil protein oligomerization motif consists of two or more α-helices oriented parallel or antiparallel, which wrap around each other in a slight left-handed superhelical twist. The typical sequence of a coiled coil is characterized by a heptad repeat commonly denoted by the letters ***abcdefg***, [see FIGURE 1.1] where residues in positions ***a*** and ***d*** are predominantly hydrophobic, while those in positions ***b***, ***c***, ***e***, ***f***, and ***g*** are usually charged or polar.[**1,2**] Coiled coils occur in nature as the dominant motif in fibrous proteins and as mediators of oligomerization.

The oligomerization interface between supercoiled helices is made by knobs-into-holes meshing of side chains at the interface. Despite an apparent simplicity at the sequence level, the structural motif exhibits remarkable structural diversity. 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 number of helices in the assembly is mainly a function of ***a/a'*** and ***d/d'*** interactions in lower order oligomers while the specificity shifts towards ***a/g'***and ***d/e'*** interactions for higher order oligomers.[**12**] The helices in a coiled coil can associate with either parallel or antiparallel orientations. Parallel coiled coils have been studied

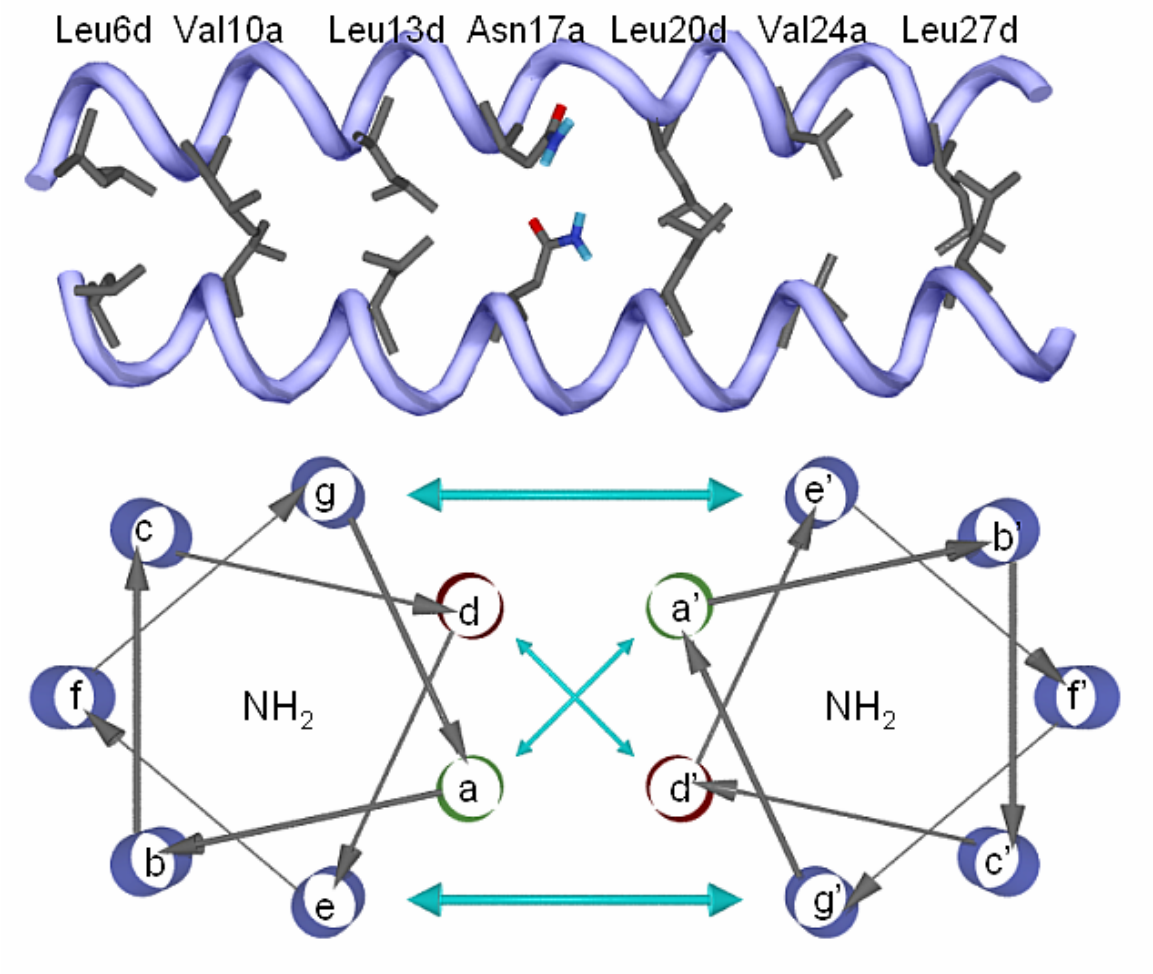


FIGURE 1.1 [**Top**] The ribbon structure of two helices interacting as a parallel coiled-coil. Only the residues at positions ***a*** and ***d*** are shown [PDB CODE 2ZTA]. [**Bottom**] The helical wheel shows a top-view from NH_2 to COOH terminus. The arrows within each circle show the residue connectivity. The arrow thickness decreases for residue positions further into the page. The small crossed arrows at the center denote the interactions of the hydrophobic residues ***a*-*a***$'$ and ***d*-*d***$'$. The large arrows at the top and bottom denote the classical salt bridge interactions between ***g*-*e***$'$ and ***e*-*g***$'$.

more extensively and are relatively well understood, but the interactions that lead to the formation of antiparallel coiled coils are less understood. As a result, it is difficult to predict topology from amino acid sequence alone.

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,[13] MULTICOIL,[14] SCORER,[15] MATCHER,[16] and MARCOIL.[17] Another program (SOCKET) can be used to detect coiled coils in a

given 3D structure.[18] Programs for the prediction of partnering specificity based on favorable interhelical interactions also exist.[19,20] 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]

1.2 Current Knowledge on Oligomeric State Specificity

The packing geometry is different in different oligomeric states.[6] The Cα−Cβ bonds in leucine ***d*** side chains in the GCN4-p1 dimer [PDB CODE 2ZTA] 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 [see FIGURE 1.2]. 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 become possible, and as a result, the helices have the least solvent-accessible surface area of all oligomers.[8]

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

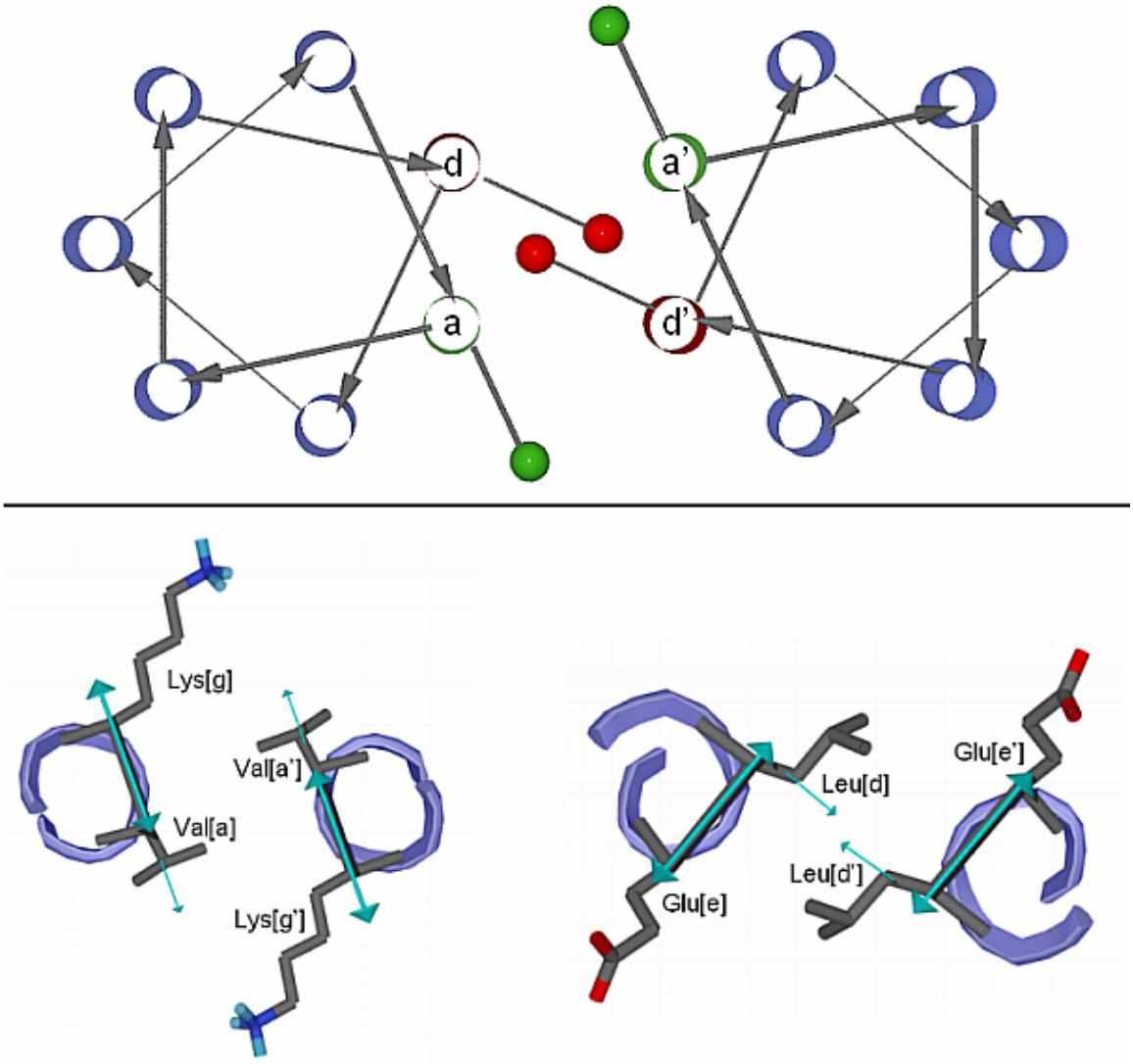


FIGURE 1.2 [**Top**] Helical wheel schematics of knobs-into-holes observed in parallel coiled coil dimers. The Cα−Cβ bonds or "knobs" are shown as small single-headed arrows and the Cα−Cα bonds or "holes" are shown as large double-headed arrows. Knobs ***d*** are perpendicular to holes ***d'-e'***. Knobs ***a*** are parallel to holes ***g'-a'***. [**Bottom left**] Parallel packing of Val at ***a*** [PDB 2ZTA]. [**Bottom right**] Perpendicular packing of Leu at ***d*** [PDB 2ZTA]. Note that all Cα−Cβ bonds are angled towards the NH_2 terminus, thus the terms parallel and perpendicular apply only to coiled coils [not to individual helices].

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[***a***] → Val[***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 Asn16***a***→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 Gln27***d***→Leu27***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,[25,37-39] 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 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***.[9] 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]

1.3 Current knowledge on topological specificity

The alignment, partner specificity, and oligomerization state of a coiled coil are mainly dictated by the interactions that occur among the ***a***, ***d***, ***e***, and ***g*** positions of the component helices. While the oligomeric state is dictated by similar rules in both parallel and antiparallel conformations the alignment of α-helices requires a broader set of both lateral and vertical interaction patterns.[50]

In the case of antiparallel coiled coil dimers, the ***a*** and ***d*** positions of the heptad repeat constitute a predominantly hydrophobic interaction interface where the residues pack in pseudo-equivalent ***a/d*** layers. The flanking residues also have different interaction patterns in parallel as compared to antiparallel coiled coils. In an antiparallel dimer, the residues at positions ***e*** and ***e'*** are on one side and residues at positions ***g*** and ***g'*** are on the other side [see FIGURE 1.3]. The antiparallel dimer is then said to have one ***e***/***e'*** face and one ***g***/***g'*** face while the parallel dimer has two equivalent ***e***/***g'*** faces. Thus a design where the ***e*** and ***g*** positions have opposite polarities would result in adjacent

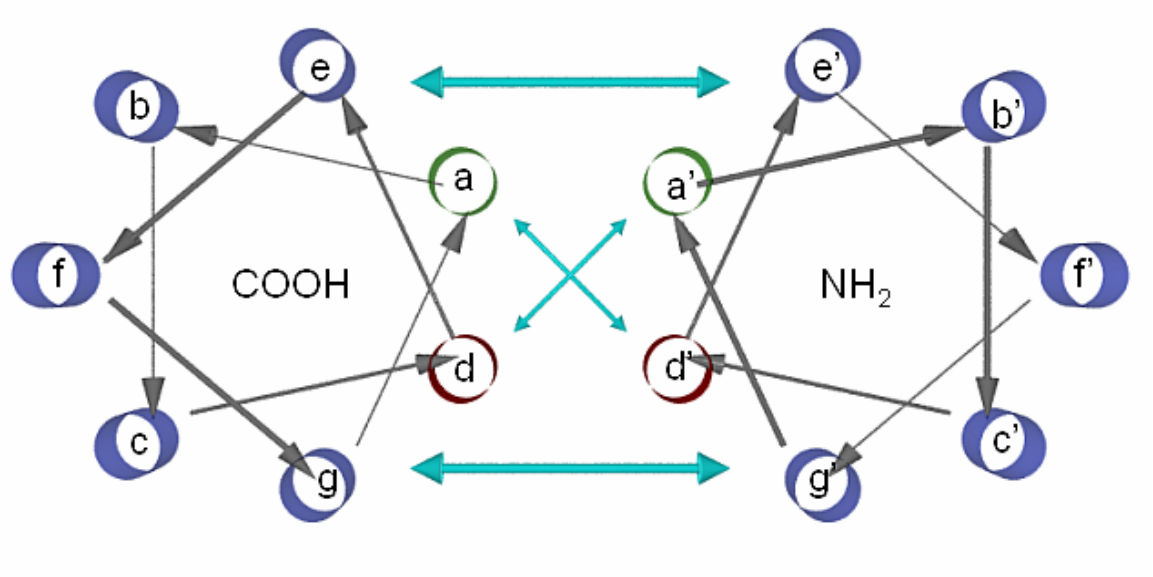


FIGURE 1.3 Helical wheel shows a top-view from COOH to NH_2 terminus in the left and from NH_2 to COOH terminus in the right. The arrows within each circle show the residue connectivity. The arrow thickness decreases for residue positions further into the page. The small crossed arrows at the center denote the interactions of the hydrophobic residues ***a-d'*** and ***d-a'***. The large arrows at the top and bottom denote the classical salt bridge interactions between ***e-e'*** and ***g-g'***.

interchain electrostatic attraction in a parallel coiled coil whereas electrostatic repulsion would result in the antiparallel counterpart [51].

Hodges *et al* [52] synthesized disulfide-bridged coiled coils of parallel or antiparallel topology with identical hydrophobic core residues. They found that the interchain electrostatic repulsions between similarly charged amino acid residues in the ***e*** and ***g*** positions of the parallel coiled coil will destabilize the structure while placing side chains of opposite polarity at different ends of the peptide will stabilize the antiparallel coiled coil due to ***g***/***g'*** and ***e***/***e'*** electrostatic attractions.

The design of homodimers that favor an antiparallel alignment of helices must also consider hydrophobic components. For example, ILE at ***d*** positions are poorly accommodated at the core of dimeric parallel coiled coils,[6] but a favorable packing environment can be created by matching ILE with ALA at some of the core ***a/d*** layers in the antiparallel conformer.[53] A pattern of ILE at ***d*** with ALA at ***a*** creates a cavity in the interior of the parallel homodimer, while it generates a well packed interface in the antiparallel homodimer.[54] The effect of steric matching is more pronounced in the

antiparallel dimer formed by a fragment of the C-terminal domain of the HIV-1 regulatory protein Vpr. The most prominent feature is a pair of TRP[***a***]/HIS[***d***] stacking interactions.[55] In a parallel coiled coil dimer, two TRP sidechains would be side by side, which is destabilizing due to the small space between the two helices.

Naturally occurring antiparallel coiled coil dimers contain polar or charged residues at core positions, which are thought to have great influence in the orientation of the helices. For example, the peptide corresponding to residues 468–497 of the C-terminal domain of the *E. Coli* osmosensor ProP contains an ARG at position ***a*** [56] and the *Agrobacterium tumefaciens* counterpart has one LYS and one ARG at different ***a*** positions. [57] It has been proposed that the pairing of two LYS or ARG at ***a*** positions would result in unfavorable electrostatic interactions but the data does not rule out other explanations. Polar and charged side chains at ***a*** or ***d*** positions of antiparallel coiled coil dimers might exhibit interhelical interactions with charged or polar residues at ***e′*** or ***g′*** positions, respectively. An example of such interaction is ARG488***a*** forming a salt bridge with ASP478***e'*** in the C-terminal domain of the *E. Coli* osmosensor ProP [56]. Cross-linking experiments suggest that an ARG488Ile substitution does not prevent dimerization of the osmosensor but rather may cause reorientation of the coiled coil helices.[58] Another interaction of this type involves a ***d*** position ARG in the helical arm of *E. coli* seryl tRNA synthetase forming a salt-bridge with a GLU residue at a ***g′*** position.[59] In the parallel orientation, these electrostatic interactions would shift from ***d***/***g'*** to ***d***/***e'***. This type of interaction occurs often in natural antiparallel coiled coils and might help specify a particular helix orientation. However, a mutagenesis study failed to see important energetic contributions of salt bridges involving interior residues to coiled coil stability and structure.[60]

Antiparallel coiled coil trimers as autonomous folding domains are very rare. One of the few available structures is an artificial construct made from the GCN4-p1 peptide with a central ASN16***a*** mutated to Ala [61]. The change in orientation was attributed to the cavity present in the parallel trimer. However, if the cavity was occupied by a benzene molecule, the parallel trimer would be stabilized [62]. The destabilization factor of creating a cavity has been estimated at 5.5 kcal/mol, a quantity large enough to switch the helix orientation [61].

The core of antiparallel coiled coil trimers exhibits two types of mixed core layers; *2**a**/**d*** and *2**d**/**a***, as well as three distinct hydrophilic faces (***e**/**g***, ***e**/**e***, and ***g**/**g***). Kennan *et al* have designed antiparallel coiled coil trimers where the choice of side-chain patterning promotes the desired orientation by means of steric matching.[**63**] Coil-Ser was a synthetic peptide designed to form a parallel coiled-coil dimer, but crystallization showed it to be an antiparallel trimer.[**64**] This was attributed primarily to hydrophobic interactions and possibly to helix macrodipole interactions, rotamer energies, and steric effects.

Antiparallel coiled coil tetramers are often characterized by the presence of apolar amino acids at either the ***e*** or ***g*** position in the heptad repeat. The ***a*** and ***d*** side chains pack into pseudo-equivalent (2***d***/2***a***, and 2***a***/2***d***) layers. As in the parallel tetramer, the interaction pattern extends to the ***b*** and ***c*** positions but the antiparallel counterpart forms two pairs of pseudo-equivalent faces, two **g/c** faces and two **b/e** faces. DeGrado *et al* have used those structural features for the design of antiparallel four helix bundles.[**65**] The resulting topology is a trade-off between electrostatic interactions at the hydrophilic faces and packing energy at the core positions. Compared to the parallel topology, side chain packing is much more efficient in the antiparallel coiled coil tetramer.[**66**] The packing efficiency is more likely to make strong contributions when the side chains occupying the ***a*** and ***d*** positions are significantly different in size. Antiparallel coiled coil tetramers such as the heterogeneous nuclear ribonucleoprotein (hnRNP) contain LEU, ILE and VAL at the core positions, thus the packing effect will likely make only minor contributions in determining helix orientation.[**67**]

A single substitution at a solvent-exposed site (GLU20***e***CYS) changed the topology of the tetrameric GCN4-LI peptide[**6**] from parallel to antiparallel.[**66**] It was found that GLU20***e*** preferentially stabilizes the parallel configuration via electrostatic interactions while CYS20***e*** stabilizes the antiparallel configuration in part from improved packing energy. The GLU20***e***Cys mutant seems to have a nearly constant supercoil radius as a result of the more even distribution of side chains afforded by mixed layers.

Recent studies in antiparallel tetrameric coiled coils have identified at least five kinds of interaction patterns directed by hydrophobic repeats of ***a-d*** (1W5K), ***a-d-e*** (2R2V) and ***a-d-g*** (2B22) type.[**68**] These patterns differ markedly from classical knobs-

into-holes packing interactions and the neighboring helices are vertically offset resulting in a variety of interdigitated packing arrangements as opposed to the discrete layers of side-chain interactions found in parallel coiled coils. The inclusion of nonpolar residues at positions ***e*** and ***g*** can also affect the angular orientation of the component helices resulting in the lateral alignment of sidechains at position ***a*** or ***d*** from adjacent helices. This effect can be observed when a large number of sidechains at positions ***e*** and ***g*** are nonpolar and possess certain packing requirements.

Experiments carried out so far have identified very few selectivity mechanisms for antiparallel coiled coil topology. This limited set of rules include **a)** steric complementarity of nonpolar side chains at ***a*** and ***d*** positions as applied to antiparallel dimers and trimers [**63,55**] **b)** the inclusion of polar or charged residues at ***a*** and ***d*** positions help specify both the dimeric state and antiparallel orientation of helices [**56,57**]. **c)** Nonpolar sidechains at ***e*** and ***g*** positions confer specificity for the antiparallel tetramer topology by means of sidechain desolvation and steric complementarity [**68**] while lateral pairing of polar or charged sidechains at such positions help stabilize dimers and trimers with hydrogen bonding and electrostatic interactions (attractive or repulsive) between ionized side chains directing helix orientation [**51,69**]. We must also consider both lateral and vertical interaction patterns to get a more complete rationalization of the structural determinants of coiled coil topology.[**70**]

1.4 Research goals and the subject of this thesis

The research presented in this thesis seeks to rationalize and extend currently accepted empirical rules on the tendency of different coiled coil sequences to form a certain topology and oligomeric state. Using a molecular dynamics approach, we employ the program CHARMM[**71**] and the effective energy function EEF1.1[**72**] for the analysis of various coiled coil sequences known to favor different topologies and oligomeric states. More precisely:

In **Chapter 2**, coiled coil model generation and free energy calculation protocols are explained.

In **Chapter 3**, four sequences are threaded onto the structures of dimeric, trimeric, tetrameric, and pentameric coiled coils. 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.

