Structural Specificity in Coiled Coils: *a* and *d* Position Polar Residues

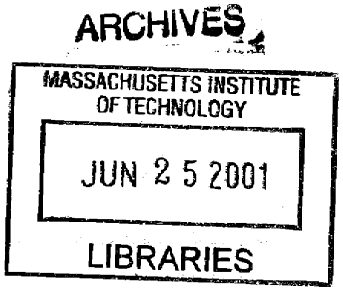
By

David L. Akey

B.S. Biochemistry
University of California, Davis, 1995

Submitted to the Department of Biology in Partial Fulfillment
of the Requirements for the Degree of

Doctor of Philosophy
at the
Massachusetts Institute of Technology



September, 2001

Signature of Author: _____
Department of Biology
June 21, 2001

Certified by: _____
Peter S. Kim
Professor of Biology
Thesis Supervisor

Accepted by: _____
Terry Orr-Weaver
Professor of Biology
Chairman, Committee for Graduate Students

Dedication

For Györgyi and Christopher

Structural Specificity in Coiled Coils: *a* and *d* Position Polar Residues

By

David L. Akey

Submitted to the Department of Biology
on (date) in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy in
Biology

ABSTRACT

Experimental studies were performed to determine the effects of single polar residues at the ***a*** and the ***d*** positions of a reference coiled coil, GCN4-pVL. The reference coiled coil is very stable in solution and exists as a mixture of dimers and trimers. The placement of single polar residues in the otherwise hydrophobic core of GCN4-pVL has dramatic effects on both stability and oligomeric specificity. The effects vary with regard to both the identity and the position (***a*** vs ***d*** position) of the polar substitution.

The ***d*** position is more sensitive to polar residues. Two residues, asparagine and glutamine, were found to be much more destabilizing when placed at ***d*** positions than any residues at the ***a*** positions. In addition to the known ability of a single asparagine at the ***a*** position to specify coiled-coil dimers, it was found that a single threonine at the ***d*** postion can specify a coiled-coil trimer.

The crystal structures of four coiled coils, with either a single serine or threonine at either the ***a*** or the ***d*** position, were determined. These structures show that threonine residues tend to form intra-helical hydrogen bonds when packed in the coiled-coil core. Serine residues tend to show more variability in their packing when placed in the core positions.

Polar residues can affect the local coiled-coil geometry. The most dramatic effect was as a result of serine residues at the ***d*** position in which in a local decrease in the supercoil radius of 0.5Å centered around the buried serine position was observed.

Thesis Supervisor: Peter S. Kim
Title: Professor of Biology

Table of Contents

Chapter 1

An Introduction to Coiled-Coil Structure

Predicting protein function from sequence has long been a holy grail of biologists. An intermediate step is to predict the fold of a protein from sequence. Most predictions of protein function and protein fold are knowledge based, that is, they are founded in the similarity of a test sequence to a sequence of known function and/or structure. All known protein folds are the result of structural studies with the exception of the coiled coil, the structure of which was proposed and simple heptad repeat noted well before experimental data confirmed these predictions.

The coiled coil is composed of two or more α-helices wrapped about each other in a supercoiled manner (figure 1). Coiled coils have been observed to form both left-handed supercoils, which have a heptad repeat $(\boldsymbol{abcdefg})_n$, and right-handed supercoils, which have an undecad repeat. The vast majority of naturally occurring coiled coils are left-handed supercoils with a heptad repeat.

The defining feature of the heptad repeat is the presence of hydrophobic residues at the buried first and fourth (***a*** and ***d***) positions (figure 1). Additionally, there are frequently charged residues at the boundary fifth and seventh (***e*** and ***g***) positions, which often form interstrand salt bridges. The simple nature of the coiled-coil repeat makes it readily identifiable by computer algorithms (*1-3*). Database analysis of representative genomes predicts that 3–5% of all residues are likely to fold as coiled coils (figure 2) (*3*), making the coiled coil the single most common fold in biological systems.

The coiled-coil structure was proposed simultaneously by Crick, and by Pauling and Corey (*4, 5*) to provide a basis for the x-ray diffraction pattern observed in α-keratin fibers. Pauling and Corey had recently proposed the α-helix with 3.7 residues per turn (now known to have 3.6 residues per turn) as one of several potential folding motifs for

polypeptide chains (*6*). The standard 3.7 α-helix as proposed should have a strong x-ray reflection at 5.4 Å displaced from the meridian. However, a main feature of the observed α-keratin diffraction pattern was a 5.15 Å reflection very close to the meridian. Crick and Pauling separately proposed that this reflection could be explained if one assumed that the α-helices were arranged at an angle to each other and wrapped about each other in a supercoiled manner. Shortly thereafter, Crick devised a geometric model for such a supercoil and proposed a "knobs-into-holes" packing arrangement for the side chains at the interface (***a*** and ***d***) positions (*7*). The heptad repeat of coiled coils falls naturally from the repeating nature of the structure proposed by Crick. Amino acid sequence analysis of fibrous proteins thought to form coiled coils confirmed the heptad repeat with nonpolar residues at the ***a*** and ***d*** positions (*8-10*).

The first high resolution crystal structure of a coiled-coil dimer, the leucine zipper of the bZIP protein GCN4 (*11*), confirmed many of the details first proposed 38 years earlier by Crick (*7*), including the "knobs-into-holes" packing and the ability to describe the overall supercoil structure by a simple set of geometric parameters. These parameters (figure 3) define the radius and pitch of both the supercoil and α-helix, and the phase angle of the ***a***-position α-carbon atom. In addition to the geometric regularity of the structure, two other features of the GCN4 leucine zipper structure (also known as GCN4-p1) are now recognized as important determinants of coiled-coil structure and specificity: interhelical salt bridges between the ***g*** and ***e'*** (i to i '+ 5) residues, and buried polar interactions between the composite helices.

COILED COILS IN BIOLOGICAL SYSTEMS

The earliest studied coiled coils are long, fibrous proteins including intermediate filaments, the motor protein myosin, and muscle regulatory protein tropomyosin. These proteins are involved in structural and contractile functions. The intermediate filament (IF) family includes keratins, vimentin, desmin, and neurofilament proteins, and the closely related nuclear lamins and tektins (*12, 13*). IFs generally form extended 10-nm diameter filaments thought to consist of 16 coiled-coil dimers in cross section. The IFs are generally very stable and have many disparate, structural functions: keratins, expressed in epithilial cells, form skin, hair, and nails; neurofilaments form the core of neuronal axons; lamins form the nuclear lamina that lines the nuclear envelope, the assembly and dissambly of which is regulated by phosphorylation and is an important part of the cell cycle; and vimentins, an important structural element of the cellular cytoskeleton, anchor desmosomes and hemidesmosomes and are important for cell adhesion and stability.

Coiled coils play an important role in a variety of motor proteins, including myosins, dyneins, and kinesins. In both the myosins and kinesins, coiled coils form the stalk of a functional dimer (*14, 15*). In muscle myosin, the dimer then assembles into the higher order thick-filament structure. The coiled coil of kinesin separates the N-terminal motor domain from the C-terminal cargo domain. The role of the coiled-coil region of dynein is less clear, although it is thought to form the stalk separating the head domain from the tubulin binding domain (*16*). The length of this stalk is conserved and seems to be an important determinant in the arrangement of dynein along the microtubule surface.

The association of the coiled-coil protein tropomyosin with actin fibers is an important regulatory mechanism in muscle contraction.

Another major functional role of coiled coils is transcriptional regulation. In 1989, it was shown that the leucine zipper region of the bZIP transcriptional regulators were coiled coils (*11, 17*). The bZIP proteins are of particular interest because as a class, they not only form homotypic dimers (e.g., yeast transcriptional regulator GCN4), but also a variety of heterotypic dimers (e.g., oncogenes Fos and Jun, Myc and Max (*18*)). Of particular note is the ability of Jun to form a variety of heterodimers with the Fos and ATF1-4 proteins (*19*). A more complete understanding of the formation of heterotypic structures has important implications in transcriptional regulation.

In addition, coiled coils are a central component of a variety of membrane fusion proteins. The structure of influenza hemagglutinin (HA) includes a central coiled-coil trimer that connects three identical subunits (*20, 21*). This motif is also seen in the structure of the core particle of the TM domain of the Moloney murine leukemia virus envelope protein, the GP41 envelope protein of both HIV and SIV, and in the envelope proteins of the Ebola virus (*22-31*). The repertoire of membrane fusion proteins that utilize coiled coils has since been expanded to include the SNAP-SNARE vesicle fusion complex, which has an unusual four-helix heterotypic structure (*32*).

COILED COIL PREDICTION AND DESIGN

Although the basic structure of the coiled coil has long been known there are variations, including the number of strands, orientation of the composite strands, and composition of the strands (figure 1C,D,E). Crystal structures show coiled coils that

consist of 2–5 individual strands (*11, 33-35*). The composite strands of a coiled coil may be arranged such that the N-termini of all strands are aligned in a parallel manner, or the N-terminus of one strand is aligned with the C-terminus(i) of a partner strand(s) in an antiparallel manner. Additionally, coiled coils may be homotypic, i.e., all strands with the same sequence, or heterotypic, i.e., different strands with different sequences. Although the variations that lead to these differences are not well enough understood to make reliable predictions in coiled-coil structure, there has been successful design of coiled coils with predetermined properties (e.g., strand number, orientation, and composition),.

The simplest aspect of coiled-coil prediction is determining if a protein sequence is likely to form a coiled coil. Algorithms to address this issue are based on the distribution of residues at each position in the coiled-coil heptad observed in databases of known coiled-coil sequences (*1-3*). The COILS program aligns a test sequence to a heptad repeat and compares the occurrence of each residue at the given positions in the heptad to the observed frequencies of those residues at the equivalent positions in the database (*1*). PairCoil (and later MultiCoil) improved on this method by comparing pairwise frequencies in the submitted sequence against the pairwise frequencies in databases of known coiled coils, and have a lower false-positive rate (*2, 3*). MultiCoil also predicts oligomeric state (dimer or trimer) of the submitted sequence.

An ancillary benefit of the MultiCoil program was the creation of oligomer-specific coiled-coil databases. These databases of coiled-coil dimers and trimers highlight the differences in residue distribution at the core ***a*** and ***d*** positions as a function of oligomeric state. Interestingly, a shared feature of both dimers and trimers is the

relatively high number of polar and charged residues at the ***a*** and ***d*** positions, with approximately 20% of these buried positions charged or polar (table 1). Although the presence of buried polar and charged groups is a conserved feature, the exact roles of these residues is not fully understood. There is potential for buried polar and charged residues to affect oligomeric state, strand orientation, partner selectivity, and proper register alignment of the composite strands of a coiled coil.

Oligomeric Specificity

The most important factor in determining the oligomeric state of coiled coils is the identity of the residues at the ***a*** and ***d*** positions. Variants of GCN4-p1, with different hydrophobic residues at the ***a*** and ***d*** positions, have different oligomeric states (table 2) (*35*). For example, a GCN4-p1 variant with all isoleucines at the ***a*** positions and leucines at the ***d*** positions (here termed GCN4-pIL) is dimeric in solution, while a variant with isoleucines at both the ***a*** and ***d*** positions (GCN4-pII) is trimeric in solution (*35*). The trimeric state for GCN4-pII correlates with the residue distribution observed in comparisons of dimeric and trimeric coiled coils (*3, 36*) (table 1), where isoleucines in the ***d*** positions of trimers are found 2.7 times more frequently than isoleucines in the ***d*** positions of dimers. While these experiments show the oligomeric state of coiled coils with homogeneous cores, most coiled coils have a mixture of residues at the ***a*** and ***d*** positions; thus, more experiments are required to determine the relative strength of the oligomeric preferences observed for residues in the core positions and how these preferences are balanced.

The effects of mixed cores on oligomeric state for coiled coils has been addressed by studies in which model coiled coils with valines at the ***a*** positions and leucines at the ***d*** positions are used as probes of oligomeric specificity (*37-40*). The GCN4-p1-based variant with this core composition (GCN4-pVL) is a mixture of dimers and trimers in solution (*35*) (table 2), and changes in oligomeric state affected by single substitutions at either an ***a*** or ***d*** position can be observed. As the wild type dimeric GCN4-p1 sequence differs from GCN4-pVL by a single ***a***-position asparagine, it is inferred that ***a***-position asparagines are important dimer determinants. This is supported by the structure of GCN4-p1 in which the buried asparagines are hydrogen bonded in the core of the dimer (figure 4A) (*11*). Similarly, ***a***-positions lysines that occur over 13 times more frequently in dimers than in trimers (Table 1) also specify dimers in the GCN4-pVL test case (*39*). The crystal structure of the ***a***-position lysine variant reveals a different mechanism for dimer specificity in which the lysine residues are directed toward the surface of the coiled coil (figure 4B) (*39*). Lysines in the ***a*** position of trimers would have more limited access to the surface and thus would be more destabilizing.