In **Chapter 4**, we consider six coiled coil sequences that form antiparallel coiled coil dimers, trimers, and tetramers. We build both parallel and antiparallel configurations for these sequences using Crick's mathematical model [**73**] and evaluate their relative stability using molecular dynamics simulations with the effective energy function EEF1.1 [**72**]. Finally, individual residue contributions are computed to obtain new insights that might contribute to the design of coiled coils with a well defined and predictable three-dimensional structure.

2 COILED COIL MODEL GENERATION AND FREE ENERGY CALCULATION PROTOCOL

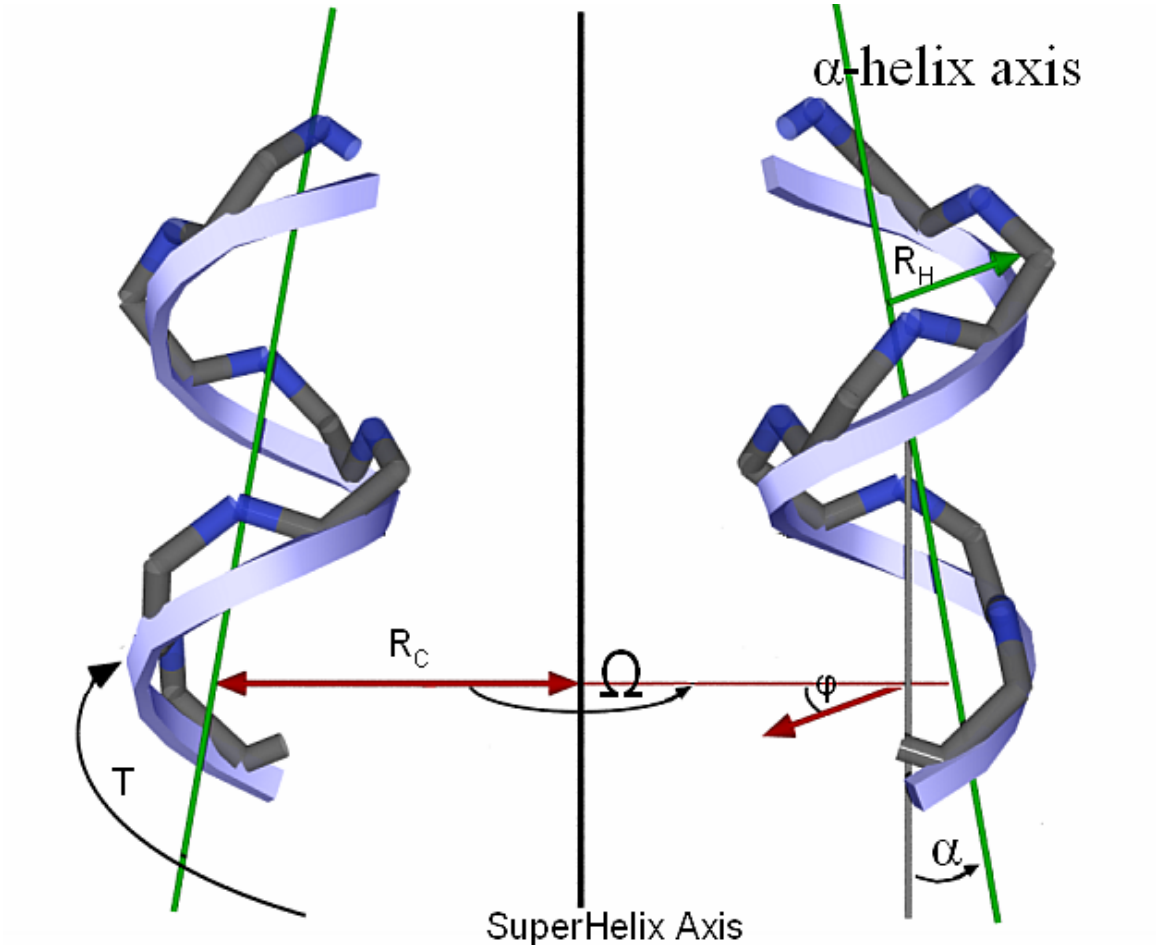
In this chapter, two methods of coiled coil model generation are discussed: threading onto crystal-structure templates and *de novo* building of backbone coordinates from a mathematical model. Also our free energy calculation protocol is introduced and some remarks are made about the limitations of the methods employed.

2.1 Generation of Coiled Coil Structures

The comparison of relative stabilities using effective energy function (EEF1.1) is meaningful only if the structures under investigation have the same number of atoms. The length of the coiled coils that can be studied by computational techniques is limited even further because the sequences being compared must start and end at similar heptad positions. For example, our investigation on the oligomeric state of coiled coils[**12**] employed a severely truncated structure of the five-stranded coiled coil because three- and four-stranded structures of comparable length could not be found in the Protein Data Bank. It was assumed that the truncation did not affect the oligomeric state.

The results from sequence threading onto crystal-structure templates could be biased depending on the differing resolution of the crystal structures. For example, a high resolution crystal structure will likely contain side chains in rotameric states that optimize its free energy. Clearly, a structure where the side chains have been built by a computer algorithm will have less optimized rotamers. As a result, the high resolution structure will give the lowest free energy.

In order to ensure reproducibility of molecular dynamics results is that all structures must be treated equally. In this thesis, all structures have had their side chains built onto the respective backbone using SCRWL3.0 and its backbone-dependent rotamer library.[**74**] For the construction of side chains on the helix bundles obtained from the PDB, SCRWL3.0 is given the atomic coordinates of the neighboring helices (steric boundaries), sequence files are written in upper-case letters; thus all side chains are built by SCRWL3.0, not just the mutated ones.



Initial Helix Parameters

Atom Type	R_H, Å	dZ, Å
N	1.61	-0.05
Cα	2.27	+1.23
C	1.68	+0.36

Coiled Coil Parameters

Oligomeric State	R_C, Å	P_C, Å	Ω
Dimer	4.9	148	180^0
Trimer	6.7	175	120^0
Tetramer	7.6	205	90^0

Figure 2.1 Scheme of the mathematical model for ideal coiled coil construction and definition of the parameters employed. The coiled coil frame rotates in the direction of the parameter **T**, which is obtained from $T = 2\pi\, dZ / P_C$.

The coiled coil pitch P_C, is related to the tilting angle α, of the component helices by $\alpha = 2\pi\, R_C / P_C$.

Conceptually, coiled coil structures are simple: the α-helical coiled coil results from systematic bending of an α-helix such that every seventh residue is structurally equivalent. The distance required for the superhelix to complete one full turn is called the pitch P_C, and the angle between two neighboring helices is called the crossing angle α. Because all Cα−Cβ bonds are angled towards the NH_2 terminus, the core residues will interlock if the vertical displacement of the helices has the right value. Thus the motif can be parameterized and equations for the coordinates of the backbone atoms can be derived.

For the investigation on the energetic determinants of coiled coil topology, [**50**] backbone coordinates were built using Crick's mathematical model for an ideal coiled coil.[**73**] The parameters of the model are summarized in Figure 2.1. For simplicity, we will illustrate our general approach using a parallel dimeric coiled coil as an example. A straight, 3.5 residues/turn helix is generated along the Z axis. The 3.5 residue/turn periodicity positions each ***a*** and ***d*** residue such that they line up along the helical axis rather than spiraling around the helix as in a canonical α-helix. The atomic coordinates X_i, Y_i, and Z_i, for the backbone atoms in the straight helix are given by,

$$X_i = R_H \sin(\theta t + \varphi)$$
$$Y_i = R_H \cos(\theta t + \varphi)$$
$$Z_i = Z_{i-1} + dZ \qquad (2.1)$$

Where the value of θ is 4π/21 radians, the parameter t runs over integral values generating three atoms per residue, and the values of R_H and dZ depend on the atom type (N, Cα, C). The parameter φ is chosen to position the ***a*** and ***d*** residues toward the center of the coiled coil and is determined by the heptad assignment of the first residue in the sequence.

The helix generated by Equation 2.1 is then translated along the X axis by a value equal to the coiled coil radius and given a left-handed twist using Crick's Equations for an ideal coiled coil. The second helix is generated by translating the initial helix along the X axis by a value equal to the coiled coil radius, rotating it by 180^0 about the coiled coil axis and giving it a left-handed twist creating the dimeric coiled coil. The atomic coordinates X_C, Y_C, and Z_C, of the coiled coil backbone atoms are given by,

$$X_C = [R_C \cos (T + \Omega)] + [X_i \cos (T + \Omega)]$$
$$+ [Y_i \cos \alpha \sin (T + \Omega)]$$
$$Y_C = [-R_C \sin (T + \Omega)] - [X_i \sin (T + \Omega)]$$
$$+ [Y_i \cos \alpha \cos (T + \Omega)]$$
$$Z_C = Z_i + Y_i \sin \alpha \qquad (2.2)$$

Where the angular and spatial separation of the component helices in the coiled coil are denoted by Ω and R_C, respectively. The parameter α is the tilting angle of the component helices and is obtained by $\alpha = 2\pi R_C / P_C$, where P_C is the coiled coil pitch. The angular displacement of the initial helix in the coiled coil frame is given by $T = 2\pi (Z_H - Z_0) / P_C$. The angular separation Ω for a dimeric coil takes on values of 0^0 and 180^0 for helix A and B respectively. For trimeric coils, Ω has values of 0^0, 120^0 and 240^0, while in tetrameric coils Ω has values of 0^0, 90^0, 180^0, and 270^0.

The antiparallel coiled coils are built in the same manner as their parallel counterparts but the initial helix parameters are reversed and the Z displacement offset by a multiple of 1.54 Å for helices that run in the opposite sense to ensure interlocking of ***a*** and ***d*** residues. For example, helix B in the antiparallel dimer is axially displaced relative to helix A so that each turn of helix B runs roughly midway between two adjacent turns in helix A.

2.2 Energy function

The relative stability of a structure is related to the free energy change,

$$\Delta G = \Delta W - T\Delta S \qquad (2.3)$$

where ΔW is the effective energy change and TΔS is the configurational entropy change. The effective energy term is evaluated by the energy function EEF1.1,

$$W = H_{intra} + \Delta G^{solv} \qquad (2.4)$$

where H_{intra}, and ΔG^{solv} are the intramolecular energy and the solvation free energy respectively.[72] In EEF1 the ΔG^{solv} term is approximated as a sum of contributions from all the atoms in the macromolecule,

$$\Delta G^{solv} = \sum_i \Delta G_i^{solv} = \sum_i \Delta G_i^{ref} - \sum_i \sum_{j \neq 1} \boldsymbol{f}_i(\mathbf{r}_{ij}) \boldsymbol{V}_j \qquad (2.5)$$

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 $\sum_{j \neq 1} \boldsymbol{f}_i(\mathbf{r}_{ij})\boldsymbol{V}_j$ accounts for the exclusion of solvent around group **i** due the shielding by groups **j**. EEF1.1 is an updated parameterization[75] based on potentials of mean force calculated in explicit solvent.[76]

For the purpose of free energy calculations, all structures are subjected MD simulations using the *Nose-Hoover* algorithm[77] with nonbonded interactions being updated every 20 dynamics steps and the coordinate frames saved every 1000 steps. Structures are allowed to equilibrate during the first 100,000 steps and the production phase starts beyond that point. The reported effective energy is the average value measured during the production phase. We find a simulation length of about 500,000 steps [1 ns] to be sufficient to obtain a good estimate of the effective energy.

2.3 Entropic terms

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

$$T\Delta S = T\,[\Delta S^{trans} + \Delta S^{rot} + \Delta S^{conf}] \qquad (2.6)$$

The $T\Delta S^{trans}$ term is evaluated from all of the 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.[78] 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

$$\Delta S^1 = R \ln [\Delta x\Delta y\Delta z / 1660\ A^3] \qquad (2.7)$$

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 datapoints introduces an additional entropy term per spatial dimension into ΔS^{trans},

$$\Delta S^{x} = -R[\int_{\Delta x} p_i(x)\ln p_i(x)\, dx - \int_{\Delta x} p_i^{flat}(x)\ln p_i^{flat}(x) dx]$$
$$= -R[\int_{\Delta x} p_i(x)\ln p_i(x)\, dx + \ln \Delta x] \qquad (2.8)$$

where $p_i^{flat}(x) = 1/\Delta x$ 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:

$$\Delta S^{trans} = \Delta S^{1} + \Delta S^{x} + \Delta S^{y} + \Delta S^{z} \qquad (2.9)$$

The rotational entropy loss, $T\Delta S^{rot}$, is 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 align the protein with respect to the initial minimized structure using the COOR ORIENT RMS command, then we use 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 compute the rotational entropy loss,

$$\Delta S^{rot} = -R[\int_{-\pi}^{\pi} p(\phi)\ln p(\phi) d\phi - \int_{-\pi}^{\pi} p_\phi \ln p_\phi d\phi]$$
$$-R[\int_{-1}^{1} p(\theta)\ln p(\theta)\sin\theta d\theta - \int_{-1}^{1} p_\theta \ln p_\theta \sin\theta d\theta]$$
$$-R[\int_{-\pi}^{\pi} p(\psi)\ln p(\psi) d\psi - \int_{-\pi}^{\pi} p_\psi \ln p_\psi d\psi] \qquad (2.10)$$

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 $[(N-1)\Delta S^{trans}/N + (N-1)\Delta S^{rot}/N]$. 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

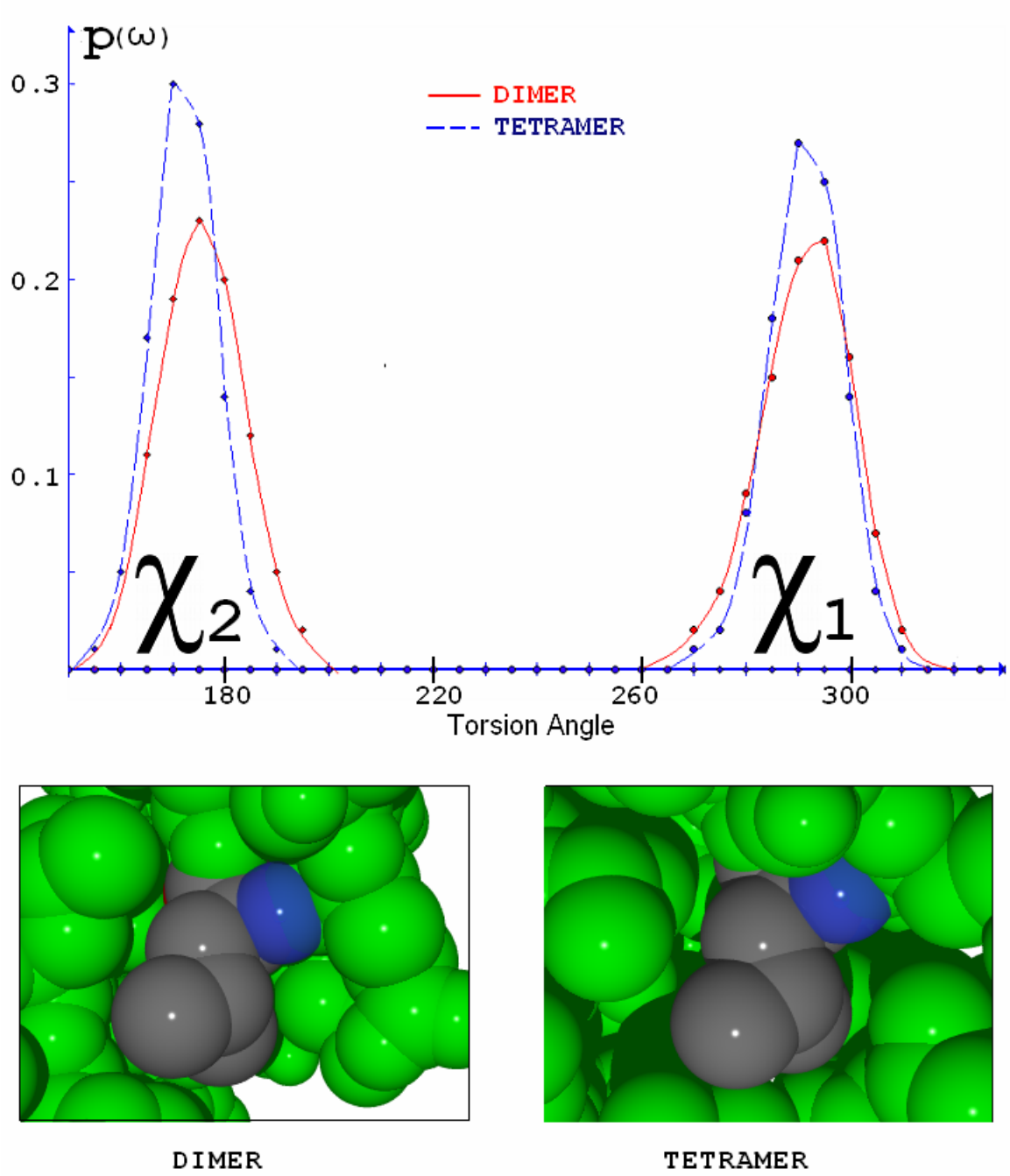


Figure 2.2 [**Top**] The probability distribution of χ1 and χ2 for Leu13***d*** in the dimer and tetramer models of PDB 2ZTA. The conformational space of Leu13***d*** in the tetramer is more restricted as compared to the dimer as shown by a narrower distribution of the torsion angles. The result is a side chain entropy loss [compared with the unfolded state] of 0.31kcal/mol for the dimer and 0.53 kcal/mol for the tetramer. [**Bottom**] Space fill models of dimer and tetramer structures viewed from NH_2 to COOH terminus. Only one Leu13***d*** is shown [colored in CPK mode]. Note that the Leu13***d*** in the dimer structure has more room around it.

defined. The ambiguity in the separation of the two is large for highly flexible molecules.[79] Here, the helices are quite rigid, and the separation obtained should be "reasonable".

The S^{conf} term is 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 are 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 compute the probability of each conformation,

$$p(\omega) = \frac{\exp[-W(\omega)/RT]}{\int \exp[-W(\omega)/RT] d\omega} \qquad (2.11)$$

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

$$S^{conf} = -R \sum_{N(\chi)} [\int p(\omega) \ln p(\omega) d\omega] \qquad (2.12)$$

where the outer summation is over the total number $N(\chi)$ of torsional angles in the protein. Doing this calculation every 20,000 steps [25 measurements per 1 ns] seems sufficient to obtain a good estimate of the conformational entropy. The conformational entropy loss ΔS^{conf} is the difference between the S^{conf} term of the multimer and the monomer, thus the ΔS^{conf} of the monomer is equal to zero.

Our conformational entropy calculations yield the expected results. We note that as the side chains become increasingly buried, they lose more conformational entropy. The effect is more apparent for side chains that participate in knobs-into-holes interactions [see Figure 2.2].

3 ENERGETIC DETERMINANTS OF COILED COIL OLIGOMERIC STATE SPECIFICITY

In this chapter, we employ the program CHARMM[71] and the effective energy function EEF1.1[72] 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.

3.1 Related theoretical work

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.[80] 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.[81-83] Analysis of packing energies of leucine and isoleucine in dimers and tetramers showed the importance of packing in determining oligomeric state.[83] Crick's parametrization for the backbone together with rotamer enumeration was used to predict coiled coil structures of different oligomeric states.[84] 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.[85] 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.[86] 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[87] and different Monte Carlo methods for calculation of the partition functions.[88-89] 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.

3.2 Structures and protocols

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 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.[74] 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 upper-case letters, thus all side chains were built by SCRWL3.0, not just the mutated ones. Table 3.1 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 VAL27***d*** (using our numbering scheme) resulted in unusually high bonded energy terms for the side-chains in question. Thus, the peptide segment LYS28***e***-LYS29***f***-LEU30***g***-LEU31***a*** of helix A was rebuilt using ideal ϕ and ψ values.

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

Table 3.1 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. GLY32 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. GLY appears in all three pdb files; due to the absence of interpretable electron density GLU appears only in the B strand of the 1CE0 trimer, and ARG 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 fourteen-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.

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 here are approximate. The nonbonded interactions were updated every 20 dynamics steps, and the coordinate frames were saved every 1000 steps, which gives 400 frames per simulation run.

All frames were used to evaluate the translation and rotation entropy. For the purpose of effective energy and conformational entropy measurements, each MD simulation was divided into a 100 000 step equilibration phase and a 300 000 step production phase. The effective energy was measured for all frames obtained from the production phase [300 frames] while the conformational entropy was measured every twenty frames [15 frames total], and the results were averaged.

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.

3.3 Results in outline

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 3.2. 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 3.2). 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.

TABLE 3.2 Free energy estimates for all sequences studied[a]

Oligomeric order	I	II	III	IV	V
GCN4-P1		**A**			
Average 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		**B**			
Average 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		**C**			
Average 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		**D**			
Average 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, translational and rotational entropy terms are averages during 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.

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 to the experimental value.[90] 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 3.3, 3.4, 3.5 and 3.6), 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). 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.

3.4 GCN4-p1 dimer

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 3.2. LEU side chains at all ***d*** positions exhibit the characteristic perpendicular packing mode while VAL side chains at ***a*** positions exhibit parallel packing except for VAL31***a***, where the deviation from parallel is evident in helix 1. We found that this side-chain tends to flip during dynamics. Experimentally, VAL31***a*** amide protons have been found to exchange much faster than its counterparts at other ***a*** positions.[**91**] This is most likely the result of fraying at the helix end making the packing geometry more flexible.

We observe that all LEU at ***d*** sites contribute favorably to dimer specificity (Table 3.3, underlined entries). However, the relative stabilization is context-dependent (LEU6***d***: 1.1, LEU13***d***: 1.0, LEU20***d***: 0.4, and LEU24***d***: 0.3 kcal/mol). The free energy decomposition indicates that their dimer specificities arise from different contributions. LEU6***d***, LEU20***d*** and LEU27***d*** favor the dimer in terms of VDW and BOND. LEU13***d*** 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.