These examples provide a physical basis for the observed frequencies of certain residues at specific core positions, however, other examples have been less clear. A GCN-pVL variant with a single ***a***-position glutamine , GCN4-pV_QL, does not have oligomeric specificity, even though ***a***-position glutamines are found 2.6 times more frequently in trimers than in dimers (Table 1). Crystal structures of this peptide were solved as both a dimer and a trimer, with asymmetric glutamines in the trimer structure (figure 4C) (*38*). Related peptides, with isoleucines at both ***a*** and ***d*** positions, except for a single ***a***-position glutamine, have a different packing arrangement than seen in GCN4-

pV_QL (figure 4C) (*41, 42*). The isoleucine core variant structures have symmetric glutamines coordinating a central chloride ion, while the GCN4-pV_QL structure has asymmetric glutamines and two coordinated water molecules. The differences in glutamine packing may simply be due to the lack of halides in the GCN4-pV_QL crystal conditions, or to the different core composition, illustrating the importance of context when considering effects of core-position residues.

Partner Selection

Many coiled-coils assemble as heterotypic complexes and are selective as to with which partner a given strand will pair. Some coiled-coil proteins, particularly bZIP transcription factors, have multiple partners (*16*). Studies of the Fos-Jun heterodimer show that ***g*** to ***e'*** salt bridges are important determinants in coiled-coil selectivity (*43*). These studies transferred the partner selectivity of the Fos-Jun heterodimer to GCN4-p1 solely by swapping residues at the ***e*** and ***g*** positions. Several groups used ***g*** to ***e'*** salt bridges as determinants of partner selectivity in the design of obligate heterotypic dimers and trimers (*44, 45*). In addition to surface charge interactions, core residues also affect partner selection. The most obvious determinant is that core-position polar residues should align opposite a similar residue on the partner strand. A particularly intriguing example of buried polar and charged residues is the coordination of a buried arginine with three glutamines in the SNAP-SNARE complex (figure 5D). This cluster will not accommodate a second arginine, but will function with the substitution of a fourth glutamine for the conserved arginine (*46, 47*).

Genetic selections for heterotypic dimers confirm the importance of both surface ***g*** to ***e'*** interactions and alignment of buried polar residues in partner selection (*48, 49*). Interestingly, these experiments do not always maximize the number of complementary charge-charge interactions. Destabilization of respective homotypic dimers is also a strong selective force indicating that repulsive interactions are also strong determinants. These selections almost invariably result in paired asparagines at equivalent positions in partner strands. While this result stresses the importance of matching aspargines in coiled-coil cores, not all polar residues are likely to have the same effect. For example, in the Fos-Jun structure an ***a***-position threonine is aligned opposite a valine residue (*50*). In this structure, the terminal threonine hydroxyl forms an intra-chain hydrogen bond. Similarly, basic lysine and arginine residues probably do not have a role in partner selection, as suggested by structures with ***a***-position lysines. ***a***-position lysines direct the terminal charged nitrogen away from the core (figure 4B), burying only the hydrophobic region of the side chain (*39*).

Strand Orientation

Many of the same elements that affect partner selection also affect strand orientation. To a first approximation, strand orientation is a problem of partner selection; instead of choosing among different strands, the choice is among two orientations of a single partner strand. Strand orientation can be directed by ***a***- and ***d***-position residues. Antiparallel orientation can be specified by polar residue placement. A heterodimer, designed such that a parallel alignment will pack asparagine against hydrophobic residues

but an antiparallel alignment will pack asparagine with a complementary asparagine, is antiparallel (*51*).

Molecular Modeling of Coiled Coils

The small size and repeating pattern of the coiled-coil fold make it a good system for molecular modeling. Coiled coils have been modeled using traditional molecular dynamics techniques, the dead-end elimination algorithm, and by a flexible backbone method that takes advantage of the algebraic parameterization devised by Crick (*52-54*). The most impressive of the modeling results used Crick's parameterization in combination with molecular dynamics and culminated in the de novo design of right-handed coiled coils. This model not only correctly specified the oligomeric state, but also the precise packing of the core residues (*55*). These efforts are impressive, but difficulty arises when attempting to predict the packing of buried polar and charged residues. These residues may not always be directed toward the interior. Buried polar and charged residues coordinate water molecules and ions in ways that are not simple to model or predict. Packing is not necessarily conserved from structure to structure. Also, as outlined in the second chapter of this thesis, buried polar residues have topological effects on the backbone geometry of the coiled coil.

Traditionally, the most successful structure predictions relied on knowledge-based methods. To extend coiled-coil predictions to partner selectivity and strand orientation would require a sizeable database of heterotypic coiled coils and of antiparallel coiled coils. However, there are currently too few examples of these variations to build

statistically meaningful residue distributions. Progress can be made in several areas. High-throughput screens for coiled-coil interactions can be useful for increasing the number of known heterotypic coiled coils. Continued work with designed sequences will hopefully allow us to understand partnering issues more fully. Continual improvement of modeling methods may allow for computational screens of coiled-coil interactions. Hopefully, a combination of the above approaches will result in the prediction of the variations on coiled-coil interactions with reasonable confidence.

References

1. Parry, D. A. (1982) *Biosci. Rep. 2*, 1017-24.

2. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. USA 92*, 8259-63.

3. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Sci. 6*, 1179-89.

4. Crick, F. (1952) *Nature 170*, 882-883.

5. Pauling, L., and Corey, R. B. (1953) *Nature 171*, 59-61.

6. Pauling, L., Corey, R. B., and Branson, H. R. (1951) *Proc. Nat. Acad. Sci. 37*, 205-211.

7. Crick, F. H. C. (1953) *Acta Crystallogr. 6*, 689-697.

8. Sodek, J., Hodges, R., Smillie, L., and Jurasek, L. (1972) *Proc Natl Acad Sci U S A 69*, p3800-4.

9. Parry, D., Crewther, W., Fraser, R., and MacRae, T. (1977) *J Mol Biol 113*, p449-54.

10. McLachlan, A. (1978) *J Mol Biol 124*, p297-304.

11. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science 254*, 539-544.

12. Fuchs, E., and Weber, K. (1994) *Annu Rev Biochem 63*, 345-82.

13. Parry, D., and Steinert, P. (1999) *Q Rev Biophys 32*, p99-187.

14. Barral, J., and Epstein, H. (1999) *Bioessays 21*, p813-23.

15. Hirokawa, N., Noda, Y., and Okada, Y. (1998) *Curr Opin Cell Biol 10*, p60-73.

16. Vallee, R., and Gee, M. (1998) *Trends Cell Biol 8*, p490-4.

17. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science 243*, 538-42.

18. Cohen, C., and Parry, D. (1994) *Science 263*, 488-9.

19. Hai, T., and Curran, T. (1991) *Proc. Nat. Acad. Sci. U S A 88*, 3720.

20. Weis, W. I., Brunger, A. T., Skehel, J. J., and Wiley, D. C. (1990) *J. Mol. Biol. 212*, 737-761.

21. Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994) *Nature 371*, 37-43.