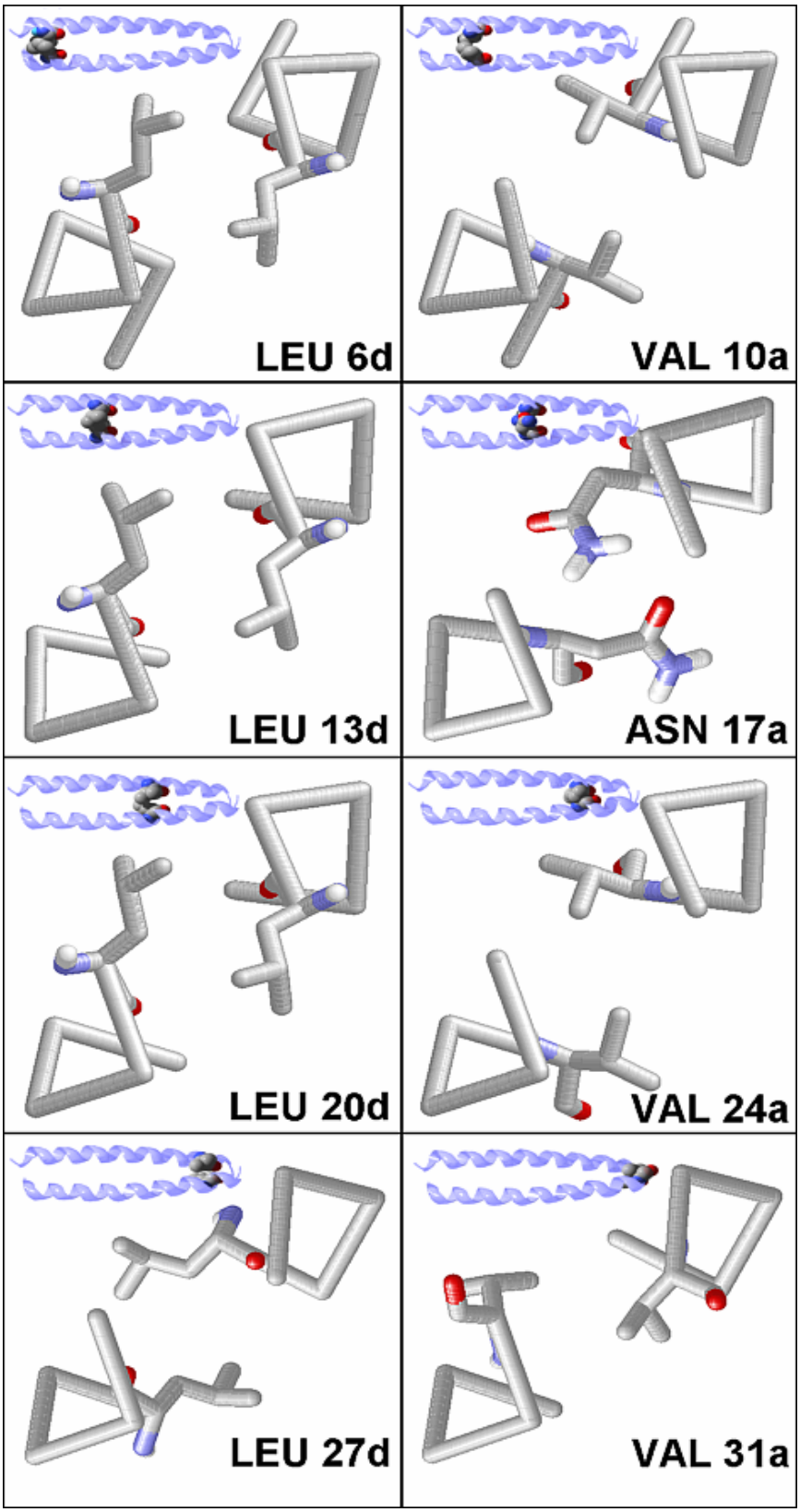


Figure 3.1. Structural details of the ***a*** and ***d*** residues of the GCN4-p1 dimer viewed from NH_2 to COOH-terminus. Images at the upper right corners indicate the location of the side chain within the protein.

ASN17***a*** is highly conserved in bZIP transcription factors and is thought to impart dimeric specificity because the orientation of the ASN 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 ASN adopts an asymmetric conformation. Goodman and Kim have reported that amide proton exchange is much faster at the region near ASN17***a*** 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 ASN switches rapidly between two conformations.[**91**] In agreement with the NMR data of the Jun leucine zipper, the two ASN17***a*** 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 ASN17***a*** 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.

LYS16***g*** and GLU21***e*** 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 GLU and LYS in the higher oligomers. Higher order structures allow the formation of LYS16***g***-GLU21***e'*** 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 GLU with GLN increased the stability of the coiled coil.[**92**] A similar situation exists for the LYS28***e***-GLU23***g'*** 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. ARG2***g*** favors dimer in all three GCN4-based peptides. The dimer specificity is 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.

Table 3.3 Average Free Energy ($W - T\Delta S^{conf}$) Per Residue for the GCN4-p1 Sequence during Four 0.6 ns (300 Frames) Nose-Hoover MD Simulations[a]

RESIDUE	I	II	III	IV	V
Ace	-4.04	-3.31	-4.58	-4.60	-4.58
Arg 2g	4.04	3.39	6.29	6.93	5.34
Met 3a	-1.97	-1.62	-1.18	-0.75	-1.94
Lys 4b	0.86	1.26	1.05	0.88	0.88
Gln 5c	-8.50	-8.61	-8.49	-8.28	-8.27
Leu 6d	-4.21	-6.92	-5.90	-5.49	-6.00
Glu 7e	-3.49	-4.53	-4.12	-3.66	-4.67
Asp 8f	-10.47	-10.84	-10.64	-10.33	-10.82
Lys 9g	-1.17	-0.99	-0.66	-0.56	-0.34
Val 10a	-2.53	-4.63	-4.82	-4.30	-4.47
Glu 11b	-4.15	-4.79	-4.45	-3.97	-4.43
Glu 12c	-4.45	-4.86	-4.44	-4.06	-4.40
Leu 13d	-4.68	-7.70	-6.08	-6.78	-7.18
Leu 14e	-5.78	-5.81	-5.86	-6.45	-5.40
Ser 15f	-8.21	-7.96	-7.81	-8.41	-7.98
Lys 16g	0.14	0.71	1.52	1.04	1.51
Asn 17a	-11.57	-11.19	-12.22	-10.81	-10.68
Tyr 18b	-1.42	-1.57	-1.18	-0.51	-0.83
His 19c	-3.32	-3.28	-3.41	-3.70	-4.14
Leu 20d	-4.52	-6.60	-6.31	-6.49	-5.92
Glu 21e	-4.19	-4.63	-4.36	-4.60	-3.68
Asn 22f	-11.63	-11.35	-11.49	-11.14	-11.12
Glu 23g	-3.97	-3.74	-4.16	-4.04	-3.99
Val 24a	-2.61	-3.68	-4.46	-3.87	-4.40
Ala 25b[b]	-5.51	-6.39	-6.21	-5.93	-6.14
Arg 26c	1.24	1.87	1.96	1.76	2.14
Leu 27d	-4.03	-6.81	-6.45	-6.35	-6.70
Lys 28e	1.30	0.59	1.10	1.33	1.33
Lys 29f	1.75	1.61	1.85	1.65	2.71
Leu 30g	-1.78	-1.70	-1.70	-2.40	-2.86
Val 31a	2.20	2.17	1.08	1.32	0.31
Cbx	-1.41	-1.39	-1.44	-1.62	-1.97
$\sum[W - T\Delta S^{conf}]$	-108.08	-123.32	-117.58	-114.15	-118.69

[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 Ala25***b*** 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 Ala25***b*** and Lys29***f***.

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 LYS9***g*** in the trimer sequence (Table 3.4); 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 LYS9***g*** in the H38-p1 sequence must be attributable to the core substitutions LEU13***d***→THR and LEU6***d***→GLN but its exact origin is unclear.

Small amounts of stability and specificity for dimer come from solvent-exposed sites, such as GLN5***c***, ASP8***f***, GLU11***b***, GLU12***c***, and TYR18***b***. Their origin has not been analyzed in detail, although it could. For example, TYR18***b*** 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 LYS16***g***' in higher oligomers.

3.5 H38-p1 trimer

Figure 3.2 shows the core interface details of the H38-p1 trimer structure after dynamics. In the crystal structure, LEU side-chains exhibit threefold, propeller-like symmetry (acute packing); their Cα-Cβ bonds point towards the solvent while their -Cγ-$(C\delta)_2$ units are directed towards the core. ILE exhibit similar packing symmetry with their -Cγ's pointing towards the solvent and the -Cγ-Cδ units directed towards the hydrophobic core. Even though the trimeric specificity of the H38-p1 peptide was reproduced in our calculations (Table 3.2), not all the structural features of the crystal structure were maintained in the simulations. For example, it was found that the hydrogen bond network involving the GLN6***d*** side-chains changes after dynamics. Shu *et al.* have suggested that GLN6***d*** imparts trimer specificity due to the formation of a three-fold symmetric hydrogen bond network between the side chain amide and the backbone carbonyl of the opposite helix.[27] We observed that the hydrogen bond 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 GLN.

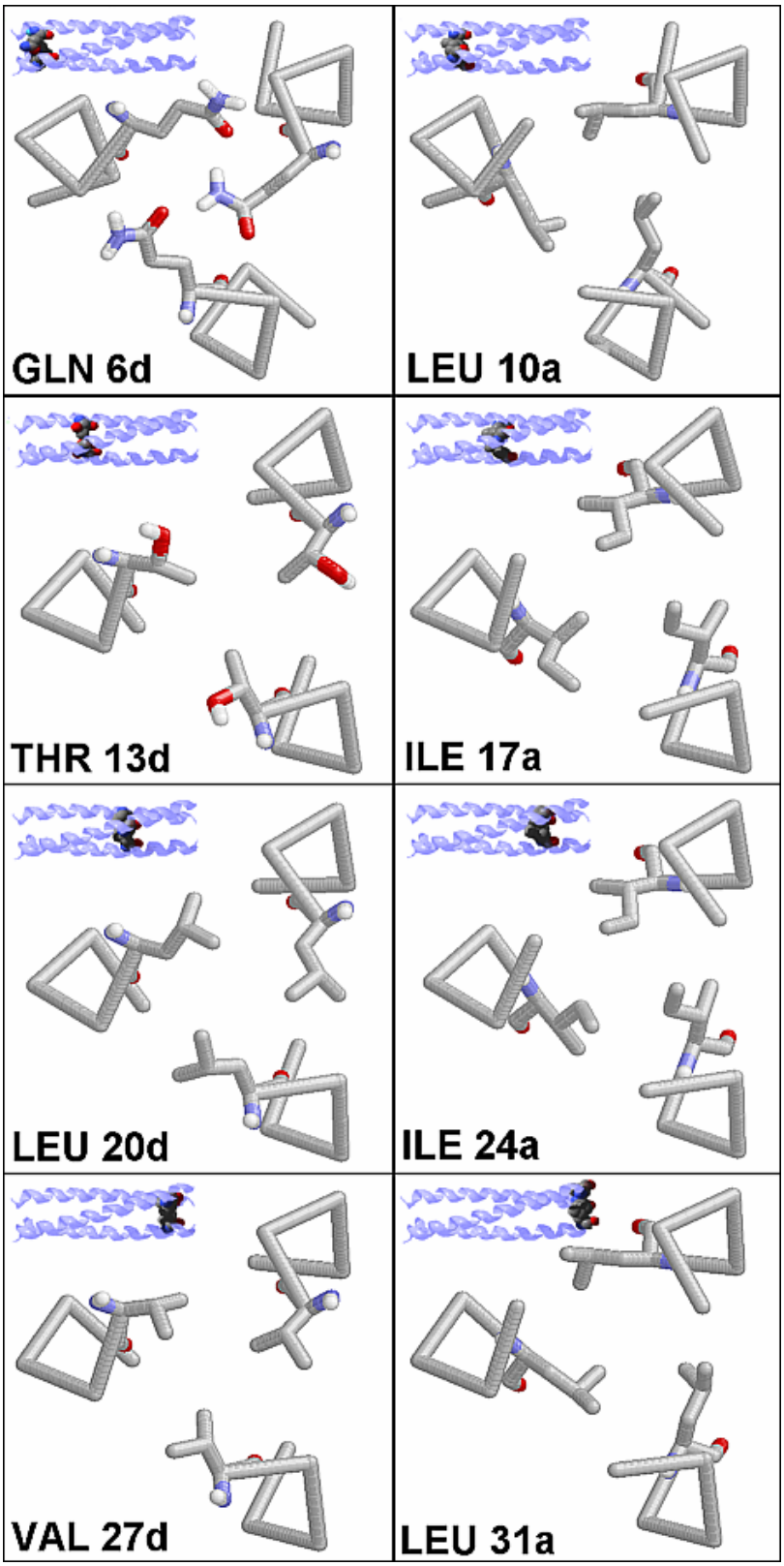


Figure 3.2. Structural details of the ***a*** and ***d*** residues of the H38-p1 trimer viewed from NH_2 to COOH-terminus. Images at the upper right corners indicate the location of the side chain within the protein.

As a result, one of the GLN side chains moves its amide group into the core to fill that space. We found that GLN6***d*** 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 3.4, underlined entries). VAL27***d*** 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*. VAL has a weak preference for the parallel packing exhibited by the ***d*** sites of the tetramer.[7] Thus the oligomeric preference of VAL 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 VAL at ***d*** interacts. For example, VAL20***d*** in the COMP sequence favors the tetramer.

ILE3***a***, ILE17***a***, and ILE24***a*** favor trimer by 0.7, 1.0 and 1.0 kcal/mol respectively. ILE3***a***, and ILE24***a*** derive their specificity from a combined effect of VDW and bonded energy, while the situation is more complex for ILE17***a***. ILE 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 ILE rotamer there is only one position that directs the -Cγ-Cδ unit towards the complementary cavity in the correct orientation and that is position ***a***. Even with this packing constraint, one out of three ILE in the H38-p1 tends to flip during some simulations, which suggests that the core in this oligopeptide is poorly packed.

THR13***d*** has been proposed as critical for H38-p1 trimerization because it is part of the HIV-1 gp41 core.[27] A LEU → THR mutation at position 13***d*** in the GCN4-pVL peptide changed the oligomeric distribution from a mixture of dimers and trimers to mostly trimers.[35] Nevertheless, our calculations indicate that THR13***d*** favors the tetramer. It is possible that THR at ***d*** is not trimer specific but simply less dimer specific than LEU. Alternatively, THR may favor trimer by affecting the contributions of neighboring residues. For example, LEU14***e*** was tetramer specific in the GCN4-p1 sequence and is trimer specific in the H38-p1 sequence.

Table 3.4 Average Free Energy ($W - T\Delta S^{conf}$) Per Residue for the H38-p1 Sequence during Four 0.6 ns (300 Frames) Nose-Hoover MD Simulations[a]

RESIDUE	I	II	III	IV	V
Ace	-4.67	-3.46	-4.91	-4.80	-4.65
Arg 2g	3.79	2.10	3.71	3.79	3.72
Ile 3a	-0.20	-0.98	-1.32	-0.07	-0.79
Lys 4b	0.58	1.35	0.80	1.03	0.90
Gln 5c	-8.16	-8.20	-8.43	-7.78	-8.41
Gln 6d	-9.33	-10.37	-10.86	-10.84	-10.45
Glu 7e	-4.03	-3.95	-4.10	-3.79	-4.43
Asp 8f	-10.62	-10.87	-10.63	-10.44	-10.87
Lys 9g	-2.35	-1.22	-2.45	-2.45	-1.69
Leu 10a	-4.30	-4.22	-6.24	-6.28	-6.21
Glu 11b	-4.81	-4.92	-4.74	-4.21	-4.90
Glu 12c	-4.95	-5.47	-5.13	-4.83	-4.98
Thr 13d	-4.85	-5.16	-5.35	-5.65	-5.32
Leu 14e	-5.24	-5.24	-5.53	-5.05	-4.00
Ser 15f	-8.11	-7.75	-7.78	-8.13	-8.09
Lys 16g	0.03	0.64	1.18	1.23	1.22
Ile 17a	-3.11	-3.12	-4.66	-4.02	-3.71
Tyr 18b	-1.30	-1.19	-1.21	-0.23	-0.71
His 19c	-3.51	-3.67	-3.45	-3.68	-4.05
Leu 20d	-4.76	-7.36	-6.93	-7.08	-6.53
Glu 21e	-3.56	-3.26	-3.49	-4.18	-2.97
Asn 22f	-11.28	-11.43	-11.40	-11.15	-10.87
Glu 23g	-4.10	-3.79	-4.29	-5.01	-4.09
Ile 24a	-1.89	-1.57	-3.77	-3.22	-3.46
Ala 25b	-5.61	-6.16	-6.13	-6.01	-6.09
Arg 26c	1.22	1.95	1.17	1.22	1.40
Val 27d	-2.71	-3.45	-4.74	-4.64	-4.27
Lys 28e	1.36	0.73	1.07	1.31	1.59
Lys 29f	1.96	1.94	1.88	1.84	2.00
Leu 30g	-1.92	-1.94	-2.05	-3.15	-3.09
Leu 31a	1.21	1.94	-0.03	-1.24	-1.41
Cbx	-1.39	-1.26	-1.56	-1.66	-1.91
$\sum$[W - $T\Delta S^{conf}$]	-106.62	-109.35	-121.36	-119.17	-117.13

[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.

Overall, among the 8 sequence substitutions going from GCN4-p1 to H38-p1, the most important for the switch from dimer to trimer seem to be the three LEU at ***d***, which favor the dimer while their replacements either favor the trimer or disfavor the dimer.

3.6 GCN4-LI tetramer

Figure 3.3 shows the core interface details of the native GCN4-LI tetramer after dynamics. The hydrophobic core remains unchanged after dynamics.

Three out of four ILE at ***d*** sites show the expected tetramer preference: ILE6***d***, ILE13***d***, and ILE20***d***. 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. ILE27***d*** exhibits a slight preference for pentamer. One of the ILE27***d*** 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.

LEU at position ***a*** in the GCN4-LI sequence destabilizes dimers and trimers. Two LEU residues at ***a*** positions favor the tetramer (LEU17***a*** and LEU31***a***), while LEU10***a*** and LEU24***a*** favor pentamer by a small amount. Thus, in the tetramer sequence the preferred ILE rotamer at position ***d*** directs tetramer formation in a slightly context dependent manner while LEU at position ***a*** acts as a destabilizing factor for lower order oligomers, also in a context dependent manner.

LEU14***e*** favors tetramer by 1.3 kcal/mol in terms of VDW. This is likely due to interactions with ILE13***d*** from a neighboring chain. GLU21***e*** 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 LYS16***g*** and/or HIS19***c*** from a neighboring helix. The change in behavior of this residue may be an indirect effect of the replacement of ASN17***a***. GLU23***g*** also favors the tetramer, which exhibits low BOND and VDW terms, but less desolvation than the pentamer.

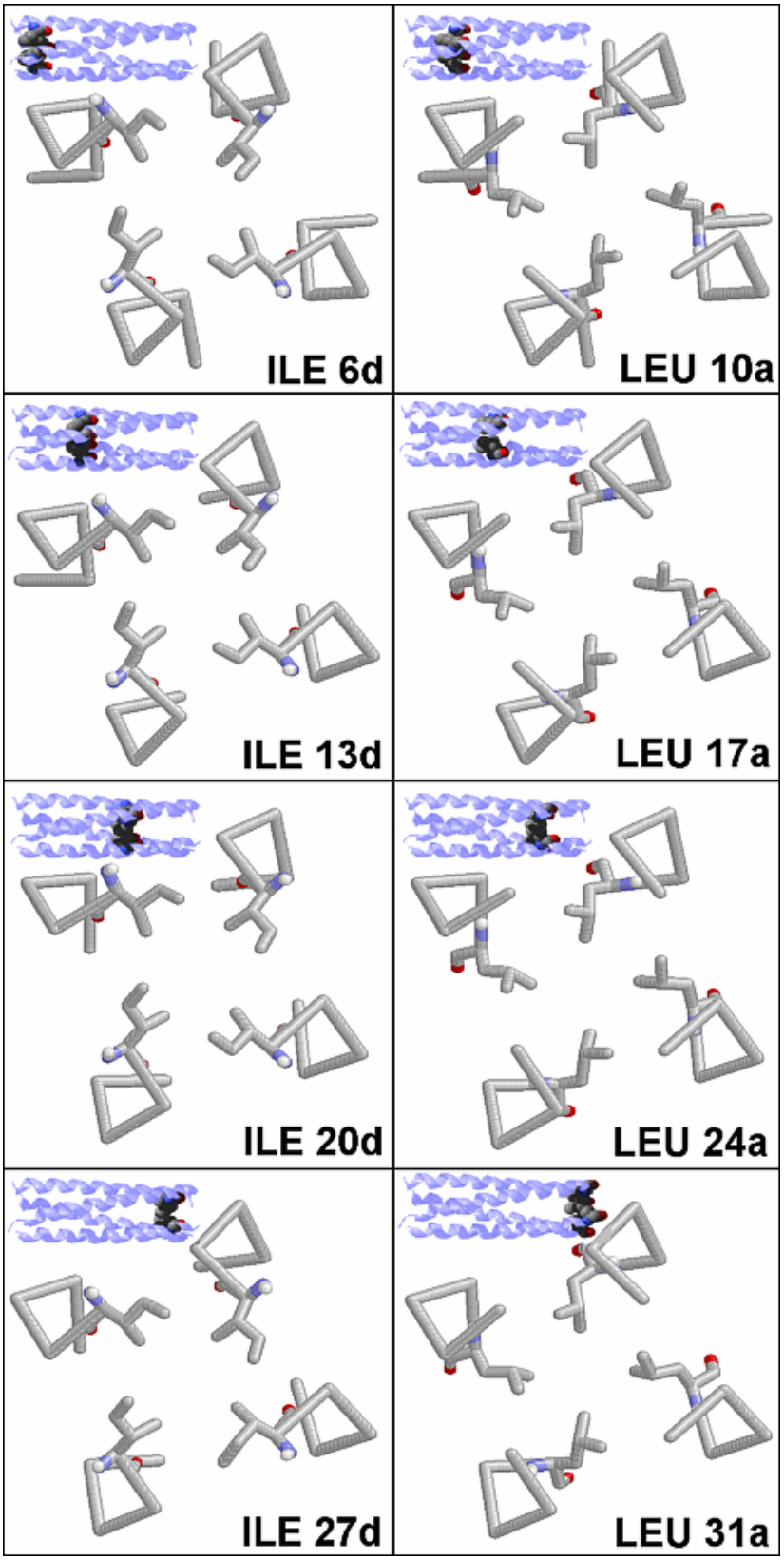


Figure 3.3. Structural details of the ***a*** and ***d*** residues of the GCN4-LI tetramer viewed from NH_2 to COOH-terminus. Images at the upper right corners indicate the location of the side chain within the protein.