22. Fass, D., Harrison, S. C., and Kim, P. S. (1996) *Nat. Struct. Biol. 3*, 465-469.

23. Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) *Cell 89*, 263-273.

24. Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) *Nature 387*, 426-430.

25. Malashkevich, V. N., Schneider, B. J., McNally, M. L., Milhollen, M. A., Pang, J. X., and Kim, P. S. (1999) *Proc Natl Acad Sci U S A 96*, p2662-7.

26. Caffrey, M., Cai, M., Kaufman, J., Stahl, S., Wingfield, P., Covell, D., Gronenborn, A., and Clore, G. (1998) *EMBO J 17*, p4572-84.

27. Malashkevich, V., Chan, D., Chutkowski, C., and Kim, P. (1998) *Proc Natl Acad Sci U S A 95*, p9134-9.

28. Blacklow, S., Lu, M., and Kim, P. (1995) *Biochemistry 34*, p14955-62.

29. Dessen, A., Volchkov, V., Dolnik, O., Klenk, H., and Weissenhorn, W. (2000) *EMBO J 19*, p4228-36.

30. Malashkevich, V., Schneider, B., McNally, M., Milhollen, M., Pang, J., and Kim, P. (1999) *Proc Natl Acad Sci U S A 96*, p2662-7.

31. Weissenhorn, W., Carfi, A., Lee, K., Skehel, J., and Wiley, D. (1998) *Mol Cell 2*, p605-16.

32. Sutton, R., Fasshauer, D., Jahn, R., and Brunger, A. (1998) *Nature 395*, p347-53.

33. Harbury, P. B., Kim, P. S., and Alber, T. (1994) *Nature 371*, 80-83.

34. Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T., and Engel, J. (1996) *Science 274*, 761-765.

35. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science 262*, 1401-1407.

36. Woolfson, D. N., and Alber, T. (1995) *Protein Sci. 4*, 1596-607.

37. Wagschal, K., Tripet, B., Lavigne, P., Mant, C., and Hodges, R. (1999) *Protein Sci. 8*, 2312-29.

38. Gonzalez Jr., L., Brown, R. A., Richardson, D., and Alber, T. (1996) *Nat. Struct. Biol. 3*, 1002-1009.

39. Gonzalez Jr., L., Woolfson, D. N., and Alber, T. (1996) *Nat. Struct. Biol. 3*, 1011-1018.

40. Tripet, B., Wagschal, K., Lavigne, P., Mant, C., and Hodges, R. (2000) *J. Mol. Biol. 300*, 377-402.

41. Nautiyal, S., and Alber, T. (1999) *Protein Sci 8*, p84-90.

42. Eckert, D. M., Malashkevich, V. M., and Kim, P. S. (1998) *J. Mol. Biol. 284*, 859-865.

43. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992) *Cell 68*, 699-708.

44. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) *Current Biology 3*, 658-667.

45. Nautiyal, S., Woolfson, D. N., King, D. S., and Alber, T. (1995) *Biochemistry 34*, 11645-11651.

46. Ossig, R., Schmitt, H., de, G. B., Riedel, D., Keranen, S., Ronne, H., Grubmuller, H., and Jahn, R. (2000) *EMBO J 19*, p6000-10.

47. Katz, L., and Brennwald, P. (2000) *Mol Biol Cell 11*, p3849-58.

48. Arndt, K., Pelletier, J., Muller, K., Alber, T., Michnick, S., and Pluckthun, A. (2000) *J Mol Biol 295*, p627-39.

49. Zeng, X., Herndon, A. M., and Hu, J. C. (1997) *Proc. Natl. Acad. Sci. USA 94*, 3673-8.

50. Chen, L., Glover, J. N., Hogan, P. G., Rao, A., and Harrison, S. C. (1998) *Nature 392*, 42-48.

51. Oakley, M. G., and Kim, P. S. (1998) *Biochemistry 37*, 12603-12610.

52. Rozzelle, J., Tropsha, A., and Erickson, B. (1994) *Protein Sci 3*, p345-55.

53. Dahiyat, B. I., and Mayo, S. L. (1996) *Protein Sci. 5*, 895-903.

54. Harbury, P. B., Tidor, B., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. USA 92*, 8408-12.

55. Harbury, P. B., Plecs, J., Tidor, B., Alber, T., and Kim, P. S. (1998) *Science 282*, 1462-7.

	a Position			*d* Position		
Residue	Dimer%	Trimer%	Trimer/Dimer	Dimer%	Trimer%	Trimer/Dimer
Leu	28.7	22.7	0.8	39.9	24.2	0.6
Ile	13.4	14.6	1.1	3.5	9.6	2.7
Val	12.7	13.5	1.1	6.4	11.8	1.8
Met	3.1	2.4	0.8	3.2	3.0	0.9
Phe	1.7	1.8	1.0	2.1	3.1	1.5
Tyr	3.1	3.2	1.0	3.4	2.0	0.6
Gly	0.3	1.2	4.6	0.4	0.8	2.0
Ala	10.4	21.3	2.1	20.3	21.1	1.0
Cys	2.1	0.0	0.0	0.3	0.0	0.0
Trp	0.0	0.4	36.6	0.7	0.4	0.6
Pro	0.0	0.0		0.0	0.0	
Lys	**9.3**	**0.8**	**0.08**	**1.1**	**3.1**	**2.9**
Arg	**5.3**	**1.1**	**0.2**	**0.1**	**1.7**	**17.7**
His	**1.3**	**0.9**	**0.7**	**0.7**	**2.1**	**3.0**
Glu	**1.2**	**1.5**	**1.3**	**5.5**	**1.0**	**0.2**
Asp	**0.0**	**0.8**		**1.3**	**0.6**	**0.4**
Gln	**0.9**	**2.4**	**2.6**	**1.8**	**1.6**	**0.9**
Asn	**3.9**	**0.7**	**0.2**	**1.3**	**3.1**	**2.4**
Ser	**2.0**	**5.1**	**2.5**	**4.1**	**2.7**	**0.7**
Thr	**0.7**	**5.6**	**8.1**	**3.6**	**8.2**	**2.3**
Polar/Charged:	**24.5**	**18.8**		**19.5**	**24.1**	

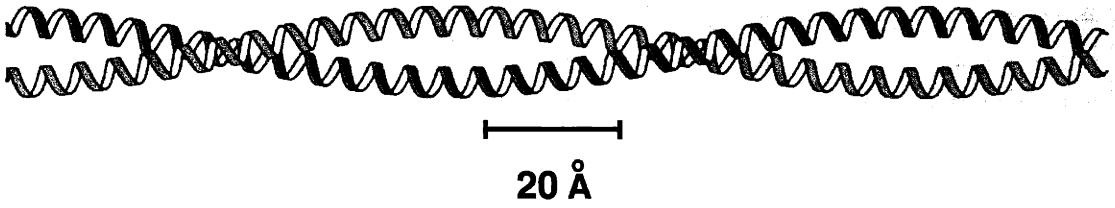
Table 1. Residue distribution at ***a*** and ***d*** positions in coiled-coil dimers and trimers. Data are from (*3*). Polar and charged residues are shown in boldface.