Table 3.5 Average Free Energy ($W - T\Delta S^{conf}$) Per Residue for the GCN4-LI Sequence during Four 0.6 ns (300 Frames) Nose-Hoover MD Simulations[a]

RESIDUE	I	II	III	IV	V
Ace	-4.78	-3.39	-4.45	-4.56	-4.30
Arg 2g	5.27	4.10	5.85	6.49	6.11
Met 3a	-2.12	-1.23	-0.54	-1.22	-1.65
Lys 4b	0.02	1.55	0.78	0.50	0.89
Gln 5c	-9.21	-8.75	-8.43	-8.31	-8.46
Ile 6d	-1.68	-3.20	-3.91	-3.92	-3.66
Glu 7e	-4.07	-4.10	-4.06	-3.72	-4.20
Asp 8f	-10.19	-10.72	-10.52	-10.14	-10.38
Lys 9g	-1.40	-1.41	-0.50	-0.86	-0.24
Leu 10a	-4.17	-4.87	-5.71	-7.18	-7.33
Glu 11b	-4.73	-4.70	-5.01	-4.55	-4.96
Glu 12c	-4.32	-4.76	-4.53	-4.18	-4.45
Ile 13d	-2.21	-3.04	-4.39	-4.97	-4.41
Leu 14e	-5.72	-6.05	-6.14	-6.53	-5.50
Ser 15f	-8.06	-7.93	-7.73	-7.95	-7.93
Lys 16g	-0.14	-0.32	1.38	0.66	0.43
Leu 17a	-5.56	-6.49	-7.09	-7.71	-7.56
Tyr 18b	-1.36	-1.44	-1.33	-0.60	-1.01
His 19c	-3.53	-3.47	-3.35	-3.78	-4.47
Ile 20d	-2.40	-2.52	-4.23	-4.95	-4.62
Glu 21e	-3.38	-3.78	-3.56	-4.21	-3.41
Asn 22f	-11.30	-11.56	-11.51	-11.08	-10.87
Glu 23g	-3.82	-3.85	-3.90	-4.67	-3.87
Leu 24a	-4.15	-4.63	-5.74	-7.31	-7.73
Ala 25b	-5.71	-6.29	-6.31	-6.23	-6.39
Arg 26c	0.89	1.41	1.63	1.23	1.74
Ile 27d	-1.61	-2.54	-4.22	-4.21	-4.22
Lys 28e	1.60	0.86	0.65	1.01	1.13
Lys 29f	1.80	2.06	1.33	1.82	1.86
Leu 30g	-2.09	-2.06	-1.96	-3.09	-3.17
Leu 31a	0.77	2.04	-0.14	-1.60	-1.56
Cbx	-1.35	-1.20	-1.58	-1.68	-1.94
$\sum$[W - $T\Delta S^{conf}$]	-98.70	-102.28	-109.22	-117.52	-116.15

[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.

3.7 COMP pentamer

The pentameric COMP variant exhibits the least disruption of the hydrophobic core upon molecular dynamics. As can be seen from Figure 3.4, the side chains at the pentamer core have the highest level of symmetry as compared to other structures. One exception is GLN27***d***, whose amide group flips during the simulations. GLN27***d*** was built by SCWRL3.0 within a symmetric hydrogen bond network with its amide groups pointing towards the COOH-terminus. The crystal structure contains a chloride ion trapped between five GLN and the hydrogen bond network is symmetric[**10**] but if the ion is replaced by all-trans retinol, four of the GLN amides point towards the NH_2-terminus.[**93**] Thus the conformation of GLN27***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 GLN27***d*** to be 1.4 kcal/mol mainly from ELEC. The side chain entropy loss for GLN27***d*** is the largest in the pentamer. This means that the GLN27***d*** hydrogen bond network is very stable regardless of the resulting conformation. However, the oligomeric specificity of GLN at ***d*** might depend on the sequence context. For example, GLN at ***d*** exhibited trimer specificity in H38-p1. Based on sequence analysis GLN27***d*** is supposed to be important since it is conserved in COMP from several sources.[**10**] Nevertheless, the GLN27***d***→LEU27***d*** mutation increases the temperature of melting up to above 120^0C.[**93**, **94**]

Consistent with the previous observations on GCN4-based peptides, LEU10***a***, LEU17***a***, and LEU24***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's usually lowest for the dimer.

PRO is seldom found in coiled coil sequences probably due to its helix-breaking property. PRO4***b*** is slightly tetramer specific due to bonded and ELEC energies (VDW favors the pentamer) but it favors lower order oligomers the least (Table 3.6).

MET6***d*** slightly favors the pentamer by a combined effect of SOLV, BOND, and ELEC. MET3***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

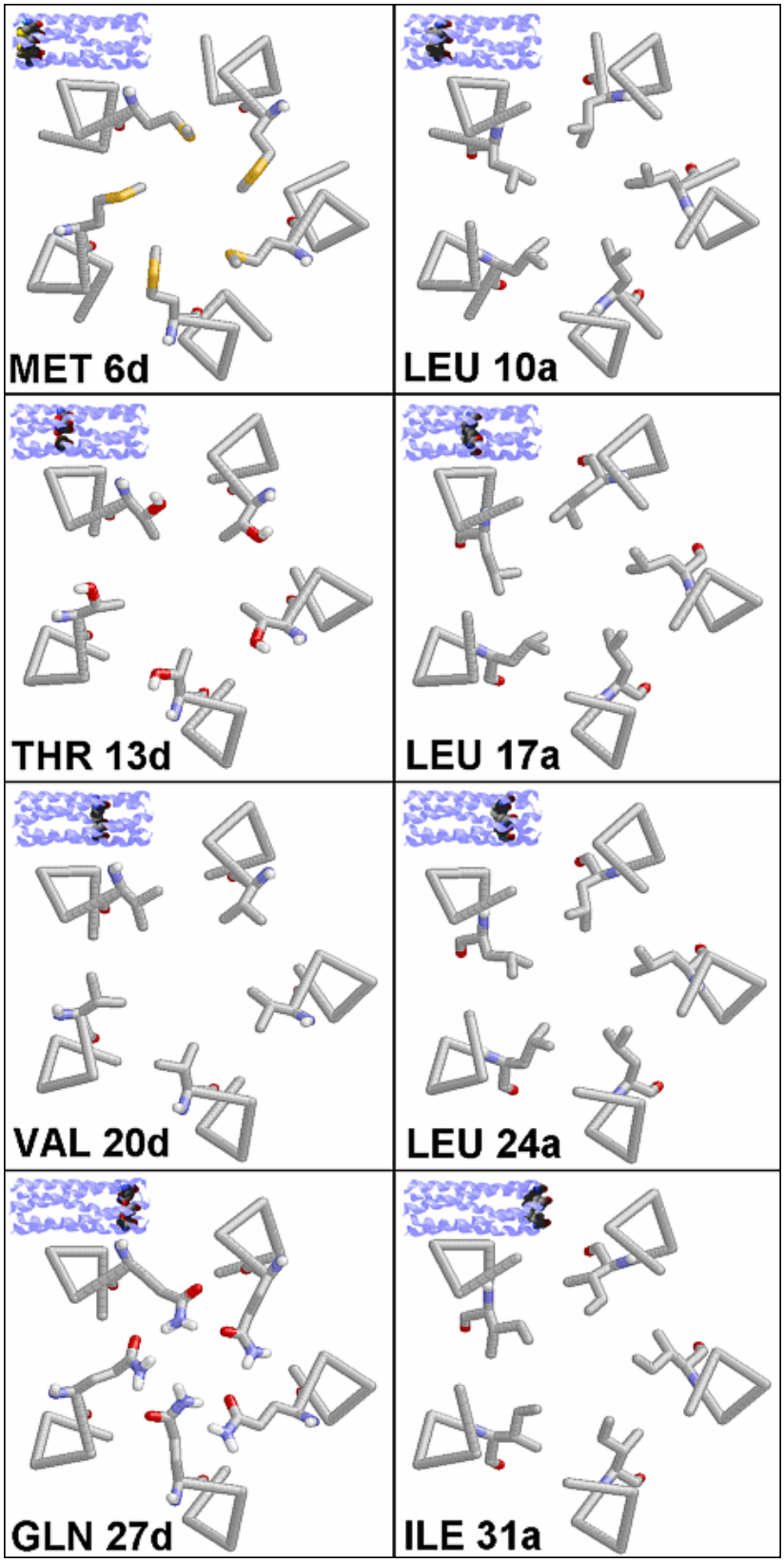


Figure 3.4. Structural details of the ***a*** and ***d*** residues of the COMP pentamer viewed from NH_2 to COOH-terminus. Images at the upper right corners indicate the location of the side chain within the protein.

MET at the core seems to favor pentamer. The other two residues at ***d*** positions (THR13***d*** and VAL20***d***) do not favor pentamer. THR13***d*** favors trimer and the specificity comes mostly from 0.7 kcal/mol ELEC. This is the effect of interactions of THR13***d*** Oγ with the ASN14***e*** Nδ from a neighboring helix; this residue also favors trimer in terms of ELEC. VAL20***d*** favors tetramer; the source of specificity is not clear but the BOND term is lowest in the tetramer.

The VDW term of ALA16***g*** provides most of its 0.7 kcal/mol pentamer specificity, which means that ALA at ***g*** provide the best fit for a kind of extended knobs-into-holes interaction in the pentamer.[**10**] ALA16***g*** interacts with LEU17***a*** from another helix, which is also pentamer specific. A similar pattern exists for the ALA3***a***- LEU2***g'*** pair where the ALA residue shows a relatively high VDW stabilization term in the pentamer. The inverse interaction, observed in LEU23***g***- LEU24***a'***, is also pentamer specific. The fact that LEU2***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 LEU23***g*** with that of LEU24***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 ASN14***e***, which favors pentamer in terms of VDW but favors the trimer in terms of ELEC. ARG21***e*** has the advantage of a long nonpolar segment that allows it to serve as a knob while retaining the advantage of electrostatic interactions. ARG21***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 LEU7***e***, LEU23***g*** and VAL28***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 LEU at position ***g*** in the pentamer. These residues are part of an extended core and their behavior indicates

Table 3.6 Average Free Energy ($W - T\Delta S^{conf}$) Per Residue for the COMP Sequence during Four 0.6 ns (300 Frames) Nose-Hoover MD Simulations[a]

RESIDUE	I	II	III	IV	V
Ace	-3.16	-3.33	-3.75	-3.68	-3.71
Leu 2g	-3.27	-2.92	-2.71	-2.42	-2.74
Ala 3a	0.09	-0.41	-0.48	0.14	-0.93
Pro 4b	5.56	5.48	5.25	4.84	4.98
Gln 5c	-8.04	-8.38	-8.46	-8.46	-8.65
Met 6d	-4.51	-3.87	-4.33	-4.09	-4.57
Leu 7e	-5.05	-4.76	-5.23	-5.07	-5.23
Arg 8f	1.46	1.55	1.44	2.21	1.84
Glu 9g	-5.91	-6.04	-6.36	-6.28	-5.65
Leu 10a	-4.57	-4.92	-6.29	-6.71	-6.76
Gln 11b	-8.65	-8.44	-8.60	-8.49	-8.94
Glu 12c	-5.20	-5.42	-5.52	-4.94	-5.43
Thr 13d	-4.57	-4.39	-6.06	-5.23	-5.34
Asn 14e	-10.59	-10.99	-11.43	-10.56	-11.26
Ala 15f	-6.72	-6.81	-6.68	-6.64	-6.18
Ala 16g	-6.56	-7.12	-7.20	-7.42	-7.92
Leu 17a	-4.15	-4.78	-5.09	-5.55	-5.64
Gln 18b	-8.99	-9.20	-9.16	-8.50	-8.95
Asp 19c	-10.33	-10.17	-10.63	-10.45	-9.65
Val 20d	-3.07	-4.01	-4.35	-5.27	-4.92
Arg 21e	2.48	2.36	3.30	2.54	1.73
Glu 22f	-4.44	-4.24	-4.44	-4.20	-4.31
Leu 23g	-4.95	-5.27	-5.20	-6.27	-7.06
Leu 24a	-4.40	-5.37	-6.46	-6.86	-6.88
Arg 25b	2.95	2.23	2.44	3.62	3.00
Gln 26c	-9.63	-8.74	-9.50	-10.01	-9.25
Gln 27d	-8.00	-7.63	-9.67	-9.31	-10.24
Val 28e	-1.49	-2.27	-2.09	-2.98	-4.01
Lys 29f	1.85	1.90	1.89	1.31	2.40
Glu 30g	-1.38	-1.78	-1.60	-1.77	-0.95
Ile 31a	3.57	3.53	2.45	2.16	1.69
Cbx	-1.48	-1.29	-1.48	-1.58	-2.03
$\sum$[W - $\mathbf{T\Delta S^{conf}}$]	-121.15	-125.51	-135.98	-135.94	-141.56

[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.

that VDW interactions of the types ***a-g***'and ***d-e***' are one of the main determinants of pentamer specificity.

3.8 Discussion

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

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.[92] This needs to be kept in mind when we compare our findings to experiment.

In many cases the physical origin of the small energy differences observed are 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 TYR residue could come from the backbone, the nonpolar part of the sidechain, or the hydroxyl group. Such an analysis could be done in the future for residues of special interest.

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

(1) LEU 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]

(2) ILE at ***a*** confers trimer specificity.

(3) LEU at ***a***, favors tetramers, disfavors lower oligomers and should be indicative of tetramer or pentamer.

(4) VAL at ***a*** favors trimer. We do not have a lot of data for VAL at ***d***. In one case it favors trimer and in another tetramer. Interestingly, VAL is found at ***a*** as often as at ***d*** positions in other trimeric coiled coils.[95]

(5) ILE at ***d*** favors 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.
(7) MET at ***a*** or ***d*** favors pentamer. Interestingly, MET***a*** happens to mark the end of the dimerization domain in GCN4-p1.[5] However, a designed peptide with PHE at all core positions forms a pentamer and when one PHE is changed to MET, it forms a tetramer.[96] Perhaps there is context dependence in the oligomeric propensity of MET.
(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 trade-off between [SOLV+BOND+CONF] and [VDW+ELEC]; if the latter terms outweigh the former the specificity will shift towards higher oligomers.
(9) Bulky side chains at positions ***c*** and ***b*** should destabilize higher oligomers for steric reasons. One example is TYR18***b*** 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$^{17\boldsymbol{a}}$ 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.[7] Mutation to ALA, ABU, or GLN also had the same effect.[30] However, these mutation data only show that ALA, ABU, VAL or LEU at ***a*** are less dimer specific or more trimer specific than ASN. The absolute specificity is not revealed by mutation experiments. Interestingly, ASN at ***a*** was not sufficient to impart dimer specificity in the nonpolar environment of a membrane.[29] This is consistent with our result of ELEC favoring trimers. Perhaps the SOLV contribution is underestimated in our implicit model and this leads to ASN favoring trimers also in an aqueous environment.

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, GLN6***d*** imparts trimer specificity to the H38-p1 peptide but the energetic advantage of trimer is very small compared to 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 GLN. In the absence of a water molecule the distance between these side chains will change and affect the calculation results of EEF1.1.

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 LYS9***g***, LYS16***g*** and GLU23***g*** → ALA in the GCN4-LI peptide). This and many other results from this work are experimentally testable.

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.

4 ENERGETIC DETERMINANTS OF COILED COIL TOPOLOGY

In this chapter, we use the Effective Energy Function (EEF1.1) to provide an energetic basis for the tendency of six coiled coil peptides to align their α-helices in antiparallel orientation. In addition, we examined the effect of mutations known to disrupt the topology of these peptides.

4.1 Structures and Protocols

In this work, backbone coordinates were built using Crick's mathematical model for an ideal coiled coil.[73] Side chains were built with SCRWL3.0 and its backbone-dependent rotamer library [74] using the atomic coordinates of the neighboring helices as steric boundaries. The parameters of Crick's mathematical model and our general approach were summarized in Section 2.1.

Structural analysis and figures were done using the minimized average structures. The average of the atomic coordinates from each simulation was taken and the resulting structure subject to 300 steps of ABNR minimization.

The agreement between the minimized average models and the X-Ray or minimized average NMR structures is quite good with the RMSD of the backbone atoms being: 1R48: 2.44, 1X9V: 1.24, 1COS: 1.06, 1RB4: 1.01, 1W5K: 1.23, and 1TXP: 1.55 angstroms. The 1R48 dimer has a relatively large RMSD because the minimized average NMR is apparently bent towards the ***b***/***e*** surface while our model was built with ideal coiled coil parameters.[56]

Dimer and tetramer structures were subjected to 1 ns (500 000 steps) Nose-Hoover MD simulations. The trimer simulation times were extended to 1.8 ns (900 000 steps). Harmonic constraints with a force constant of 2.0 kcal/Å^2 were applied to the backbone α-carbons to keep the structures close to the desired fold. The nonbonded interactions were updated every 20 dynamics steps, and the coordinate frames were saved every 1000 steps.

The relative stability of a structure is related to the free energy change as obtained from equation 2.3. The configurational entropy would normally be approximated as a

sum of translational, rotational, and conformational contributions. However, we are comparing structures with the same number of helices and different topology. Under these conditions, the rotation and translation entropy can be assumed to be invariant. Then the configurational entropy change ΔS can be approximated by the change in conformational side chain entropy contribution ΔS^{conf} alone (the term is obtained from equation 2.12).

4.2 Results in outline

TABLE 4.1 Free energy estimates for all sequences studied. The effective energy is an average during two 0.8 ns (400 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.

Topology		Parallel	Anti Parallel
1R48 DIMER	**A**		
Average Eff. Int. (400,000 steps)		-150.57	-155.52
$-T\Delta S^{conf}$		6.44	5.95
Free Energy		-144.13±0.29	-149.56±0.31
1X9V DIMER	**B**		
Average Eff. Int. (400,000 steps)		-67.45	-67.94
$-T\Delta S^{conf}$		4.23	2.76
Free Energy		-63.22±0.16	-65.18±0.17
1RB4 TRIMER	**C**		
Average Eff. Int. (900,000 steps)		-129.08	-128.39
$-T\Delta S^{conf}$		9.64	6.87
Free Energy		-120.01±0.26	-121.51±0.26
1COS TRIMER	**D**		
Average Eff. Int. (900,000 steps)		-141.61	-141.15
$-T\Delta S^{conf}$		11.01	6.10
Free Energy		-130.61±0.26	-135.05±0.25
1W5K TETRAMER	**E**		
Average Eff. Int. (400,000 steps)		-117.67	-120.54
$-T\Delta S^{conf}$		13.59	11.76
Free Energy		-104.07±0.24	-108.78±0.24
1TXP TETRAMER	**F**		
Average Eff. Int. (400,000 steps)		-121.35	-121.78
$-T\Delta S^{conf}$		12.45	9.28
Free Energy		-108.90±0.26	-112.50±0.29

The overall free energy differences between parallel and antiparallel constructs are given in table 4.1. The antiparallel topology is correctly predicted for the six sequences studied, although the free energy difference between the parallel and antiparallel conformations is small. Tables 4.2-4.7 give the decomposition of the free energy into residue contributions. We start by singling out those residues that favor the configuration in question. We denote parallel and antiparallel preference with a positive and negative ΔΔG respectively. Then the effective energy (INTE) term from the residue of interest is decomposed into contributions from the interaction of that residue with each residue in the protein. The effective energy is further decomposed into van der Waals (VDW), electrostatics (ELEC), solvation (SOLV), and bonded free energy (BOND) to correlate with the structural features observed in the average structure.

4.2 Antiparallel dimers (1R48 and 1X9V)

1R48 dimer: The peptide corresponds to residues 468–497 of the C-terminal domain of the osmosensor ProP. Our model contains 30 residues, from GLY2***g*** to HIS31***a***, in addition to the ACE and CBX caps. The antiparallel topology is 5.54 kcal/mol lower in energy than the parallel one (Table 4.1A). The side chain entropy makes a contribution of -0.49 kcal/mol to the observed structural preference.

The results for individual residue contributions are presented on table 4.2. The combined ILE6***d*** and HIS31***a*** packing energy is by far the largest determinant for the antiparallel alignment of the osmosensor ProP. The energetic contribution of ILE6***d*** is very large (-4.37 kcal/mol) with a substantial BOND term. We did energy calculations on a peptide model made of two copies of the heptad where ILE6***d*** is located and the results were similar. Thus we concluded that the unfavorable BOND term of ILE6***d*** is a result of the sequence context, not location (end or center) in the coiled coil. This means that the sequence around ILE6***d*** cannot be accommodated in the parallel dimer unless the backbone geometry deviates significantly from the ideal parameters.

Further decomposition of the BOND term of ILE6***d*** identified the χ1 bonds as the source of the highest energy (0.984 and 1.243 kcal/mol for χ1 bond from helix A and B, respectively). Both χ1 bonds have much lower energies in the antiparallel structure.

Table 4.2 Average Free Energy (W - $\mathbf{T}\Delta S^{conf}$) per residue for the 1R48 dimer sequence during two 0.8 ns (400 frames) Nosé-Hoover MD simulations[a]

	II	anti_II	ΔΔG
Ace	-3.95	-4.51	-0.56
Gly 2g	-6.12	-6.87	-0.75
Gly 3a	-6.23	-6.02	0.22
Asp 4b	-8.51	-8.98	-0.47
Asn 5c	-10.71	-10.79	-0.09
Ile 6d	1.46	-2.91	-4.37
Glu 7e	-2.81	-2.79	0.01
Gln 8f	-8.43	-8.63	-0.20
Lys 9g	-0.61	-0.80	-0.19
Ile 10a	-1.99	-2.63	-0.63
Asp 11b	-10.57	-10.62	-0.05
Asp 12c	-10.18	-10.23	-0.05
Ile 13d	0.02	-1.52	-1.55
Asp 14e	-8.21	-8.24	-0.04
His 15f	-2.45	-2.65	-0.19
Glu 16g	-3.98	-4.31	-0.33
Ile 17a	-4.04	-3.21	0.83
Ala 18b	-5.71	-5.44	0.27
Asp 19c	-10.09	-9.47	0.62
Leu 20d	-5.35	-3.41	1.94
Gln 21e	-9.85	-8.87	0.98
Ala 22f	-5.56	-5.96	-0.41
Lys 23g	1.07	0.16	-0.91
Arg 24a	0.73	0.27	-0.46
Thr 25b	-4.06	-4.31	-0.26
Arg 26c	2.07	1.15	-0.92
Leu 27d	-5.12	-2.85	2.28
Val 28e	-1.70	-1.06	0.64
Gln 29f	-6.98	-6.68	0.30
Gln 30g	-6.73	-6.40	0.32
His 31a	1.69	0.34	-1.35
Cbx	-1.25	-1.30	-0.05
$\sum$[W - $\mathbf{T\Delta S^{conf}}$]	-144.13	-149.56	-5.44

[a] All entries are in units of Kcal/mol•helix. The first and second column are the results for parallel and antiparallel topology, respectively third. The third column (ΔΔG) is the free energy that favors a particular helix orientation, negative values for antiparallel and positive values for parallel. 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.