Positions		$-[\Theta]_{222}$	T_m	No. of helices	
a	***d***	(deg cm^2 dmol^{-1})	(°C)	Unmodified	SS
I	L	32,400	>100	2	2
I	I	32,400	>100	3	6
L	I	30,600	>100	4	4
V	I	22,500	73	-	(4,6)
L	V	30,600	81	3	(2,-)
V	L	32,400	95	(2,3)	2
L	L	31,500	>100	3	(2,4,6,-)

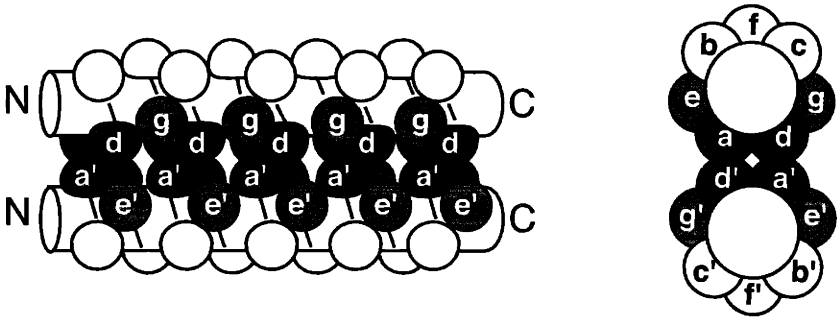
Table 2. Oligomeric state and stability of core position variants of GCN4-p1. The data are from (*35*). Residues were inserted at four ***a*** and ***d*** positions of GCN4-p1. The number of helices in the solution complex formed by unmodified peptides and by disulfide bonded peptides. Parentheses indicate multiple species were present; a dash indicates the presence of a species that could not be assigned.

Figure 1. Coiled coil structure. (A) Coiled-coil fold showing two α-helices wrapped around each other with a left handed supercoil. (B) Heptad repeat of a coiled coil. The ***a*** and ***d*** interface positions (red) are generally hydrophobic. The ***e*** and ***g*** positions (blue) are frequently charged and can form inter-helical salt bridges. (C,D,E) Common coiled-coil variations. (C) Crystal structures of coiled coils as a dimer (GCN4-p1)(*11*), trimer (GCN4-pII)(*33*), tetramer (GCN4-pLI)(*35*) or pentamer (COMP)(*34*). (D) Coiled coils may be homo-typic (all strands identical) or hetero-typic (composed of different strands. (E) Coiled coils may be parallel (N-termini aligned) or anti-parallel (N-terminal aligned with C-terminal of partner strand).

A.



B.



C.

D.

E.

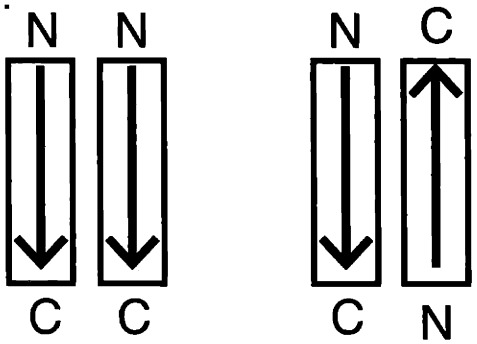


Figure 2. Predicted percentage of open reading frames containing coiled coils in a few representative genomes. Predictions are based on the program MultiCoil(*3*) run against the open reading frames of the indicated genomes. A 50% probability cutoff was used in determining which proteins are likely to contain coiled coils.

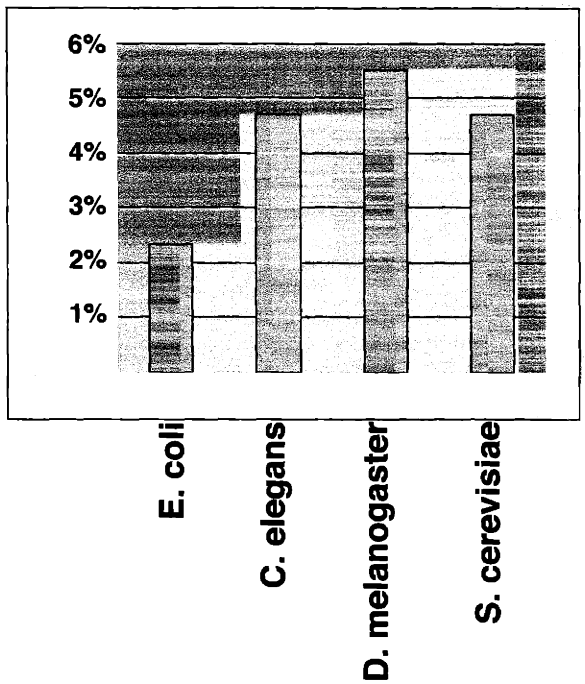

6%
5%
4%
3%
2%
1%
E. coli
C. elegans
D. melanogaster
S. cerevisiae

Figure 3. Crick supercoil parameters(*7, 54*). Coiled coils can be described by a simple set of parameters: R_0 and R_1, the supercoil and α-helix radii, respectively; ω_0 and ω_1, the supercoil and α-helix pitches, respectively; ϕ, the α-carbon phase angle, defined as the angle between the vector from the α-helix center to the supercoil center and the vector from the α-helix center to the ***a*** position C_α atom.

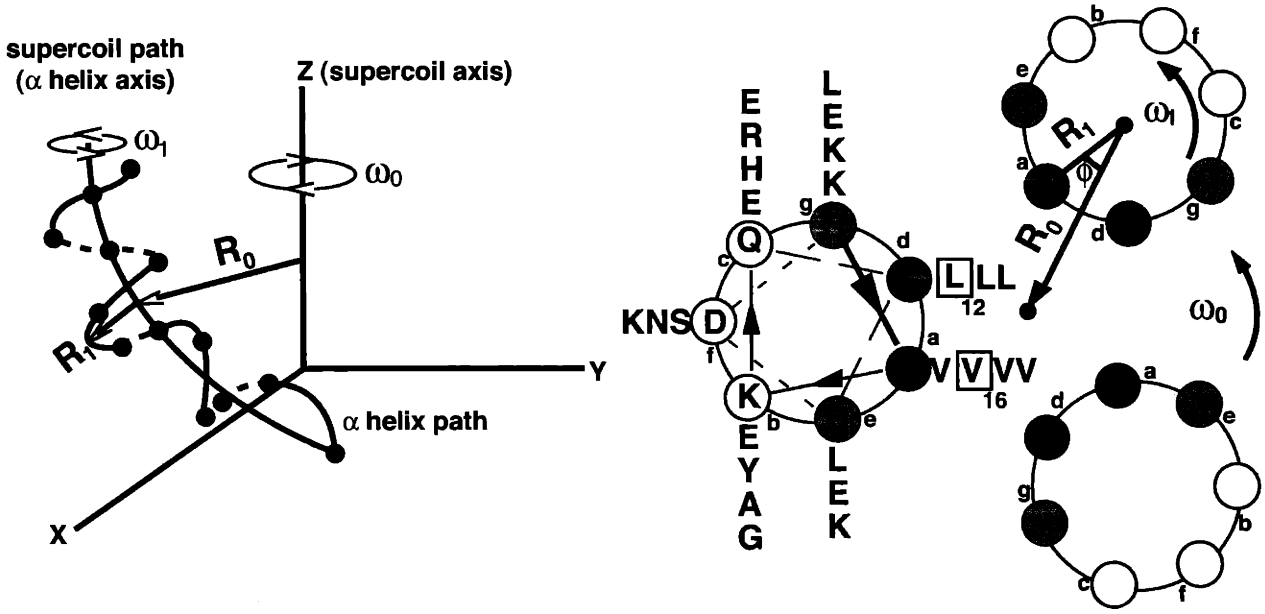
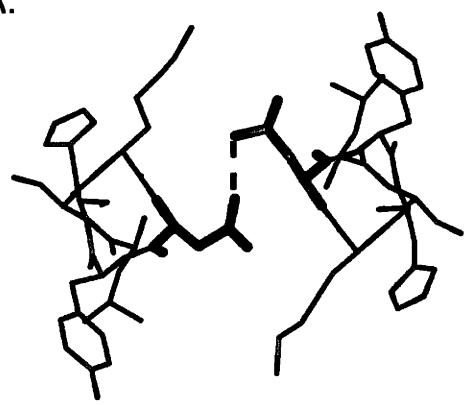
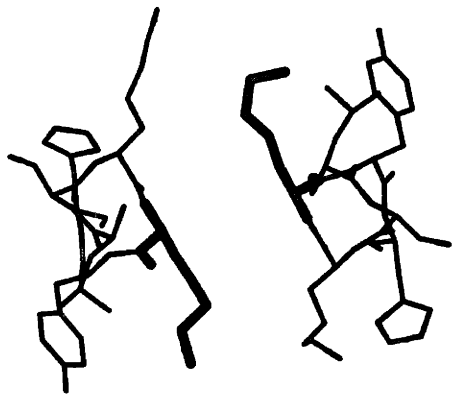

supercoil path
(α helix axis)
Z (supercoil axis)
ω1
ω0
R0
R1
Y
X
α helix path
ERHEQ
KNSD
EYAGK
LEKK
LLL
12
VVVV
16
LEK
R1
ω1
R0
ω0
φ

Figure 4. Crystal structures of coiled coils with different polar or charged residues in the core ***a*** position. (A) Wild type GCN4-p1(*11*), a dimer with the native asparagine. The asparagine residues form a buried hydrogen bond. (B) GCN4-pV_KL(*39*), dimer with the ***a*** position asparagine replaced by a lysine. The terminal charged nitrogen is directed towards the coiled-coil surface. (C) superposition of GCN4-pV_QL (purple) (*38*) and GCN4-pI_QI (grey)(*42*). Both are trimers with a single ***a*** position glutamine. GCN4-pI_QI has isoleucines at the ***a*** and ***d*** positions with the exception of the single ***a*** position glutamine (residue 16). The glutamines adopt different rotamers and the GCN4-pV_QL structure includes two coordinated non-symmetric water molecules (purple spheres), while GCN4-pI_QI has a single coordinated chloride ion (grey mesh) on the three-fold symmetry axis. (D) Coordination of a buried polar–charge interaction in the tetramericSNAP-SNARE coiled coil(*32*)between a single arginine and three glutamine residues.

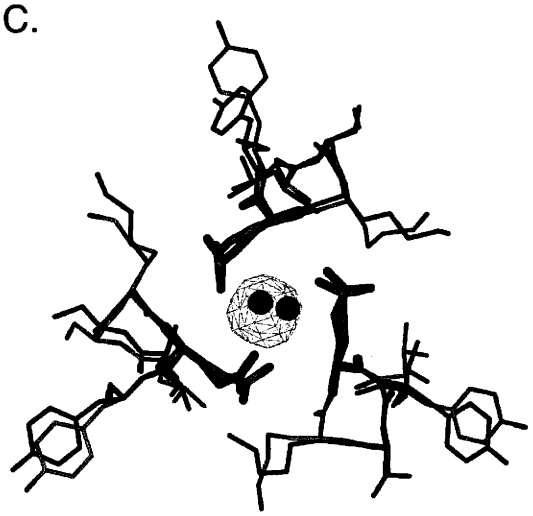
A.



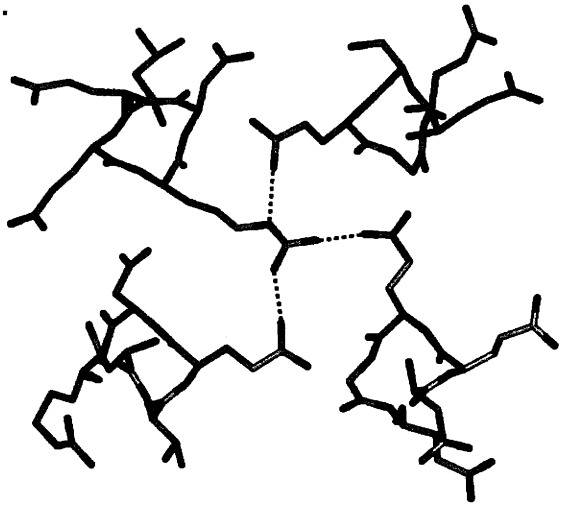
B.



C.



D.