Placing two ILE at ***d*** side by side in the parallel coiled coil dimer induces steric repulsion, which in turn makes the ILE side chains assume high energy rotamers about the $\chi 1$ bond.

This will stabilize alternative residue pairing; in the 1R48 sequence ILE6***d*** prefers a position next to HIS31***a*** and this makes the helices assume the antiparallel alignment. The trend is that any residue that is not beta branched will be preferred instead of ILE at the neighboring position. A good match can also be introduced at an analogous position in a modified sequence, which would yield heterodimers. This fact has been used successfully to design a sequence that binds the coiled coil domain of the essential yeast septin Cdc12p to form a parallel heterodimeric coiled coil. [97] Here, the residue pair ILE[***d***]-ARG[***d'***] favors the formation of heterodimers because the alternative homodimeric structures would put two ILE at ***d*** or two ARG at ***d'*** side by side.

ILE13***d*** favors antiparallel by 1.55 kcal/mol. This residue does not stand out as much as ILE6***d*** but obeys the empirical rule of beta-branched side chains being poorly packed at position ***d*** of parallel coiled coil dimers. The stabilization of the antiparallel topology comes mostly from VDW with BOND making little or no contribution. The packing mode of ILE at ***d*** in parallel coiled coil dimers can also be part of a negative design element by interfering with the alignment (for salt bridge formation) of ***e*** and ***g*** position residues near by.

For ILE at ***a*** we observe context dependence. ILE10***a***, favors the antiparallel conformer by 0.63 kcal/mol, while ILE17***a*** disfavors the conformer by 0.83 kcal/mol. The main difference between these two residues is that ILE10***a*** is part of the hole where HIS31***a*** is buried resulting in an enhanced VDW term. If it were not for its interaction with HIS31***a***, both residue pairs ILE10***a*** and ILE17***a*** would favor the parallel orientation. Thus, a sequence that places ILE at ***a*** and LEU at ***d*** in contact with each other is not useful for the design of an antiparallel homodimer. This is due to the preference of parallel coiled coil dimers to have ILE at ***a*** and LEU at ***d***.

The average minimized structure exhibits a ***g/g'*** electrostatic interaction between LYS9***g*** and GLN30***g*** in the neighboring helix. In the parallel conformer, LYS9***g*** has GLU7***e*** from the neighboring helix within interaction distance but the badly packed ILE6***d*** imposes a steric barrier (recall that ILE at ***d*** is unfavorable to the parallel dimer topology). LYS9***g*** favors the antiparallel conformer by 0.19 kcal/mol mostly from ELEC. This side chain loses 1.45 kcal/mol more entropy in the antiparallel than in the parallel structure. This is indicative of the high stability of the LYS9***g***-GLN30***g'*** electrostatic

bridge. GLN30***g*** shows little orientation preference in part because both helix orientations allow favorable electrostatic interactions.

LYS23***g*** favors the antiparallel structure by 0.91 kcal/mol. ARG24***a*** and LYS23***g'*** come into close proximity of each other in the parallel structure resulting in unfavorable electrostatic interactions (See figure 4.1B). LYS23***g*** also has the advantage of electrostatic interactions with ASP12***c*** and GLU16***g*** from the neighboring helix in the antiparallel conformer. Both LYS23***g*** and GLU16***g*** experience a loss of side chain entropy and they can be seen interacting in the antiparallel minimized average structure, which indicates this interaction is very stable. Overall, for GLU16***g***, LYS23***g*** and ARG24***a***, we see that the 1R48 sequence not only provides favorable interactions in the antiparallel conformer but also provides an unfavorable environment in the parallel conformer (See figure 4.1C and D).

ARG24***a*** favors the antiparallel conformation by 0.46 kcal/mol. One of the main contributions comes from interaction with ASP14***e*** from the neighboring helix and is mostly from VDW and ELEC. It has been suggested that an ARG24***a*** to ILE mutation might switch the topology of 1R48 to parallel.[**58**] We computed the effective interaction for ILE24***a*** with ASP14***e*** in a model of the mutant peptide. For this interacting amino acid pair, the only stabilizing interaction in the antiparallel model is a small VDW term. These results indicate that ARG24***a*** directs the antiparallel topology of the 1R48 dimer by electrostatic interactions while ILE24***a*** stabilizes the parallel topology of the ARG24***a*** to ILE mutant because of an enhanced VDW term. The main electrostatic interactions of ARG24***a*** are summarized in figure 4.1A.

We noted that the polar groups of the two ARG24***a*** do not interact but point in opposite directions and away from the protein core. This is contrary to the assumption that two ARG at ***a*** in the parallel conformer will result in unfavorable electrostatic interactions. According to our calculations, ARG24***a*** in the parallel model makes little or no contribution to the free energy of dimerization and the electrostatic effect of placing two ARG at **g** side by side is negligible. This was inferred from the free energy difference between the dimer and an isolated helix. The fact that an ARG24***a*** to ILE mutation might switch the topology of 1R48 to parallel can be due in part to the preference of parallel

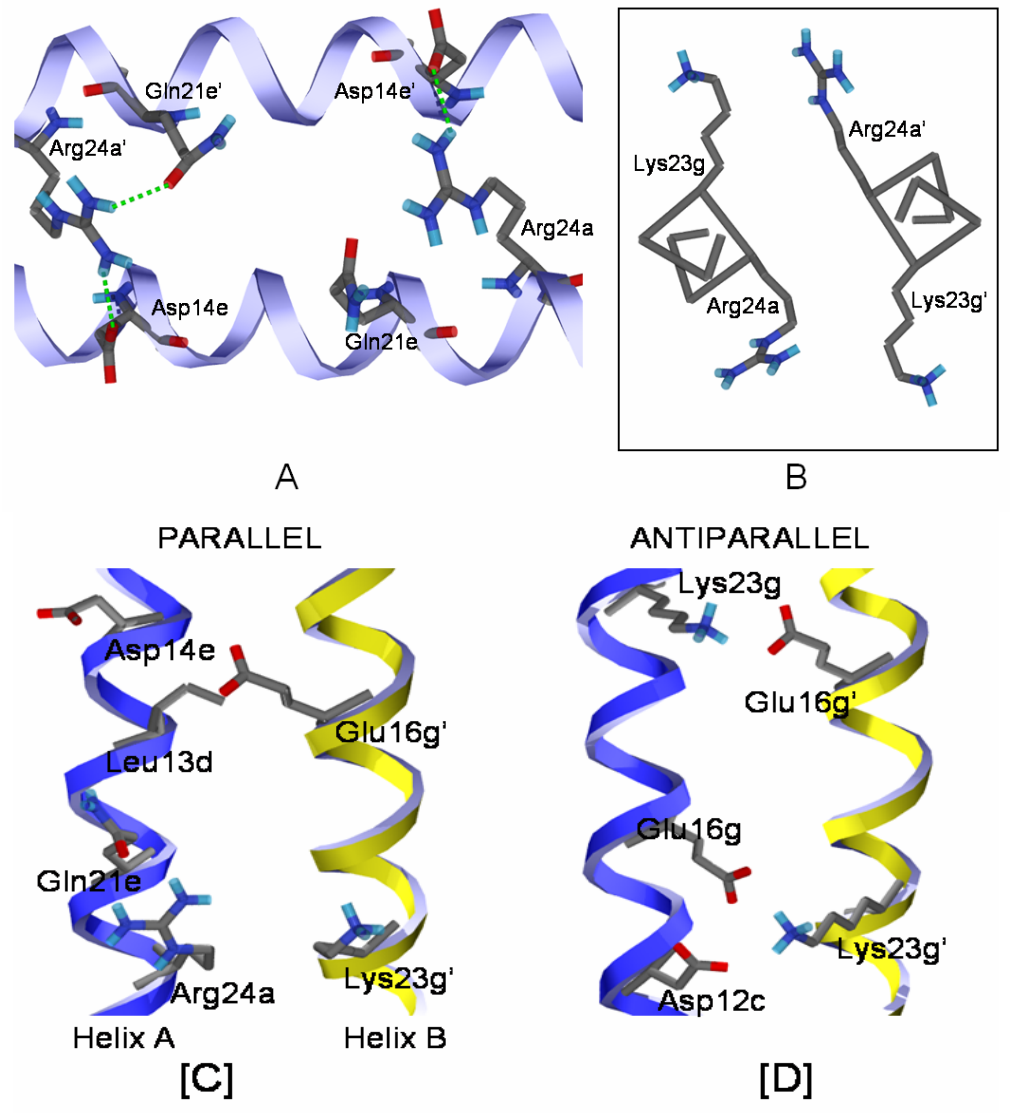


Figure 4.1. Comparison of electrostatic interactions of Arg24***a*** in parallel and antiparallel structures. **A**. Hydrogen bonds made between Arg24***a*** and the residues Asp14***e'*** and Gln21***e*** in the average structure. Hydrogen bonds (green dotted lines) drawn with Swiss Viewer with a 2.76 Å detection threshold. **B**. Top view of Arg24***a*** and Lys23***g'*** in the parallel 1R48 model. The two side chains are close enough to interact, which results in an unfavorable ELEC term. **C**. Parallel 1R48 model showing residues within interacting distance of Lys23***g*** and Glu16***g***. For Glu16***g***, the closest residue at position ***e*** is Asp14. Even if Glu16***g*** was matched with Lys or Arg at position 14***e***, the packing of Leu13***d*** would interfere with the alignment of the two side chains. For Lys23***g***, the closest residue at position ***e*** is Gln21. Even if Lys23***g*** was matched with a residue of opposite polarity at position 21***e***, Arg24***a*** would interfere with the alignment of the two side chains. For clarity, only the interactions on one side of the coiled coil are shown.
D. Antiparallel 1R48 model showing how Lys23***g*** and Glu16***g*** interact. Note that Asp12***c*** can also participate in electrostatic interactions with Lys23***g*** from the opposite helix.

coiled coil dimers to have beta-branched residues at position ***a***. Thus the observed behavior of 1R48 and the ARG24***a*** to ILE mutant is more a matter of VDW interactions. Interestingly, on an experimental study of the *Agrobacterium tumefaciens* osmosensor ProP counterpart, the investigators chose to mutate one LYS and one ARG at different ***a*** positions to ILE.[57] The results from this study indicate that LYS and ARG provide some orientation specificity but side chains of opposite polarity must be present in the neighboring helix within interacting distance. Up to date, there are no known results for the substitution of these two polar amino acids with nonpolar ones other than ILE.

Another contribution to antiparallel specificity is from HIS31***a*** with 1.35 kcal/mol from combined VDW, BOND, and CONF. HIS31***a*** favors the antiparallel configuration in part due to steric effects. Placing two HIS at the core of a coiled coil dimer is unfavorable while matching it with a smaller residue will result in a larger number of allowed rotamers (0.16 kcal/mol of CONF in antiparallel). The VDW stabilization comes mainly from contacts with ILE6***d*** from the opposite helix. ILE6***d*** also favors the antiparallel structure in terms of VDW. This follows from the empirical rule that ILE at ***d*** are poorly accommodated in parallel coiled coil dimers. We noted that when ILE at ***d*** is matched with HIS at ***a*** in the antiparallel structure, both side chains retain their initial rotamers. However, when ILE was matched with ILE and HIS with HIS (in the parallel structure) one residue from each pair flipped.

1X9V dimer: This is a fragment (residues 52-96) of the C-terminal domain of the HIV-1 regulatory protein Vpr, which dimerizes in an antiparallel orientation through hydrophobic interactions between LEU and ILE residues and a pair of TRP/HIS stacking interactions.[55] The average of the 10 NMR structures from PDB 1X9V indicate that only a short segment of the peptide is folded and the rest is in a random coil conformation. Our antiparallel coiled coil model contains 20 residues.

The antiparallel specificity is only -1.96 kcal/mol with side chain entropy slightly favoring the conformer (Table 4.1B). The random coil portion, not included in the model, might actually favor the antiparallel conformer in terms of side chain and backbone entropy because the conformational space of one monomer would overlap the other in the parallel conformer.

The free energy calculation results are summarized in **table 3**. Our antiparallel coiled coil model contains a TRP[***a***] near the N-Terminus and a HIS[***d***] near the C-terminus. This TRP/HIS pair has been described as a stacking interaction.[55] In our model, SCWRL3.0 built TRP3***a*** with a rotamer that minimizes steric clashes with HIS20***d***. The rotamer in question does not comply with the requirements of the TRP/HIS T-stacking interaction observed in the PDB structure but the changes that occurred during the MD simulation bring TRP3***a*** and HIS20***d*** very close to such arrangement. TRP3***a*** favors the antiparallel orientation by 2.38 kcal/mol mostly from BOND and ELEC. The

Table 4.3 Average Free Energy (W - $\mathbf{T}\Delta S^{conf}$) per residue for the 1X9V dimer sequence during two 0.8 ns (400 frames) Nosé-Hoover MD simulations

	II	anti_II	ΔΔG
Ace	-4.94	-5.37	-0.43
Thr 2g	-1.82	-1.71	0.12
Trp 3a	14.89	12.51	-2.38
Thr 4b	-2.93	-2.79	0.14
Gly 5c	-6.54	-6.42	0.12
Val 6d	-2.12	-5.07	-2.95
Glu 7e	-3.26	-3.13	0.13
Ala 8f	-6.39	-6.46	-0.07
Leu 9g	-4.89	-4.01	0.88
Ile 10a	-5.21	-4.13	1.08
Arg 11b	2.96	1.84	-1.12
Ile 12c	-3.08	-3.08	0.00
Leu 13d	-6.10	-4.22	1.88
Gln 14e	-8.63	-8.40	0.22
Gln 15f	-8.45	-8.85	-0.40
Leu 16g	-6.33	-5.96	0.37
Leu 17a	-6.30	-5.67	0.63
Phe 18b	-1.52	-1.33	0.18
Ile 19c	-1.10	-1.23	-0.13
His 20d	-2.03	-2.71	-0.68
Phe 21e	1.72	1.97	0.25
Cbx	-1.15	-0.96	0.18
$\sum$[W - $\mathbf{T\Delta S^{conf}}$]	-63.22	-65.18	-1.96

[a] All entries are in units of Kcal/mol•helix. The first and second column are the results for parallel and antiparallel topology, respectively third. The third column (ΔΔG) is the free energy that favors a particular helix orientation, negative values for antiparallel and positive values for parallel. 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.

two TRP3***a*** do not interact with each other in the parallel conformer. At least one of these TRP forms a hydrogen bond with the N-terminal ACE cap of the neighboring helix and disrupts the backbone. This indicates that placing two TRP at ***a*** side by side in a coiled coil dimer is energetically costly.

HIS20***d*** favors antiparallel by 0.68 kcal/mol, we noted that HIS at ***d*** is lower in energy than HIS at ***a*** (at least compared to HIS31***a*** in the 1R48 dimer). In our 1X9V parallel dimer model, the HIS20***d*** Nε binds to the backbone carbonyl of the HIS20***d*** of the neighboring helix. Our results show a large stabilization from VDW and ELEC from the interaction between the two HIS. In the 1X9V antiparallel dimer model, HIS20***d*** alternates between VDW contacts with THR2***g*** and ELEC contacts with TRP3***a*** of the neighboring helix (THR2***g*** has a favorable VDW term in the antiparallel structure). These results indicate that HIS at ***d*** would be useful to direct coiled coil dimer formation. The parallel orientation would be favored by default, and the antiparallel orientation might be directed by placing certain residues at the complementary ***a*** position.

VAL6***d*** prefers the antiparallel conformer by 2.95 kcal/mol. The orientation preference of VAL6***d*** is mostly influenced by the unfavorable packing of beta branched residues at ***d*** positions of parallel coiled coil dimers. VAL6***d*** in the antiparallel dimer is also stabilized because the side chain is flanked by LEU16***g'*** and LEU17***a'***. This positioning of LEU residues results if favorable VDW contacts for the two most populated rotamers of VAL6***d*** because there is a LEU side chain on either side. This might explain the VDW and CONF contributions. LEU17***a'*** does not favor antiparallel but the proximity of LEU16***g'*** (which favors parallel) in the parallel conformer has a significant effect in the VDW term. The available data does not yield a clear conclusion in the orientation preference of LEU at ***a*** but it is likely to be indifferent as far as dimeric coiled coils is concerned.

4.4 Antiparallel trimers (1RB4 and 1COS)

1RB4 trimer: This peptide is an ASN**16**ALA mutant of the GCN4 coiled coil parallel dimer which forms an antiparallel trimer.[**61**] The structure reveals that steric matching of large and small (LEU and ALA) side chains might play a role in directing the

antiparallel topology. With the addition of the ACE cap, our coiled coil model shifts the numbering scheme by +1 compared to the PDB version.

The antiparallel conformer is lower in energy than the parallel one by 1.50 kcal/mol (Table 4.1C). The side chain entropy contributes -2.76 kcal/mol to the antiparallel conformer stability. The asymmetry of the antiparallel trimers complicates the analysis because one of the three corresponding residues is in a different structural environment. We noted that not all three helices contribute equally to the reported average free energy, and the differences are sometimes very large. The summary of results as presented on table 4.4 is the average of three monomers, but the energy calculations for each helix was done separately. In our model, helices A and C are parallel to each other and helix B is antiparallel to A and C.

ARG2***g*** favors the antiparallel conformer by 1.24 kcal/mol. This is mostly due to mismatch in the oligomerization interface. A substantial contribution comes from the CONF term, which means that ARG2***g*** has less steric constraints in the antiparallel conformer. The VDW term also contributes due to the contacts ARG2***g*** of helix A makes with MET3***a*** of helix C. Three such contacts could form in the parallel configuration but the entropy loss would still be greater than the VDW gain.

MET3***a*** favors antiparallel by 1.12 kcal/mol mostly from VDW and BOND. The individual helix calculations showed no significant difference between the three MET3***a***, thus the contribution is not sequence dependent. MET3***a*** in the parallel conformer is located at the end of the coiled coil, thus lack VDW interactions that are only available in the antiparallel structure. Here MET3***a*** can interact with VAL24***a'***, LEU27***d'*** and VAL31***a'***, while in the parallel structure MET3***a*** can interact with only one residue at the ***a*** position of the neighboring helix.

GLU7***e*** favors antiparallel by 0.97 kcal/mol mostly from ELEC and CONF. The largest contribution to the effective energy is from helix C. Two of the GLU7***e*** make ELEC contacts with LYS28***e*** from neighboring helices while GLU7***e*** from helix C makes contacts with ARG2***g*** from helix A. LYS28***e*** also favors antiparallel by 0.11 kcal/mol. The ELEC term is large enough to outweigh the entropy loss. Thus the two LYS28***e***-GLU7***e'*** electrostatic interactions are a main determinant of the antiparallel coiled coil topology but the ARG2***g***-GLU7***e'*** electrostatic interaction between the parallel helices is

Table 4.4 Average Free Energy (W - $\mathbf{T}\Delta S^{conf}$) per residue for the 1RB4 trimer sequence during two 1.8 ns (900 frames) Nosé-Hoover MD simulations[a]

1RB4	II	anti_II	ΔΔG
Ace	-4.46	-4.74	-0.28
Arg 2g	6.40	5.15	-1.24
Met 3a	-1.22	-2.34	-1.12
Lys 4b	0.98	0.38	-0.61
Gln 5c	-8.66	-8.55	0.11
Leu 6d	-6.16	-6.04	0.12
Glu 7e	-3.14	-4.11	-0.97
Asp 8f	-10.22	-10.01	0.21
Lys 9g	-0.63	-0.56	0.06
Val 10a	-4.25	-4.18	0.08
Glu 11b	-3.54	-3.73	-0.19
Glu 12c	-4.21	-3.64	0.57
Leu 13d	-6.51	-5.94	0.57
Leu 14e	-5.48	-5.65	-0.17
Ser 15f	-7.34	-7.44	-0.10
Lys 16g	1.04	0.58	-0.46
Ala 17a	-8.08	-8.22	-0.15
Tyr 18b	-0.44	-0.66	-0.22
His 19c	-2.52	-2.29	0.22
Leu 20d	-6.51	-6.10	0.41
Glu 21e	-2.72	-3.13	-0.41
Asn 22f	-11.14	-11.15	-0.01
Glu 23g	-4.19	-4.38	-0.20
Val 24a	-4.20	-4.09	0.12
Ala 25b	-5.67	-5.74	-0.07
Arg 26c	0.64	0.75	0.11
Leu 27d	-7.22	-6.19	1.03
Lys 28e	0.79	0.68	-0.11
Lys 29f	0.05	0.49	0.44
Leu 30g	-3.99	-3.74	0.25
Val 31a	-2.54	-1.94	0.60
Gly 32b	-5.08	-5.18	-0.10
Glu 33c	1.26	1.29	0.03
Cbx	-1.06	-1.10	-0.05
$\sum$[W - $\mathbf{T\Delta S^{conf}}$]	-120.01	-121.51	-1.50

[a] All entries are in units of kcal/mol•helix. The first and second column are the results for parallel and antiparallel topology, respectively third. The third column (ΔΔG) is the free energy that favors a particular helix orientation, negative values for antiparallel and positive values for parallel. 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.

stronger. With two guanidinium groups, ARG2***g*** can assume different configurations and still interact with GLU7***e'***.