Chapter 2

Buried Polar Residues in Coiled-Coil Interfaces

Chapter 2 has been published as:

David L. Akey, Vladimir N. Malashkevich and Peter S. Kim,

"Buried Polar Residues in Coiled-Coil Interfaces" Biochemistry (in Press)

[1]Abbreviations: MALDI-TOF, matrix-assisted laser desorption ionization time-of-flight; CD, circular dichroism; CSP and LCSP, Crick supercoil parameters and local Crick supercoil parameters, respectively; T_m, midpoint of thermal unfolding transition; RMSD, root-mean-square deviation. [†]Coordinates have been deposited to the RCSB Protein Data Bank, www.rcsb.org: pVL_S, 1IJ0; pVL_T, 1IJ1; pV_TL, 1IJ2; pV_SL, 1IJ3.

ABSTRACT: Coiled coils, estimated to constitute 3–5% of the encoded residues in most genomes, are characterized by a heptad repeat, $(\boldsymbol{abcdefg})_n$, where the buried ***a*** and ***d*** positions form the interface between multiple α-helices. Although generally hydrophobic, a substantial fraction (~20%) of these ***a***- and ***d***-position residues are polar or charged. We constructed variants of the well-characterized coiled coil GCN4-p1 with a single polar residue (Asn, Gln, Ser, or Thr) at either an ***a*** or a ***d*** position. The stability and oligomeric specificity of each variant were measured, and crystal structures of coiled-coil trimers with threonine or serine at either an ***a*** or a ***d*** position were determined. The structures show how single polar residues in the interface affect not only local packing, but also overall coiled-coil geometry as seen by changes in the Crick supercoil parameters and core cavity volumes.

Coiled coils are an oligomerization motif commonly occurring at the interface between separate protein chains. Coiled coils are found in many cytoskeletal and contractile systems (e.g., intermediate filaments, nuclear lamins, and myosin), transcription regulators (e.g., Myc and Max, Fos and Jun, GCN4), viral envelope proteins (e.g., MoMLV, HIV, SIV, influenza), and other systems (*1*). Genome database searches with coiled-coil prediction programs suggest that 3–5% of all protein residues exist as coiled coils (*2*). This implies a role for the coiled-coil motif in a wide range of biological functions. A more complete understanding of the interactions significant to coiled-coil formation will be useful in analyzing protein–protein interactions, predicting candidate partners, and assigning potential biological functions to novel proteins.

Coiled coils are comprised of two or more α-helices that wrap around each other. The hallmark of coiled coils is a heptad repeat of amino acids, $(\boldsymbol{abcdefg})_n$, with a predominance of hydrophobic residues at the buried first and fourth (***a*** and ***d***) positions and charged residues frequently at the fifth and seventh (***e*** and ***g***) positions (Figure 1). Despite the prevalence of hydrophobic residues, approximately 20% of the ***a*** and ***d*** residues are polar or charged (*2-6*).

Previous work shows that buried polar residues in coiled coils can be important determinants of structural uniqueness (*7-12*), influencing both oligomeric state (e.g., dimers vs trimers) and strand orientation (i.e., parallel vs antiparallel). For example, ***a***-position asparagines and lysines, both more frequent in coiled-coil dimers than trimers (Table 1), have been found to specify dimeric coiled coils (*8, 10*). In addition, it has been shown that buried asparagines can direct strand orientation (*9*).

To address the role of buried polar residues in coiled-coil assembly, we constructed and characterized a series of peptides based on the GCN4-p1 sequence (*13, 14*) (a well-studied model system for coiled coils) with a single polar residue at either the ***a*** or the ***d*** position (Figure 1). We characterized these peptides in terms of their helical content, thermal stability, and oligomeric specificity. To investigate how polar residues are accommodated in the interior of coiled coils, we determined the crystal structures of four coiled-coil trimers containing a serine or threonine at either the ***a*** or the ***d*** position.

EXPERIMENTAL PROCEDURES

Peptide Synthesis and Purification. The sequence of the GCN4-pVL variants is Ac-RMKQLEDKVEE-**L/X**-LSK-**V/X**-YHLENEVARLKKLVGER, where the ***a*** and ***d*** positions are underlined. Position 12(***d***) (boldface) is either a leucine or a polar residue (N, Q, S, or T), and position 16(***a***) (boldface) is either a valine or a polar residue. Peptides were synthesized as described previously (*15*) using FMOC chemistry, desalted over a Sephadex G-25 column in 5% acetic acid, and purified by reverse-phase HPLC using a C18 preparative column. Purity was checked by analytical HPLC with a 0.1%/min water/acetonitrile gradient with 0.1% trifluoroacetic acid, and the mass of the peptides was verified by MALDI-TOF mass spectroscopy (PerSeptive Biosystems, Inc., Framingham, MA).

Circular Dichroism (CD) and Sedimentation Equilibrium. CD measurements were carried out in 150 mM NaCl, 50 mM Na_2HPO_4, pH 7.0, at 4 °C with an AVIV 62A or AVIV 62A/DS CD spectrometer (AVIV Instruments, Lakewood, NJ) using a 10 mm cell. Peptide concentrations were determined from the tyrosine absorption maxima at 276 nm in 6 M guanidine hydrochloride (*16*). Peptide concentrations were 10 μM for CD experiments. Thermal melts were performed using the same buffer conditions and peptide concentrations. Data for the thermal melts were acquired by averaging the CD signal over 30 s with a 90 s equilibration time after each 2 °C step. The T_m, midpoint of thermal unfolding transition, reported is the maximum of the first derivative of the thermal melt data.

Sedimentation equilibrium experiments were performed in the same buffer conditions as for CD measurements using a Beckman XLA-90 analytical ultracentrifuge

(Beckman Coulter, Palo Alto, CA). Sedimentation equilibrium experiments were performed at peptide concentrations of 20, 50, and 100 μM. Data were collected after spinning for 18–24 h at speeds ranging from 18,000 to 35,000 rpm at 4 °C. For all speeds and time points, a second data set was acquired 3 h after the first. The two data sets were compared to ensure that the samples were at equilibrium. A single ideal species model was fit to the data. The partial specific volume was calculated from the residue-weighted average of the amino acid sequence (*17*). Solvent density was calculated from the solvent composition (*17*).

Crystallization and Data Collection. Crystallization conditions were found using a sparse matrix screen (Hampton Research, Laguna Niguel, CA). The appropriate conditions were further optimized to yield crystals large enough for diffraction studies. The final crystallization conditions were as follows: pVL_T , 0.1 M sodium cacodylate, pH 6.0, 0.2 M zinc acetate, 20% PEG 8000; pVL_S, 0.1 M sodium cacodylate, pH 7.0, 0.2 M zinc acetate, 16% PEG 8000; pV_TL, 0.1 M HEPES, pH 7.0, 25 mM cadmium sulfate, 1 M sodium acetate; pV_SL, 0.1 M MES, pH 6.5, 50 mM cesium chloride, 10% PEG 20,000. Three of the crystals were soaked for a brief period in stabilizing solutions containing mother liquor and added cryoprotectant (pVL_S , 40% MPD; pV_TL, 20% MPD; pV_SL, 40% MPD) before freezing. Crystals were flash-frozen and data collected at 100 K using either an in-house source (RU-300 rotating anode with an R-AXIS IV detector), beam-line X4A at the National Synchrotron Light Source (NSLS), Brookhaven, NY, with an R-AXIS IV detector, or beam-line 5.0.2 at the Advance Light Source (ALS), Lawrence National Laboratory, Berkeley, CA, with an ADSC-4 CCD detector (Table 2). Diffraction intensities were integrated using the programs DENZO

and SCALEPACK (*18*) and reduced to structure factors with the program TRUNCATE from the CCP4 program suite (*19*).