We noted that LYS28***e*** can make an electrostatic bridge with a GLU in both parallel and antiparallel conformers but the ELEC term makes all the difference and the antiparallel one wins even though its CONF term is not favorable. There is a small difference (about 1 Å) in the distance between the alpha carbons of LYS28***e*** and the interacting GLU (GLU7***e*** in antiparallel, GLU23***g*** in parallel), the distance is longer in the antiparallel. However, the orientation of the side chains seems to be the most important factor. The LYS28***e***-GLU7***e'*** pair in the antiparallel conformer point in opposite directions and towards each other in an extended conformation while a LYS28***e***-GLU23***g'*** pair in the parallel structure places both residues pointing in roughly the same direction. This has nothing to do with the residue type, by default the side chains tend to orient themselves towards the N-terminus of the helix.

LYS16***g*** behaves a lot like LYS9***g***, and favors antiparallel by 0.33 kcal/mol from CONF. Overall, for the ***g*** and ***e*** position residues, the net result of the parallel to antiparallel switch is that a larger number of interacting polar side chains are optimally aligned.

TYR18***b*** provides a small contribution to the antiparallel alignment (0.46 kcal/mol). Energy calculations on the individual helices indicate that TYR18***b*** from helices A and B contribute to the antiparallel preference while TYR18***b*** from helix C slightly opposes the conformation. The hydrogen bond between the TYR18***b*** of helix A and helix B is stable enough to be detected in the average structure (See Figure 4.2B) and the VDW interactions with LEU14***e*** from the neighboring helix also contribute. We noted that, LEU14***e*** from helices A and B slightly favor antiparallel in terms of VDW.

GLU21***e*** favors antiparallel by 0.41 kcal/mol mainly from CONF. This is likely because residues at ***e*** are directed more towards the core in parallel coils than in antiparallel ones but the exact source is unclear.

Overall, the ***a*** and ***d*** residues of the 1RB4 peptide do not favor the antiparallel trimer topology in terms of the reported average free energy but some individual side chains do contribute. For example, LEU6***d*** from helix B.

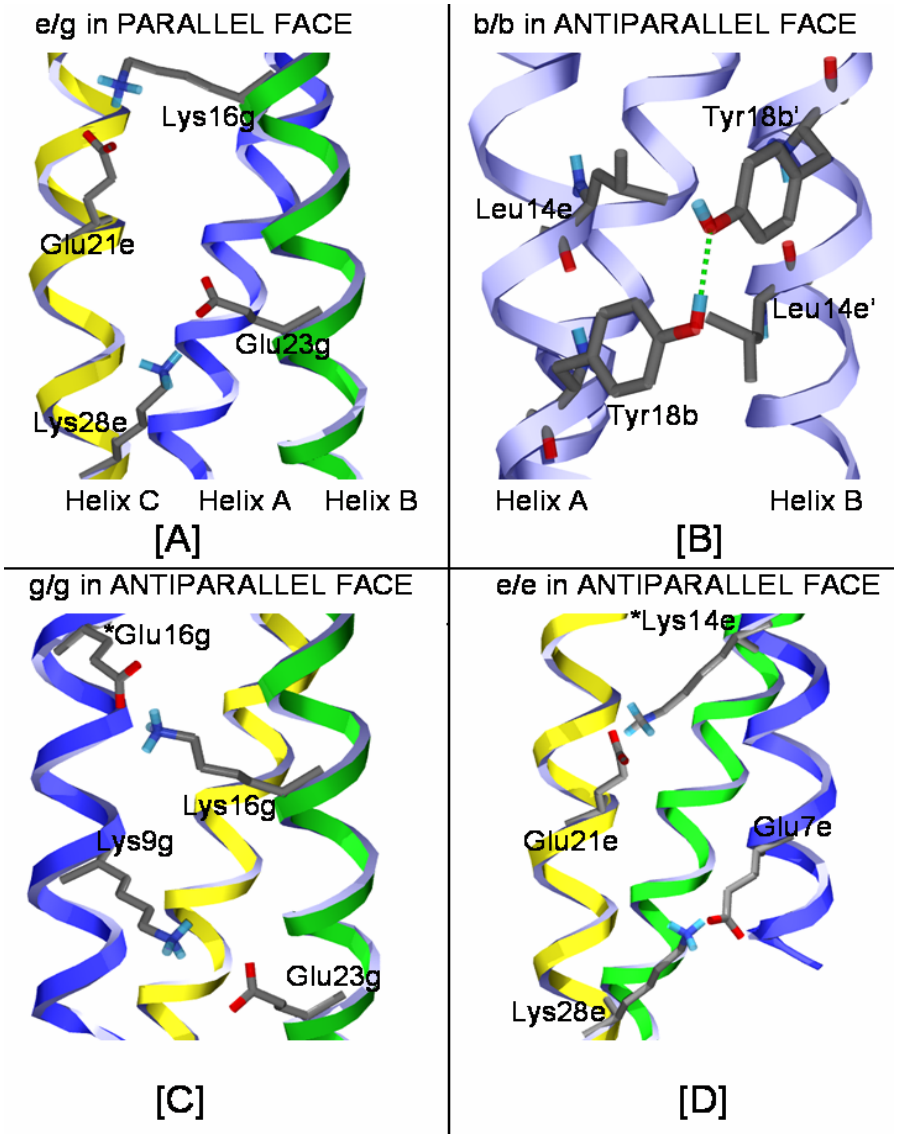


Figure 4.2. Directionality implications in the electrostatic interactions of the trimeric coiled coil 1RB4. Compared to its antiparallel counterpart, the parallel conformer has more constraints in the side chain type and even the order in which they are placed. **3A.** Glu23***g*** favors parallel Glu21***e*** does not. The Lys-Glu salt bridge has different stability depending on the directionality of the interacting side chains. **3B.** Hydrogen bond between Tyr17band Tyr17b'. The side chains interact favorably by assuming their default rotamers. **3C**. Salt bridge between Lys9***g*** and Glu23***g'*** in antiparallel 1RB4 compared with a similar salt bridge between Lys16***g*** and a theoretical Lys16***g'*** (this interaction would occur in a heterotypic coiled coil). **3D**. Salt bridge between Glu7***e*** and Lys28***e'*** in antiparallel 1RB4 compared with a similar salt bridge between Glu21***e*** and a theoretical Lys14***e'***. In average, the electrostatic interactions envolving ***e*** and ***g*** residues are stronger in the antiparallel trimers than in the parallel counterparts. The reason is that the side chain pairs are correctly aligned in the default orientation for the antiparallel topology regardless of order. Similar interactions in the parallel topology require larger deviations from the default rotamers as there are additional constraints.

We noted what might be a common trend in antiparallel coiled coil trimers. The LEU13***d*** and the ALA17***a*** form one ALA/ALA/LEU and one LEU/LEU/ALA layer. However, ALA17***a*** favors the antiparallel topology in terms of VDW while LEU13***d*** does not. Our calculations show that the VDW stabilization actually comes from LEU14***e***. Due to the asymmetry of the structure, LEU13***e*** and ALA16***a*** (from helix A and helix B) form two identical ALA/LEU layers very similar to what we would see in the core of an antiparallel dimer. The result from individual helix calculations showed that the ALA16***a*** from helix A and helix B favor the antiparallel topology while the ALA16***a*** from helix C does not, which accounts for the apparently negligible stabilization. Thus the design of an antiparallel coiled coil trimer would involve optimization of ***a/e'*** and ***d/g'*** VDW interactions rather than simple steric matching of ***a*** and ***d*** residues.

1COS trimer: The 1COS peptide contains four heptads similar to (L***a***E***b***A***c***L***d***E***e***G***f***K***g***)n, which was originally designed to mimic the structural features of the tropomyosin coiled coil dimer. A few modifications to the sequence resulted in an antiparallel coiled coil trimer.[64] Our model includes the full 1COS sequence. Crystallization took place at pH 5. The ***e/e*** face of the coiled coil has three pairs of GLU that come close enough to interact, so we modeled GLU as protonated. We also did calculations for this sequence with the GLU in the unprotonated form (results not shown) and obtained a smaller energy difference between parallel and antiparallel (2.95 kcal/mol) as compared to the protonated peptides (4.45 kcal/mol). The difference is mostly due to the unprotonated GLU side chains contributing less (in terms of ELEC) to the stability of the antiparallel topology. It seems possible that the topology preference of 1COS was influenced by the protonation state of its ionizable side chains. As such, the antiparallel topology will be favored at low pH while the parallel topology will be favored at high pH. The summary of results as presented on table 4.5 is the average of three monomers, but the energy calculations for each helix were done separately to isolate the effect of placing one of the three corresponding residues in a different structural environment. In our model, helices A and C are parallel to each other and helix B is antiparallel to A and C.

Our results (Table 4.1) show that the preference of 1COS for the antiparallel topology is entropy driven. LYS9***g***, LYS16***g*** and LYS23***g*** all favor the antiparallel

Table 4.5. Average Free Energy (W - $\mathbf{T}\Delta S^{conf}$) per residue for the 1COS trimer sequence during two 1.8 ns (900 frames) Nosé-Hoover MD simulations[a]

1COS	II	anti_II	ΔΔG
Ace	-4.35	-4.38	-0.03
Glu 2g	-3.44	-3.54	-0.09
Trp 3a	13.24	11.77	-1.48
Glu 4b	-6.83	-6.20	0.63
Ala 5c	-6.06	-6.00	0.06
Leu 6d	-5.75	-5.80	-0.05
Glu 7e	-6.79	-6.77	0.02
Lys 8f	0.11	-0.12	-0.23
Lys 9g	0.00	-0.57	-0.56
Leu 10a	-6.69	-6.50	0.19
Ala 11b	-7.43	-7.30	0.13
Ala 12c	-6.73	-6.46	0.27
Leu 13d	-6.22	-6.13	0.08
Glu 14e	-7.64	-8.31	-0.67
Ser 15f	-6.21	-6.05	0.16
Lys 16g	0.92	0.40	-0.52
Leu 17a	-6.57	-6.63	-0.06
Gln 18b	-7.90	-7.69	0.21
Ala 19c	-6.48	-6.28	0.20
Leu 20d	-6.22	-6.16	0.06
Glu 21e	-7.95	-8.02	-0.07
Lys 22f	-1.23	-1.02	0.21
Lys 23g	-0.24	-1.25	-1.01
Leu 24a	-6.47	-6.86	-0.39
Glu 25b	-8.04	-7.79	0.25
Ala 26c	-6.80	-6.92	-0.12
Leu 27d	-3.84	-4.02	-0.18
Glu 28e	-5.14	-6.02	-0.88
His 29f	-0.04	-0.49	-0.45
Gly 30g	-2.79	-2.86	-0.07
Cbx	-1.04	-1.09	-0.06
$\sum$[W - $\mathbf{T}\Delta\mathbf{S}^{conf}$]	-130.61	-135.05	-4.45

[a] All entries are in units of kcal/mol•helix. The first and second column are the results for parallel and antiparallel topology, respectively third. The third column (ΔΔG) is the free energy that favors a particular helix orientation, negative values for antiparallel and positive values for parallel. 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.

conformer by 0.56, 0.52, and 1.01 kcal/mol, respectively. The corresponding side chain entropy contributions are -1.68, -1.45 and -1.39 kcal/mol for LYS9***g***, LYS16***g*** and

LYS23***g***, respectively. In the parallel conformer, these three LYS side chains interact with three GLU at position ***e***, namely GLU14***e***, GLU21***e*** and GLU28***e*** in three faces of the coil. Whereas in the antiparallel conformer these interactions occur only on one out of three faces (the ***e***/***g*** face). Thus the antiparallel structure loses stabilization from ELEC in exchange for entropic free energy gain.

Two out of four GLU at position ***e*** favor antiparallel. GLU7***e*** and GLU21***e*** are nearly indifferent because their microenvironment does not change much going from parallel to antiparallel. GLU14***e*** favors antiparallel by -0.67 kcal/mol. Most of the stabilization is from the **e**/**e** antiparallel interface between helix A and B.

In summary, the GLU[***e***]-LYS[***g***] pairs are involved in electrostatic interactions in the parallel conformer but the stabilization gained from these interactions is not large enough to counter their CONF and SOLV cost. The antiparallel conformer has three possible GLU-LYS interactions in the ***e/g*** face, three GLU-GLU interactions in the ***e/e*** face, and three LYS-LYS interactions in the ***g/g*** face. On the other hand, all these nine interacting amino acid pairs are of ***e/g*** type in the parallel conformer. Thus the antiparallel conformer is favored because only one of its three faces has charges paired to form electrostatic bridges. Although salt bridges are normally stabilizing, the effect of the CONF and SOLV terms will sometimes outweigh the ELEC term.

The microenvironment of GLU28***e*** is significantly different from the other three GLU at ***e***. The CONF term of GLU28***e*** disfavors antiparallel by 0.59 kcal/mol but the overall free energy favors antiparallel by 0.88 kcal/mol, mainly due to VDW and ELEC. The favorable contributions to the free energy come from helices A and B (antiparallel interface) while helix C opposes the antiparallel topology. We noted that GLU28***e*** interacts with TRP3***a***, forming some kind of antiparallel dimer interface between helices A and B. GLU28***e*** also interacts with GLU7***e*** between helices A and B, while GLU28***e*** from helix C interacts with LYS23***g***. Since TRP3***a*** was added by the investigators to facilitate NMR and UV spectroscopic studies,[**64**] the side chain is not considered relevant in determining the topology of the 1COS peptide. However, GLU28***e*** from all three helices are involved in favorable electrostatic interactions, yet this residue has an unfavorable free energy in helix C. We concluded that the GLU23***e***-TRP3***a*** VDW term is what makes the difference in helices A and B. The free energy of LEU27***d*** also favors the

antiparallel topology (0.18 kcal/mol) and comes only from helix B. LEU27***d*** of helix B faces two TRP3***a*** (from helix A and C), which explains why the main contribution comes from VDW.

TRP3***a*** favors antiparallel by 1.48 kcal/mol. Besides the above mentioned interaction in the antiparallel (A-B) interface, TRP at ***a*** would be unfavorable for the parallel trimer topology due to steric reasons. The three TRP3***a*** in the parallel conformer are arranged in three-fold symmetry with the nitrogens in the pentameric ring interacting with each other, leaving the hexameric ring pointing towards the N-termini. Such arrangement would be even less favorable if the TRP side chains were not located at the end of the peptide.

LEU6***d*** favors antiparallel by 0.05 kcal/mol, mostly from VDW with a CONF contribution. This value is small on average but only the LEU6***d*** from helix B contributes while the LEU6***d*** of helices A and C slightly oppose the conformation. LEU6***d*** is paired in complementary layers with LEU24***a***, which favors antiparallel by 0.39 kcal/mol mostly from VDW and CONF. We examined the LEU24***a***, and found that helices B and C contribute to the antiparallel preference while helix B opposes it. Thus it is the layer that contains two LEU at ***a*** and one LEU at ***d*** that favor antiparallel. A simple explanation is that in the parallel coiled coil trimer, the packing of LEU at ***a*** is not as good as the packing of LEU at ***d***. Thus the layer that contributes more to the stabilization of the antiparallel structure is the one that removes the most LEU at ***a*** side chains from a badly packed environment. This explanation is consistent with the one given by Eisenberg *et al* .[**64**]

One of the most consistent patterns in the 1COS peptide is the antiparallel preference of LEU at ***a*** in terms of VDW. In our calculations, only LEU24***a*** stands out but all LEU at ***a*** have favorable VDW terms in the antiparallel structure. Thus, LEU at ***a*** could be useful for the design of trimeric antiparallel coiled coils. However, LEU at ***a*** favors antiparallel tetramers in the context of the 1W5K peptide. This means that the residues which destabilize the competing topology (for example, LEU at ***d*** in the antiparallel tetramer) are equally important. Overall, LEU at ***a*** gave results consistent with experimental data from the 1COI peptide, which contains VAL instead of LEU at ***a***. For this peptide, it was found that LEU at ***a*** disfavor parallel coiled coil trimers.[**98**]

4.5 Antiparallel tetramers (1W5K and 1TXP)

1W5K tetramer: The peptide corresponds to the GCN4-LI sequence (a parallel tetrameric coiled coil) with the substitution of a single solvent-exposed glutamic acid residue (GLU20***e*** to CYS). This is one of the few examples of coiled coil peptides with available structural data for both parallel and antiparallel topologies. Overall, the antiparallel conformer is 4.70 kcal/mol lower in energy than the parallel one (Table 4.1E). The summary of results for 1W5K is shown in Table 4.6.

As expected, CYS20***e*** (the substitution which triggers the antiparallel topology) favors the antiparallel topology by 0.38 kcal/mol. Two CYS20***e*** fit better than two GLU20***e*** within the available packing space. The TYR17***b*** residues pack against the surface of the coiled coil in 1W5K (Figure 4.1A), whereas if CYS20***e*** is replaced with GLU the TYR rings point towards the solvent Figure 4.1B. We computed the INTE contributions for residues TYR17***b*** and CYS20***e***. We found that the VDW term from CYS/TYR and TYR/TYR contacts are much more favorable in the antiparallel conformation. Thus we can conclude that the clustering of two TYR and two CYS in the antiparallel conformation provides significant stabilization from VDW. When the same calculation was done on the GCN4-LI sequence (replace CYS20***e*** with GLU) the VDW term for these residues is less favorable and GLU21***e*** incurs a ELEC and SOLV cost by switching to antiparallel topology. Since GLU is much bulkier than CYS, the two TYR17***b*** do not gain much VDW stabilization by switching to antiparallel topology in the GCN4-LI sequence. The ELEC term is also unfavorable for the antiparallel conformer because it places the two GLU21***e*** carbonyl groups close to each other. Even in protonated form, the two GLU21***e*** carbonyl groups would not interact favorably because the distance is not right. This is in contrast to the 1COS antiparallel trimer where the distance between two protonated GLU at ***e*** is optimal for electrostatic interactions.

In the antiparallel 1W5K tetramer, the organization of the core residues is not strictly 2***a***/2***d*** layers. Each side chain occupies the space between two core residues in the opposite helix. This arrangement affords more efficient packing than the parallel structure core. As a result, all LEU at ***a*** favor the antiparallel conformer as do some ILE at ***d***, but to a smaller extent.

Table 4.6. Average Free Energy (W - $\mathbf{T}\Delta S^{conf}$) per residue for the 1W5K tetramer sequence during two 0.8 ns (400 frames) Nosé-Hoover MD simulations[a]

1W5K	II	anti_II	ΔΔG
Ace	-4.66	-4.73	-0.06
Met 2a	0.37	-0.17	-0.54
Lys 3b	1.43	1.40	-0.03
Gln 4c	-7.57	-7.53	0.04
Ile 5d	-3.36	-3.62	-0.26
Glu 6e	-4.04	-3.50	0.54
Asp 7f	-10.49	-9.99	0.49
Lys 8g	-1.18	-0.82	0.36
Leu 9a	-5.47	-6.08	-0.61
Glu 10b	-3.14	-3.36	-0.22
Glu 11c	-4.10	-4.41	-0.31
Ile 12d	-3.88	-3.86	0.03
Leu 13e	-5.94	-5.52	0.41
Ser 14f	-6.94	-7.66	-0.72
Lys 15g	0.81	1.00	0.19
Leu 16a	-5.88	-6.07	-0.20
Tyr 17b	-0.44	0.05	0.49
His 18c	-2.80	-2.90	-0.10
Ile 19d	-4.46	-4.79	-0.33
Cys 20e	-4.78	-5.16	-0.38
Asn 21f	-10.46	-10.39	0.07
Glu 22g	-3.99	-4.40	-0.41
Leu 23a	-5.08	-5.91	-0.83
Ala 24b	-5.91	-6.10	-0.19
Arg 25c	1.37	1.06	-0.32
Ile 26d	-4.05	-3.97	0.08
Lys 27e	3.63	1.91	-1.72
Lys 28f	1.50	1.07	-0.43
Leu 29g	-3.32	-2.63	0.69
Leu 30a	-0.50	-0.65	-0.15
Cbx	-0.76	-1.06	-0.30
$\sum$[W - $\mathbf{T\Delta S^{conf}}$]	-104.07	-108.78	-4.70

[a] All entries are in units of kcal/mol•helix. The first and second column are the results for parallel and antiparallel topology, respectively third. The third column (ΔΔG) is the free energy that favors a particular helix orientation, negative values for antiparallel and positive values for parallel. 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.

ILE5***d*** favors antiparallel by 0.26 kcal/mol with the SOLV and CONF terms being the most significant. These energy contributions come mainly from interactions with

LEU30***a*** in the opposite helix but the interaction range extends to the next nearest core residue (in this case ILE26***d'***) and the flanking ***g'*** position residue (in this case LEU29***g***). LEU30***a*** also favors antiparallel by 0.15 kcal/mol mostly from VDW and BOND. Interestingly, the VDW term has nearly the same value for the LEU30***a***-GLU6***e*** and the LEU30***a***-ILE5***d*** interacting pairs. This means that a combination of both vertical (nearest ***a'*** residues) and lateral (flanking ***e'*** residues) interactions help encode the antiparallel tetrameric coiled coil topology.

ILE19***d*** favors antiparallel by 0.33 kcal/mol from VDW and CONF. This is mostly due to ILE19***d***- LEU16***a*** VDW interactions but ILE19***d***-LYS15***g*** VDW interactions also contribute.

LEU9***a*** favors antiparallel by 0.61 kcal/mol mostly from BOND. The other residue in the layer is ILE26***d***, which does not favor antiparallel. The LEU9***a***- ILE26***d*** SOLV term still exhibits the same trend as the ILE5***d***-LEU30***a*** pair but its significance vanishes when the effect from all other residues is taken into account. The VDW term has contributions from ILE26***d***, LEU23***a*** and LYS27***e***. From this, we noted that the VDW term from the nearest core residue (ILE26***d***) is comparable in magnitude to the VDW term from LYS27***e***, this is another example of lateral interactions playing an important role.

LEU16***a*** favors the antiparallel tetramer but the contribution is not very large (0.20 kcal/mol). The energy term that favors the antiparallel tetramer the most is ELEC. This could only come from the backbone. The Sγ of CYS has tendency to form a hydrogen bond with the i-4 (LEU16***a*** in this case) peptide carbonyl oxygen. We noted that the distance from the main chain amine of LEU16***a*** to the main chain carbonyl of CYS20***e*** is slightly longer in the parallel structure. The most likely reason is that CYS20***e*** incurs a higher SOLV cost in the parallel structure and counters this by satisfying its hydrogen bonding potential. This has the effect of weakening the main chain hydrogen bond between CYS20***e*** and LEU16***a***.

LEU23***a*** favors antiparallel by 0.83 kcal/mol from SOLV and BOND. The energy terms come from interactions with ILE12***d*** and LEU13***e***. The fact that ILE12***d*** and LEU13***e*** do not favor antiparallel is consistent with previous results which showed that

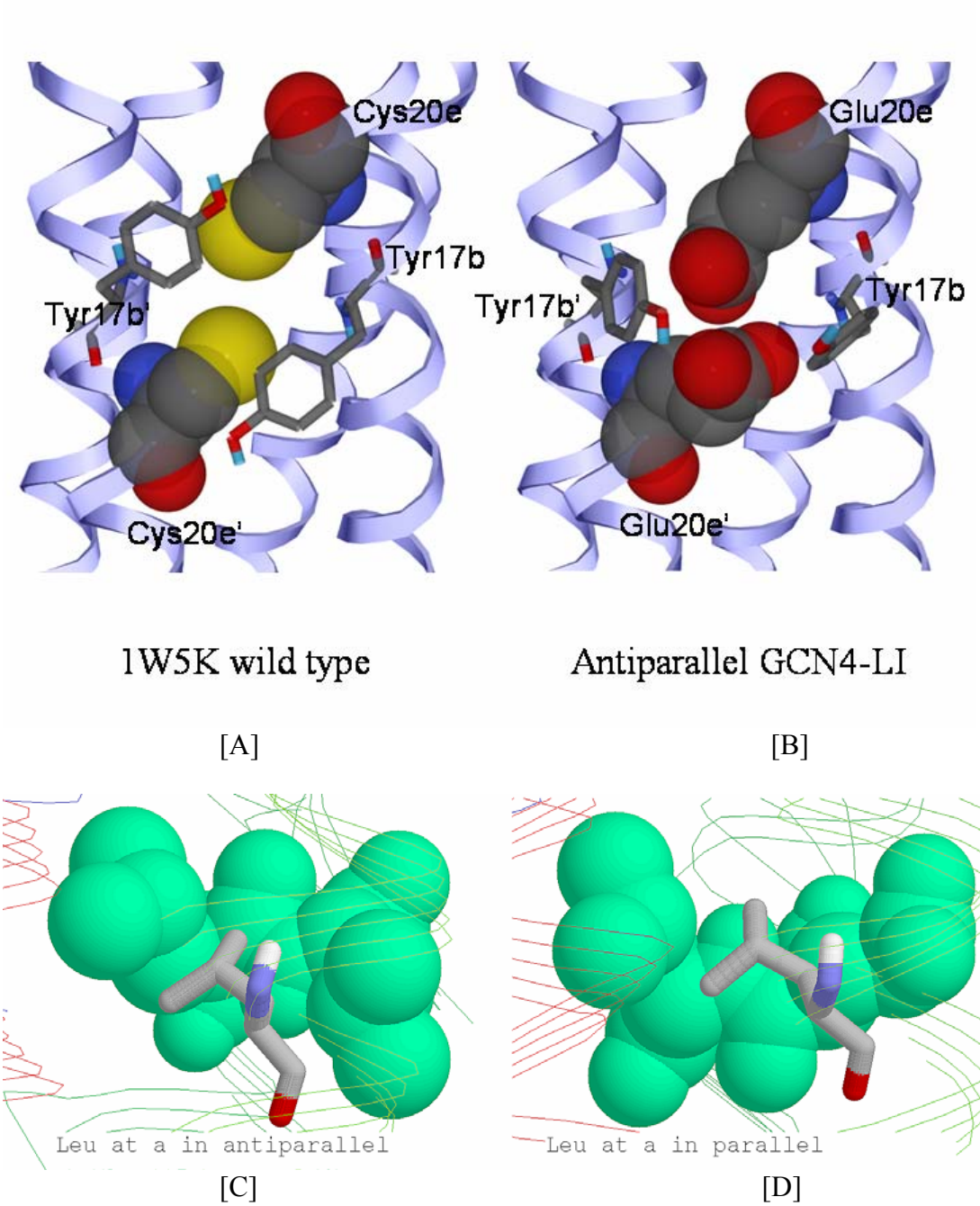


Figure 4.3 Structural determinants of the 1W5K antiparallel tetramer**. A.** Side view of the packing of Tyr17b and Cys20e in the 1W5K tetramer. There are two such clusters in an antiparallel coiled coil tetramer. For clarity, only the front cluster is shown. **B.** For comparison: packing of Tyr17**b** and Glu20**e** in an antiparallel GCN4-LI model. Notice that Glu20**e** is too bulky for its available packing space. **C**. Leu at ***a*** packing in antiparallel and [**D**] parallel tetramer. The flanking residue at position (**e** in antiparallel, **g** in parallel) might be more significant in the antiparallel conformer in terms of VDW.

ILE12***d*** interacts with Leu13***e*** from the neighboring helix and this combination confers specificity for the parallel coiled coil tetramer.

MET2***a*** favors the antiparallel conformer by 0.54 kcal/mo, mostly from CONF. The structure places MET2***a*** in non interacting pairs at either end of the coiled coil. We would expect quite different results if this residue was not at the end of the helices.

In summary, LEU at ***a*** seems to have preference for the four stranded state while the preference of ILE at ***d*** is not well defined. Then, it is possible that the LEU at ***a*** make 1W5K tetrameric and the ILE at ***d*** adjusts to the topology. We propose that one main determinant for the antiparallel orientation of 1W5K is the extent of VDW contacts that the flanking ***e*** residue makes with LEU at ***a***. To illustrate this, we compare LEU9***a*** with LEU16***a***, which also favors antiparallel. We selected these two residues because they both have flanking LYS residues. In the parallel structure, LEU16***a*** makes VDW contacts with LYS15***g***. In the antiparallel structure, LEU9***a*** makes VDW contacts with LYS27***e***. However, the contacts are better in the antiparallel structure due to the default directionality of the side chains (Figure.4.1C and D).

1TXP tetramer: Corresponds to the oligomerization domain of the hnRNP tetramer. Our model coiled coil contains the full length of the 1TXP peptide plus ACE and CBX caps, which shifts the numbering scheme by one as compared to the PDB version. An average of the 20 NMR structures retrieved from the PDB indicates that all but the last four residues in the sequence are helical.[7] These last residues are beyond the oligomerization interface. In the absence of a random coil building algorithm, we built them in a helical conformation.

The antiparallel conformer has a ΔΔG of -3.60 kcal/mol, with -3.17 kcal/mol from side chain entropy. Thus the formation of the antiparallel coiled coil tetramer is mostly entropy driven (Table 4.1F).

Table 4.7 shows the summary of results for the 1TXP sequence. Apparently, there is no clear cut pattern as to which residue combination is best as far as the ***a*** and ***d*** residues are concerned. However, we noted that certain ***a/e'*** and ***d/g'*** amino acid pairs stabilize the antiparallel coiled coil tetramer.

Table 4.7. Average Free Energy (W - $\mathbf{T}\Delta S^{conf}$) per residue for the 1TXP tetramer sequence during two 0.8 ns (400 frames) Nosé-Hoover MD simulations[a]

	II	anti_II	ΔΔG
Ace	-4.73	-4.72	0.02
Ile 2a	0.72	-0.10	-0.82
Gln 3b	-6.39	-6.78	-0.39
Ala 4c	-6.23	-6.17	0.06
Ile 5d	-3.34	-4.46	-1.12
Lys 6e	-0.41	-1.00	-0.58
Lys 7f	-0.40	-0.62	-0.21
Glu 8g	-3.38	-4.15	-0.77
Leu 9a	-6.92	-7.02	-0.10
Thr 10b	-3.55	-3.41	0.14
Gln 11c	-7.74	-7.27	0.48
Ile 12d	-4.57	-4.42	0.15
Lys 13e	1.94	0.47	-1.47
Gln 14f	-8.23	-7.74	0.49
Lys 15g	-0.09	0.19	0.28
Val 16a	-4.00	-4.43	-0.42
Asp 17b	-9.73	-10.38	-0.66
Ser 18c	-6.38	-7.51	-1.14
Leu 19d	-5.55	-5.20	0.35
Leu 20e	-4.88	-4.38	0.50
Glu 21f	-3.23	-2.77	0.46
Asn 22g	-9.78	-9.21	0.57
Leu 23a	-5.89	-6.29	-0.41
Glu 24b	-3.91	-3.60	0.31
Lys 25c	-0.42	-0.25	0.16
Ile 26d	-2.21	-1.75	0.46
Glu 27e	-2.48	-2.67	-0.19
Lys 28f	2.57	2.81	0.24
Glu 29g	1.49	1.37	-0.13
Cbx	-1.19	-1.06	0.13
$\sum$[W - $\mathbf{T\Delta S^{conf}}$]	-108.90	-112.50	-3.60

[a] All entries are in units of kcal/mol•helix. The first and second column are the results for parallel and antiparallel topology, respectively third. The third column (ΔΔG) is the free energy that favors a particular helix orientation, negative values for antiparallel and positive values for parallel. 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.

ILE2***a*** favors the antiparallel conformer by 0.82 kcal/mol mostly from VDW. The closest core residues are LEU23***a*** and ILE26***d*** but the VDW term with GLU27***e*** is the most stabilizing. It might not be appropriate to conclude that such a residue combination

confers antiparallel specificity because ILE2***a*** does not have any flanking residues in the parallel structure (because it is too far at the end of the coil). GLU27***e*** also favors antiparallel. This residue can make a hydrogen bond in either orientation; with LYS6***e'*** in parallel and ASN22**g**' in antiparallel. However, the CONF term that makes GLU27***e*** favor the antiparallel structure. It is possible that GLU27***e*** has favorable VDW contacts over a larger range of rotamers in the antiparallel structure.

ILE5***d*** favors the antiparallel conformer by 1.12 kcal/mol mostly from BOND. ILE5***d*** in 1TXP has a flanking LYS at ***e*** in parallel and an ASN at ***g*** in antiparallel. We compared the free energy of these two interactions because we have observed instances where a flanking LYS stabilizes the antiparallel structure. Our results indicate that ASN at ***g*** in the antiparallel structure has a more favorable VDW term than LYS at ***e***. Thus it seems like the directionality of the interacting side chains is more relevant than its identity.

As noted for other LEU at ***a***, there are residues that can complement the interaction pattern and also contribute to specificity for the antiparallel tetramer coiled. LEU23***a*** favors antiparallel by 0.41 kcal/mol mainly from VDW. LEU23***a*** is flanked by LYS6***e*** from a neighboring helix. LYS6***e*** favors antiparallel by 0.58 kcal/mol mainly due to CONF. This is very similar to the way LEU9***a*** interacts with LYS27***e*** in the antiparallel 1W5K tetramer.

LYS13***e*** favors the antiparallel topology by 1.47 kcal/mol mainly due to CONF. The ELEC component favors the parallel topology because LYS13***e*** makes an electrostatic bridge with GLN11***c'***. This interaction seems highly stable as all four LYS13***e***-GLN11***c'*** pairs are observed in the average structure. There are no counterparts to these electrostatic bridges in the antiparallel conformer. This is similar to other cases where LYS side chains gain entropic free energy by avoiding electrostatic interactions.

VAL16***a*** favors antiparallel by 0.42 kcal/mol mostly from BOND and ELEC. The main contacts are made with ILE12***d'*** and LYS13***e'*** with the VDW term from LYS13***e'*** being slightly more stabilizing. Overall, it seems like VAL at ***a*** would help specify an antiparallel orientation in a coiled coil tetramer when combined with a flanking LYS at ***e***. This is the same behavior we observed for LEU at ***a***. Probably it is the -C-$(CH_2)_2$ part of

LEU and VAL at ***a*** that provides the best VDW contacts for LYS at ***e*** in the opposite helix.

ASP17***b*** favors antiparallel by 0.66 kcal/mol mostly from ELEC. The most significant component is from interactions with LYS13***e*** from the neighboring helix. Electrostatic interactions of the ***b-e'*** type are also observed in PDB 3E7K; the structure of an antiparallel tetrameric coiled coil reported by Fujiwara and Minor Jr.[**99**] In 3E7K, ASN1251***b*** interacts ASP1261***e*** (numbering scheme in citation) and the interaction was reproduced in the minimized average of a model built using our method of coiled coil building. We found the free energy for this interaction to be favorable for the antiparallel structure as compared to the parallel one (results not shown).

SER18***c*** favors antiparallel by 1.14 kcal/mol. The largest contribution is from ELEC and the second largest is from VDW. The ELEC term is mostly from the backbone hydrogen bonds (GLN14***f*** and ASN22***g***). The Oγ of SER has tendency to form a hydrogen bond with the i-4 (GLN14***f***) peptide carbonyl oxygen. This interaction introduces a distortion of the backbone hydrogen bonding pattern. For SER18***c***, the nearest neighboring helix residue in the parallel orientation is LEU20***e*** while in the parallel orientation the nearest neighboring helix residue is GLU8***g***. It is the interaction with GLU8***g*** that is countering the distortion of the backbone hydrogen bonding pattern because SER18***c*** can make a hydrogen bond with GLU8***g*** besides the main chain carbonyl of GLN14***f***.

4.6 Discussion

In this work, we have built models of six coiled coil sequences in both parallel and antiparallel configurations using Crick's mathematical model and computed their relative free energy using CHARMM and EEF1.1. The preferential stability of the antiparallel over the parallel conformers seems to hold for the sequences considered here. However, we only consider parallel and antiparallel topologies of the experimentally observed oligomeric state.

The structural and energetic features of coiled coil topology described in this work provide new insights as to why certain amino acid types and positions can alter the

topology of the sequences studied. Our investigation on the determinants of helix orientation yields conclusions that are in agreement with previously reported observations. Because the antiparallel coiled coils incorporate mixed layers, this adds more complexity to their design as the same residue will behave differently depending on the nature of the interacting partner. Thus we have described orientation preference in terms of interacting amino acid pairs rather than single residues.

Dimers. For the two sequences studied here, the helix orientation in coiled coil dimers is encoded with the inclusion of certain residues besides the ones required to specify the oligomeric state.

[1] ARG and LYS at ***a*** have antiparallel dimer specificity. Placement of these residues at position ***d*** has nearly the same effect. The oligomeric state preference is the result of the high energetic cost of placing an ionizable side chain at the core of the helix bundle. Calculations done for other topologies indicate that, the absence of an opposite charge partner, ARG and LYS have the lowest energy when the ionizable groups are solvated.

The destabilization of competing topologies is an important element of coiled coil design. Our calculations show that ARG and LYS at ***a*** or ***d*** are on their own insufficient to destabilize the parallel dimer structure. Specificity for a single orientation might be attained by placing residues at **e'** or **g'** that form unfavorable interactions with the ***a*** or ***d*** position residues in the unwanted orientation. For example, ARG and LYS at ***a*** would require ARG or LYS at the ***e'*** position, while ARG and LYS at ***d*** would require ARG or LYS at the ***g'*** position to destabilize the competing parallel structure. In the parallel 1R48 dimer, the guanidinium group of ARG24***a*** comes within interacting distance of the ε-amino group of LYS23***g'***. Such an interaction yields an unfavorable ELEC term for the parallel model, effectively shifting the equilibrium towards the antiparallel orientation.

[2] ILE at ***d*** is better tolerated in the antiparallel orientation but the preference can be reversed if paired with a beta branched side chain in the complementary ***a*** position. The rationale for this observation is that ILE at ***d*** disfavors parallel dimers because two beta branched side chains do not fit properly in the core. If an ILE at ***a*** is matched with ILE at ***d***' in an antiparallel coiled coil dimer then the packing energy advantage of assuming the antiparallel orientation is effectively removed. We think that VAL at ***d*** behaves the same

way as ILE at ***d***. There is one instance (VAL6***d*** in 1X9V) where VAL at ***d*** prefers the antiparallel orientation.

[3] LEU at ***d*** and ILE at ***a*** strongly confer dimer specificity but the parallel orientation is lower in energy. Thus, a sequence that places ILE at ***a*** and LEU at ***d*** in contact with each other should be avoided when attempting to design an antiparallel homodimer.

[4] HIS at ***d*** favors parallel dimers. HIS at ***a*** will favor the antiparallel orientation as long as it is paired with a smaller side chain.

Trimers. There is not much information on antiparallel coiled coil trimers because very few structures have been studied. For the 1RB4 and 1COS peptides we noted the following trends:

[1] LEU at ***a*** is normally disfavored in parallel coiled coil trimers. Past studies have shown that layers of LEU at ***a*** rarely retain their 3-fold symmetry after dynamics. However, exceptions to the rule will be observed as a result of the sequence context. Our free energy calculations gave results consistent with experimental data from the 1COI peptide, that is, LEU at ***a*** disfavor parallel coiled coil trimers.[**98**] The main difference between 1COS and the 1COI is that the latter contains VAL instead of LEU at ***a***.

[2] With the exception of TRP3***a*** in 1COS, other core residues from the two sequences studied showed little preference for a particular topology. Steric matching of bulky residues with small ones has been used for *de novo* design of antiparallel coiled coil trimers.[**63**] Features such as this can be found in naturally occurring antiparallel coiled coil trimers such as the repetitive segments of Spectrin [PDB 1AJ3].[**100**] This structure contains HIS at ***a***, PHE at ***d***, and TRP at ***d***. Thus the strategic placement of such residues at core positions are is one of the most straightforward methods for the design of antiparallel coiled coil trimers.

[3] Electrostatic interactions envolving ***e*** and ***g*** residues are stronger in the antiparallel trimers than in the parallel counterparts. The reason is that the side chain pairs are correctly aligned in the default orientation for the antiparallel topology. Similar interactions in the parallel topology require larger deviations from the default rotamers. This trend is shown in Fig. 4.3. In summary, polar side chains at positions ***e*** and ***g*** of the antiparallel faces of the coiled coil will be correctly configured for optimal electrostatic interactions most of the time. Such interactions in the antiparallel faces require only that

the interacting amino acid pair have opposite polarities or hydrogen bonding potential. On the other hand, similar interactions in the parallel faces have an extra constraint. As seen in 1RB4, the interaction LYS28***g***-GLU23***e'*** is less favorable than LYS21***e***-GLU16***g'*** because the former requires LYS to assume a non standard rotamer. We noted a similar trend with other side chains. In the 1COS sequence, the protonated GLU at ***e*** are a main determinant of antiparallel topology because they can interact favorably while retaining low energy rotamers.

Tetramers. The coiled coil tetramers incorporate a larger set of interacting side chains. This means that a different interaction pattern should be considered. First, we tried layers and found no clear cut pattern to what residue combination is best if we assume that the topology of the antiparallel coiled coil tetramers is encoded in particular patterns of layers. This approach yielded very little information that was not already known. For example, there is clearly a tendency to exclude LEU at ***d***. It became clear that the approach was not going to work when we noted that ILE at ***d*** did not always favor the same structure even if similar layers were compared. We decided to treat the tetramers as a pair of antiparallel dimers connected at different points in the structure. We found that:

[1] LEU at ***a*** favors antiparallel tetramers but the extent of this specificity is heavily influenced by the flanking ***e*** residue in the neighboring helix. We found LEU at ***a*** and LYS at ***e'*** to be particularly stabilizing. We think ARG at an ***e'*** position will have the same effect.

[2] VAL at ***a*** behaves a lot like LEU at ***a*** due to similarities in their -C-$(CH_2)_2$ unit. We noted one instance where VAL16***a*** in the 1TXP tetramer provides favorable VDW contacts for LYS13***e*** in the opposite helix.

[3] ILE at ***d*** favors antiparallel tetramers when combined with certain flanking ***g*** residues in the neighboring helix. We found ILE at ***d*** combined with ASN at ***g'*** to fit this requirement.

[4] Electrostatic interactions envolving ***b*** and ***c*** residues are stronger in the antiparallel tetramers than in any other topology. The inclusion of polar or ionizable side chains at those positions should have the greatest impact in antiparallel coiled coil tetramers.

Our investigation of the individual residue contributions to the free energy provide new insights that might contribute to the design of coiled coils with a well

defined and predictable three-dimensional structure. Some trends are consistent with previous results. For example, core residues known to stabilize the oligomeric state in question (for parallel coiled coils) will disfavor the antiparallel conformer. We also noted that polar and charged side chains have a greater impact in controlling the topology of antiparallel coiled coils as compared to their parallel counterparts.

Unfortunately, the free energies reported in this thesis cannot be compared with experimental estimates. Thus we cannot predict absolute stability or dissociation constants with the methodology described here. This is probably due to the use of a folded helix, rather than a random coil, as a reference state for the monomer.

The ability to predict the association state and topology of a given coiled-coil from sequence alone remains elusive. This work is a significant step towards that objective. Our coiled coil building algorithm using Crick's mathematical model and Molecular Dynamics is able to reproduce important side chain interactions to atomic detail. For example, a model of the 3E7K antiparallel tetramer exhibited an electrostatic interaction between ARG1235***g*** and THR1277***g*** from the neighboring helix, just like the one described by Fujiwara and Minor Jr.[**99**] Also, our 1R48 antiparallel dimer model underwent some deformation during Molecular Dynamics. This deformation closely mimics the nonstandard coiled coil parameters of the average NMR structure and allowed certain interactions to be reproduced. In principle, our method requires only the sequence of a coiled coil domain and heptad assignments (optional). The association state and topology can then be inferred by comparing the free energy of the structural variants.

BIBLIOGRAPHY

[**1**] Lupas, Andrei "Coiled coils: new structures and new functions" *Trends in Biochemical Sciences* **1996**, 21(10), 375-382

[**2**] Mason, Jody M.; Arndt, Katja M. "Coiled coil domains: stability, specificity, and biological implications" *ChemBioChem* **2004**, 5(2), 170-176

[**3**] Wilson, I. A.; Skehel, J. J.; Wiley, D. C. " Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution" *Nature* 1981, 289, 366−373

[**4**] Landschulz, W. H.; Johnson, P. F.; McKnight, S. L. "The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins" *Science* **1988**, 240, 1759−1764

[**5**] O'Shea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T. "X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil" *Science* **1991**, 254, 539−544

[**6**] Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T. "A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants" *Science* **1993**, 262, 1401−1406

[**7**] Whitson, S. R.; LeStourgeon W. M.; Krezel, A. M. "Solution Structure of the Symmetric Coiled Coil Tetramer Formed by the Oligomerization Domain of hnRNP C: Implications for Biological Function" *J. Mol. Biol.* **2005***, 350*, 319−337

[**8**] Qabar, A.; Derick, L.; Lawler, J.; Dixit, V. "Thrombospondin 3 Is a Pentameric Molecule Held Together by Interchain Disulfide Linkage Involving Two Cysteine Residues" *J Biol. Chem*. **1995**, 270, 12725−12729

[**9**] Malashkevic, V. N.; Kammerer, R. A.; Efimov, V. P. "The Crystal Structure of a Five-Stranded Coiled Coil in COMP: A Prototype Ion Channel?" *Science* **1996**, 274, 761−765

[**10**] Oxenoid K.; Chou, J. J. "The structure of phospholamban pentamer reveals a channel-like architecture in membranes" *PNAS U.S.A.* **2005**, *102*, 10870−10875

[**11**] Liu, J.; Yong, W.; Deng, Y.; Kallenbach, N. R.; Lu, M. "Atomic structure of a tryptophan-zipper pentamer" *PNAS U.S.A.* **2004**, *101*, 16156−16161

[**12**] Ramos, J.; Lazaridis, T. "Energetic Determinants of Oligomeric State Specificity in Coiled Coils" *J. Am. Chem. Soc*., **2006**, 128 (48), pp 15499–15510

[**13**] A Lupas, M Van Dyke, and J Stock "Predicting coiled coils from protein sequences" *Science* **1991**, 252(5009), 1162 – 1164

[**14**] Wolf, Ethan; Kim, Peter S.; Berger, Bonnie "MultiCoil: a program for predicting two- and three-stranded coiled coils" *Protein Science* **1997**, 6(6), 1179-1189

[**15**] Wolfson, D. N.; Alber, T. "Predicting oligomerization states of coiled coils" *Protein Sci.* **1995**, 4, 1596−1607

[**16**] Fischetti, V. A.; Landau, G. M.; Schmidt, J. P.; Sellers, J. " Identifying periodic occurences of a template with applications to protein structure" *Inf. Proc. Lett*. **1993**, 45, 11−18

[**17**] Delorenzi M.; Speed T. "An HMM model for coiled-coil domains and a comparison with PSSM-based predictions" *Bioinformatics* **2002**, 18, 617−625

[**18**] Walshaw, J.; Woolfson, D. N. "SOCKET: A program for identifying and analyzing coiled-coil motifs within protein structures" *J. Mol. Biol.* **2001**, 307, 1427−1450

[**19**] Fong, J. H.; Keating, A. E.; Singh, M. "Predicting specificity in bZIP coiled-coil protein interactions" *Gen. Biol.* **2004**, 5, R11

[**20**] Deppmann, C. D.; Acharya, A.; Rishi,V.; Wobbes, B.; Smeekens, S.; Taparowsky, E. J.; Vinson, V. "Dimerization specificity of all 67 B-ZIP motifs in Arabidopsis thaliana: a comparison to Homo sapiens B-ZIP motifs " *Nucleic Acids Res*. **2004**, 32, 3435−3445

[**21**] Kammerer, R. A.; Schulthess, T.; Landwehr, R.; Lustig, A.; Engel, J.; Aebi, U.; Steinmetz, M. O. "An autonomous folding unit mediates the assembly of two-stranded coiled coils" *PNAS U.S.A.* **2005**, *95*, 13419−13424

[**22**] Van Heeckeren, W. J.; Sellers, J. W.; Struhl, K. "Role of the conserved leucines in the leucine zipper dimerization motif of yeast GCN4" *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. "Coiled coils at the edge of configurational heterogeneity. Structural analyses of parallel and antiparallel homotetrameric coiled coils reveal configurational sensitivity to a single solvent-exposed amino acid substitution" *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. "The retro-GCN4 leucine zipper sequence forms a stable three-dimensional structure" *PNAS U.S.A.* **2000**, *97*, 2562−2566.

[**25**] Moitra, J.; Szilak, L.; Krylov, D.; Vinson, C. "Leucine Is the Most Stabilizing Aliphatic Amino Acid in the d Position of a Dimeric Leucine Zipper Coiled Coil" *Biochemistry* **1997**, *36*, 12567−12573

[**26**] Lumb, K. J.; Kim, P. S. “Leucine Is the Most Stabilizing Aliphatic Amino Acid in the d Position of a Dimeric Leucine Zipper Coiled Coil” *Biochemistry* **1995**, 34, 8642−8648

[**27**] Oakley, M. G.; Kim, P. S. “A Buried Polar Interaction Can Direct the Relative Orientation of Helices in a Coiled Coil” *Biochemistry* **1998**, 37, 12603−12610

[**28**] Cristian, L.; Nanda, V.; Lear, J. D.; DeGrado, W. F. “A Buried Polar Interaction Can Direct the Relative Orientation of Helices in a Coiled Coil” *Biochemistry* **1998**, 37, 12603−12610

[**29**] Gonzalez, L., Jr.; Woolfson, D. N.; Alber, T. “Buried polar residues and structural specificity in the GCN4 leucine zipper” *Nat. Struct. Biol*. **1996**, 3, 1011−1018

[**30**] Shu, W.; Ji, H.; Lu, M. “Trimerization Specificity in HIV-1 gp41: Analysis with a GCN4 Leucine Zipper Model” *Biochemistry* **1999**, *38*, 5378−5385.

[**31**] Tripet, B.; Wagschal, K.; Lavigne, P.; Colin, C. T.; Hodges, R. S “Effects of side-chain characteristics on stability and oligomerization state of a *de Novo*-designed model coiled-coil: 20 amino acid substitutions in position d”. *J. Mol. Biol*. **2000**, *300*, 377−402.

[**32**] Wagschal, K.; Tripet, B.; Hodges, R. S. “*De novo* design of a model peptide sequence to examine the effects of single amino acid substitutions in the hydrophobic core on both stability and oligomerization state of coiled-coils” *J. Mol. Biol.* **1999**, *285*, 785−803

[**33**] Nautiyal, S.; Woolfson, D. N.; King, D. S.; Alber, T. “A Designed Heterotrimeric Coiled Coil” *Biochemistry* **1995**, 34 , 11645−11651.

[**34**] Akey, D. L.; Malashkevich, V. N.; Kim, P. S. “Buried Polar Residues in Coiled-Coil Interfaces” *Biochemistry* **2001**, 40, 6352−6360.

[**35**] Terskikh, A. V.; Potekhin, S. A.; Melnik, T. N.; Mach, J. P.; Kajava, A. “ Mutation Gln54Leu of the conserved polar residue in the interfacial coiled coil position (d) results in significant stabilization of the original structure of the COMP pentamerization domain“ *Lett. Pept. Sci*. **1997**, 4, 297−304.

[**36**] Guo, Y.; Kammerer, R. A.; Engel, J. “ The unusually stable coiled-coil domain of COMP exhibits cold and heat denaturation in 4–6 M guanidinium chloride“ *Biophys. Chem.* **2000**, *85*, 179−186

[**37**] Hu, J. C.; Newell, N. E.; Tidor, B.; Sauer, R. T. “Probing the roles of residues at the e and g positions of the GCN4 leucine zipper by combinatorial mutagenesis” *Protein Sci.* **1993**, 2, 1072−1084.

[**38**] Durr, E.; Jelesarov, I.; Bosshard, H. R. "Extremely Fast Folding of a Very Stable Leucine Zipper with a Strengthened Hydrophobic Core and Lacking Electrostatic Interactions between Helices" *Biochemistry* **1999**, 38, 870−880.

[**39**] Marti, D. N.; Bosshard, H. R. "Electrostatic Interactions in Leucine Zippers: Thermodynamic Analysis of the Contributions of Glu and His Residues and the Effect of Mutating Salt Bridges" *J. Mol. Biol.* **2003**, *330*, 621−637.

[**40**] Zheng, X.; Zhu, H.; Lashuel, H. A.; Hu, J. C. " Oligomerization properties of GCN4 leucine zipper e and g position mutants" *Protein Sci*. **1997**, 6, 2218−2226.

[**41**] Krylov, D.; Mikhailenko I.; Vinson C. "A thermodynamic scale for leucine zipper stability and dimerization specificity: e and g interhelical interactions" *EMBO J*. **1994**, 13, 2849−2861

[**42**] Meier, M.; Lustig, A.; Aebi, U.; Burkhard, P. "Removing an Interhelical Salt Bridge Abolishes Coiled-Coil Formation in a *de Novo* Designed Peptide" *J. Struct. Biol.* **2002,** *137*, 65−72

[**43**] Beck, K.; Gambee, J. E.; Kamawal A.; Bächinger H. P. "A single amino acid can switch the oligomerization state of the α-helical coiled-coil domain of cartilage matrix protein" *EMBO J.* **1997**, *16*, 3767−3777.

[**44**] Schnarr, N. A.; Kennan, A. J. "pH-Switchable Strand Orientation in Peptide Assemblies" *Org. Lett*. **2005**, 7, 395−398.

[**45**] Sottile, J.; Selegue, J.; Mosher, D. F. "Synthesis of truncated amino-terminal trimers of thrombospondin" *Biochemistry* **1991**, 30, 6556−6562.

[**46**] Frank, S.; Kammerer, A. K.; Hellstern, S.; Pegoraro, S.; Stetefeld, J.; Lustig, A.; Moroder, L.; Engel, J. " Toward a High-Resolution Structure of Phospholamban: Design of Soluble Transmembrane Domain Mutants" *Biochemistry* **2000**, 39, 6825−6831.

[**47**] Slovic, A. M.; Summa, C. M.; Lear, J. D.; DeGrado, W. F. "Computational design of a water-soluble analog of phospholamban" *Protein Sci*. **2003**, 12, 337−348.

[**48**] Slovic, A. M.; Lear, J. D.; DeGrado, W. F. "De novo design of a pentameric coiled-coil: decoding the motif for tetramer versus pentamer formation in water-soluble phospholamban" *J. Pept. Res*. **2005**, 65, 312−321.

[**49**] Slovic, A. M.; Stayrook, S. E.; North, B.; DeGrado, W. F. "X-ray Structure of a Water-soluble Analog of the Membrane Protein Phospholamban: Sequence Determinants Defining the Topology of Tetrameric and Pentameric Coiled Coils" *J. Mol. Biol.* **2005**, *348*, 777−787

[**50**] Ramos, J.; Lazaridis, T. "Structural and Energetic Basis of Coiled Coil Topology" [Publication Pending]

[**51**] McClain D. L.; Binfet J. P.; Oakley M. G. "Evaluation of the energetic contribution of interhelical coulombic interactions for coiled coil helix orientation specificity." *J. Mol. Biol.* **2001**, 313(2): 371-383

[**52**] Monera O. D.; Zhou N. E.; Kay C. M.; Hodges R. S. "Comparison of antiparallel and parallel two-stranded alpha-helical coiled-coils. Design, synthesis, and characterization" *J. Biol. Chem.* **1993**, 268(26): 19218-19227

[**53**] Gurnon D. G.; Whitaker J. A.; Oakley M. G.. "Design and Characterization of a Homodimeric Antiparallel Coiled Coil." *J. Am. Chem. Soc.* **2003**, 125(2): 7518 -7519

[**54**] Wagschal K, T. B., Lavigne P, Mant C, Hodges R. S. "The role of position a in determining the stability and oligomerization state of a-helical coiled coils: 20 amino acid stability coefficients in the hydrophobic core of proteins" *Protein Science* **1999**, 8(11): 2312-2329

[**55**] Bourbigot S.; Beltz H.; Denis J.; Morellet N.; Roques B. P.; Y.Mély, S.Bouaziz "The C-terminal domain of the HIV-1 regulatory protein Vpr adopts an antiparallel dimeric structure in solution via its leucine-zipper-like domain" *Biochem J* **2005**, 387: 333-341

[**56**] Zoetewey D. L.; Tripet B. P.; Kutateladze T. G.; Overduin M. J.; Wood J. M.; Hodges R. S. "Solution structure of the C-terminal antiparallel coiled-coil domain from Escherichia coli osmosensor ProP" *J Mol Biol*. **2003**, 334(5): 1063-1076

[**57**] Tsatskis, Y.; Kwok, S. C.; Becker, E.; Gill, C.; Smith, M. N.; Keates, R. A. B.; Hodges, R. S.; Wood, J. M. "Core Residue Replacements Cause Coiled-Coil Orientation Switching in Vitro and in Vivo: Structure-Function Correlations for Osmosensory Transporter ProP" *Biochemistry* **2008**, 47(1); 60-72

[**58**] Hillar, A.; Culham D. E.; Vernikovska Y. I.; Wood J. M.; Boggs J. M. "Formation of an Antiparallel, Intermolecular Coiled Coil Is Associated with in Vivo Dimerization of Osmosensor and Osmoprotectant Transporter ProP in *Escherichia coli*" *Biochemistry* **2005**, 44(30); 10170-10180

[**59**] Cusack S.; Berthet-Colominas C.; Härtlein M.; Nassar N.; Leberman R. "A second class of synthetase structure revealed by X-ray analysis of Escherichia coli seryl-tRNA synthetase at 2.5 Å" *Nature* **1990**, 347: 249-255

[**60**] McClain D. L.; Gurnon D. G.; Oakley M. G. "Importance of Potential Interhelical Salt-bridges Involving Interior Residues for Coiled-coil Stability and Quaternary Structure" *J Mol Biol* **2002**, 324(2): 257-270

[**61**] Holton J.; Alber T. (2004). "Automated protein crystal structure determination using ELVES." *PNAS USA* **2004**, 101(6): 1537-1542

[**62**] Gonzalez, Jr. L.; Plecs, J. J.; Alber, T. “An engineered allosteric switch in leucine-zipper oligomerization” *Nature Struct. Biol.* **1996,** 3:510−515

[**63**] Schnarr N. A.; Kennan A. J. “Strand orientation by steric matching: a designed antiparallel coiled-coil trimer“ *J Am Chem Soc*. **2004**, 126(44):14447-14451

[**64**] Lovejoy B.; Choe S.; Cascio D.; McRorie D. K.; DeGrado W.F.; Eisenberg D. "Crystal Structure of a Synthetic TripIe-Stranded alpha-Helica1 Bundle" *Science* **1993,** 259(50099) 1288-1293

[**65**] Betz S. F.; DeGrado W. F. "Controlling Topology and Native-like Behavior of de Novo-Designed Peptides: Design and Characterization of Antiparallel Four-Stranded Coiled Coils" *Biochemistry* **1996**, 35(21): 6955-6962

[**66**] Yadav, M. K., Leman, L.J., Price, D.J., Brooks III, C.L., Stout, C.D., Ghadiri, M.R. " Coiled Coils at the Edge of Configurational Heterogeneity. Structural Analyses of Parallel and Antiparallel Homotetrameric Coiled Coils Reveal Configurational Sensitivity to a Single Solvent-Exposed Amino Acid Substitution" *Biochemistry* **2006**, 45(14): 4463-4473

[**67**] Whitson S. R.; LeStourgeon W. M.; Krezel A. M. “Solution Structure of the Symmetric Coiled Coil Tetramer Formed by the Oligomerization Domain of hnRNP C: Implications for Biological Function” *J Mol Biol* **2005**, 350(2), 319-337

[**68**] Liu J.; Zheng Q.; Deng Y.; Li Q.; Neville R. K.; Lu M. “Conformational Specificity of the *Lac* Repressor Coiled-Coil Tetramerization Domain” *Biochemistry* **2007**, 46(51):14951-14959

[**69**] Monera O. D.; Kay C. M.; Hodges R. S. "Electrostatic Interactions Control the Parallel and Antiparallel Orientation of α-Helical Chains in Two-Stranded α-Helical Coiled-Coils." *Biochemistry* **1994**, 33(13): 3862-3871

[**70**] Hadley E. B.; Testa O. D.; Woolfson D. N.; Gellman S. H. “Preferred side-chain constellations at antiparallel coiled-coil interfaces” *PNAS USA* **2008** 105:530-535

[**71**] Brooks B. R.; Bruccoleri R. E.; Olafson B. D.; States D. J.; Swaminathan S.; Karplus M. “CHARMM: a program for macromolecular energy, minimization, and dynamics calculations” *J. Comput. Chem.* **1983**, *4*, 187−217.

[**72**] Lazaridis, T.; Karplus, M. “Effective energy function for proteins in solution“ *Proteins* **1999**, *35*, 133−152

[**73**] Crick, F.H.C. "The Fourier transform of a coiled-coil" Acta Crystallogr. **1953**, 6(8-9):685-689

[**74**] Canutescu, A. A.; Shelenkov, A. A.; Dunbrack, R. L., Jr. "A graph-theory algorithm for rapid protein side-chain prediction" *Protein Sci.* **2003**, *12*, 2001−2014

[**75**] Lazaridis, T. (2003). Effective energy function for proteins in lipid membranes. *Proteins: Struct, Funct, Genet.* **52**:176-192

[**76**] Masunov, A.; Lazaridis, T. "Potentials of Mean Force between Ionizable Amino Acid Side Chains in Water" *J. Am. Chem. Soc.* **2003**, *125*, 1722−1730

[**77**] Evans, D. J.; Holian, B. L. "The Nose-Hoover thermostat" *The Journal of Chemical Physics*, 1985, 83(8), 4069-4074

[**78**] Lazaridis, T.; Masunov, A.; Gandolfo, F. "Contributions to the binding free energy of ligands to avidin and streptavidin" *Proteins: Struct., Funct., Genet.* **2002**, *47*, 194−208

[**79**] Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. "The statistical-thermodynamic basis for computation of binding affinities: a critical review" *Biophys. J.* **1997**, *72*, 1047−1069

[**80**] Krystek, S. R., Jr.; Bruccoleri, R. E.; Novotny, J. "Stabilities of leucine zipper dimers estimated by an empirical free energy method" *Int. J. Pept. Protein Res.* **1991**, *38*, 229−236

[**81**] Nilges, M.; Brunger, A. T. " Automated modeling of coiled coils: application to the GCN4 dimerization region" *Protein Eng.* **1991**, *4*, 649−659.

[**82**] Nilges, M.; Brunger, A. T. " Successful prediction of the coiled coil geometry of the GCN4 leucine zipper domain by simulated annealing: Comparison to the X-ray structure" *Proteins: Struct., Funct., Genet.* **1993**, *15*, 133−146.

[**83**] DeLano, W. L.; Brunger, A. T. "Helix packing in proteins: prediction and energetic analysis of dimeric, trimeric, and tetrameric GCN4 coiled coil structures" *Proteins* **1994**, *20*, 105−123.

[**84**] Harbury, P. B.; Tidor, B.; Kim, P. S. "Repacking protein cores with backbone freedom: structure prediction for coiled coils" *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92,* 8408−8412.

[**85**] Zhang, L.; Hermans, J. "Molecular dynamics study of structure and stability of a model coiled coil" *Protein Eng.* **1993**, *16*, 384−392.

[**86**] Vieth, M.; Kolinski, A.; Brooks, C. L., III, Skolnick, J. "Prediction of quaternary structure of coiled coils. Application to mutants of the GCN4 leucine zipper" *J. Mol. Biol.* **1995**, *251*, 448−467.

[**87**] Vieth, M.; Kolinski, A.; Skolnick, J. "Method for Predicting the State of Association of Discretized Protein Models. Application to Leucine Zippers" *Biochemistry* **1996**, *35*, 966−967.

[**88**] Mohanty, D.; Kolinski, A.; Skolnick, J. "De novo simulations of the folding thermodynamics of the GCN4 leucine zipper" *Biophys. J.* **1999**, *77*, 54−59.

[**89**] Viñals, J.; Kolinski, A.; Skolnick, J. "Numerical study of the entropy loss of dimerization and the folding thermodynamics of the GCN4 leucine zipper" *Biophys. J.* **2002**, *83*, 2801−2811

[**90**] Kenar, K. T.; García-Moreno, B.; Freire, E. "A calorimetric characterization of the salt dependence of the stability of the GCN4 leucine zipper" *Protein Sci.* **1995**, *4*, 1934−1938

[**91**] Goodman, E. M.; Kim, P. S. "Periodicity of amide proton exchange rates in a coiled-coil leucine zipper peptide" *Biochemistry* **1991**, *30*, 11615−11620

[**92**] Lumb, K. J.; Kim, P. S. "Measurement of interhelical electrostatic interactions in the GCN4 leucine zipper" *Science* **1995**, *268*, 436−439

[**93**] Guo, Y.; Bozic, D.; Malashkevich, V. N.; Kammerer, R. A.; Schulthess, T.; Engel, J. "" *EMBO J.* **1998**, *17*, 5265−5272

[**94**] Hu, J. C.; Newell, N. E.; Tidor, B.; Sauer, R. T. "" *Protein Sci.* **1993**, *2,* 1072−1084

[**95**] Chen, H.; Aeschlimann, D.; Nowlen, J.; Mosher, D. F. "Expression and initial characterization of recombinant mouse thrombospondin 1 and thrombospondin 3" *FEBS Lett.* **1996**, *387*, 36−41.

[**96**] Liu, J.; Zheng, Q.; Deng, Y.; Kallenbach, N. R.; Lu, M. "Conformational transition between four and five-stranded phenylalanine zippers determined by a local packing interaction" *J. Mol. Biol.* **2006**, *361*, 168−179

[**97**] Barth P.; Schoeffler A.; Alber T. "Targeting Metastable Coiled-Coil Domains by Computational Design" *J. Am. Chem. Soc.*, **2008**, *130* (36), pp 12038–12044

[**98**] Ogihara, N.L., Weiss, M.S., Degrado, W.F., Eisenberg, D. "The crystal structure of the designed trimeric coiled coil coil-VaLd: implications for engineering crystals and supramolecular assemblies" *Protein Sci*. **1997**, 6**:** 80-88

[**99**] Fujiwara Y.; Minor Jr D. L. **"**X-ray Crystal Structure of a TRPM Assembly Domain Reveals an Antiparallel Four-stranded Coiled-coil**"** *JMB* **2008**, 383(4), 854-870

[**100**] Yan Y.; Winograd E.; Viel A.; Cronin T.;Branton C. D. "Crystal Structure of the Repetitive Segments of Spectrin" *Science, New Series*, **1993**, 262(5142): 2027-2030

INDEX