Model Building and Refinement. Molecular replacement solutions were found using the program AMORE (*19*) with either the GCN4-p1 dimer (*14*) or the GCN4-pV_QL trimer (*20*) structure as models. For molecular replacement, the model structures were truncated to alanine at position 16(***a***) for all models and also at position 12(***d***) for pVL_S and pVL_T. In all cases, AMORE identified solutions using the GCN4-pV_QL trimer with significantly higher correlation and lower *R*-factors than those found using the GCN4-p1 dimer. This confirmed that asymmetric units of the crystals contained a coiled-coil trimer.

For GCN4-pVL_S, the AMORE solution underwent simulated annealing using the program CNS (*21*) and then modified to a polyserine model, with the exception of the core residue, Ser 12(***d***), which was modified to alanine. The polyserine model was refined and water added using the program REFMAC (*19*). Water molecules were removed and side chains manually added to the model using the program O (*22*) as the densities for the side-chain rotamers became apparent. The model was further refined with multiple rounds of manual improvement using O, followed by positional and *B*-factor refinement using CNS.

After finding the molecular replacement solution using AMORE, the pVL_T, pV_TL, and pV_SL structures were refined by 10 independent rounds of simulated annealing. A $2F_O–F_C$ map was computed in CNS using the structures from all 10 rounds. Density interpretation and manual improvement were done with the program O. Crystallographic refinement was done using the CNS program. The final structures were

checked using simulated annealing omit maps (*23*). Density was not always apparent for residues 31–33 for some chains. When the density for these terminal residues was not observed they were omitted from the model. In addition, the density for some side chains was not observed and the occupancy on these side-chain atoms was reduced to zero.

RESULTS

Solution Studies. The peptide GCN4-pVL was used as the reference peptide for our studies. GCN4-pVL is equivalent to GCN4-p1 with the ***a***-position asparagine at residue 16 [denoted Asn 16(***a***)], mutated to valine. GCN4-pVL has valines at the ***a*** positions (except for a single methionine) and leucines at the ***d*** positions (Figure 1). The single methionine, Met 2(***a***), is present in wild-type GCN4-p1 and was not changed in these studies. As shown previously (*7*), GCN4-pVL is stable [(T_m) = 95 °C at 10 μM as determined by circular dichroism (CD)] and forms a mixture of dimers and trimers as determined by sedimentation equilibrium analysis (Table 1). Eight variants of GCN4-pVL with four different polar residues (Asn, Gln, Ser, or Thr) at either an ***a*** (residue 16) or a ***d*** (residue 12) position were synthesized (Figure 1). The resulting peptides are termed pV_XL for ***a***-position variants and pVL_X for ***d***-position variants, where X is N, Q, S, or T, denoting the polar substitution. Wild-type GCN4-p1, equivalent to pV_NL, is not renamed here.

Sedimentation equilibrium analyses of the peptides reveal only two variants with a unique oligomeric state (Table 1). pVL_T is a trimer, and as noted previously (*13*), wild-type GCN4-p1 is a dimer. These results correlate with the residue distributions observed in coiled-coil databases: asparagines at the ***a*** position are more often found in dimeric coiled coils, while threonines at the ***d*** position are more frequent in trimeric coiled coils. Other buried polar residues do not exhibit clear oligomeric preferences as judged by sedimentation equilibrium, and the slight preferences observed do not correlate with database residue distributions. Polar substitutions in the ***d*** position result in a greater variation in T_m and helical content than substitutions in the ***a*** position (Table 1). While

pVL_N and pVL_Q seem to favor dimer formation, the low helical content and low T_m of these peptides suggest that there may be a significant portion of predominantly unfolded monomer in solution, skewing the data toward lower molecular weight species. Thus, it is likely that the molecular weight of the folded form of these peptides is underestimated by sedimentation equilibrium analysis. In general, our data are in agreement with the oligomeric preferences seen in other model coiled-coil systems (*11, 12*). The data for pV_QL, which is found here and elsewhere (*10*) to exist as a mixture of dimers and trimers, are in conflict with other studies which suggest that ***a***-position glutamines are sufficient to specify trimeric coiled coils (*11*). These differences may illustrate the importance of context when accounting for the effects of buried polar residues on structural uniqueness.

Crystal Structures. The structures of coiled-coil trimers with threonines and serines at either the ***a*** or the ***d*** positions (pV_TL, pV_SL, pVL_T, pVL_S) were determined at a resolution of 1.9 Å or higher (Table 2). Although three of these four peptides exist as mixed oligomeric species in solution, all four peptides crystallized as a trimer. All four structures show an approximate, noncrystallographic, 3-fold symmetry. As noted, pVL_T has unique trimeric specificity and is the most stable ***d***-position polar variant (Table 1). The structure of pVL_T (Figures 2A and 3) shows symmetric packing of the Thr 12(***d***) residues with an ordered hydrogen-bond network in the core. The hydroxyl group of each Thr 12(***d***) makes an identical i to $i - 4$ backbone hydrogen bond to the peptide carbonyl of Lys 8(***g***), directing the threonine methyl groups into the core (Figure 3). Four structured water molecules, between the Val 9(***a***) and Thr 12(***d***) layers, form the vertexes of a trigonal pyramid. The three outer water molecules form potential hydrogen bonds to

the Val 9(***a***) backbone carbonyl. As there are more potential hydrogen bonds than hydrogen atoms in this network (12 vs 11), partial water molecule occupancies and transient hydrogen bonds are expected.

Similar to the ***d***-position threonine residues, ***d***-position serines in pVL_S are arranged symmetrically. In contrast to the ***d***-position threonines, the hydroxyl groups of the core serines are directed toward the center to form an internal hydrogen-bond network (Figures 2B and 3). No structured water molecules are observed in the core structure. In the vicinity of the serines there is a local decrease in the superhelical radius of about 0.5 Å (Figure 4A), changes in the supercoil pitch (Figure 4B), and an asymmetric rotamer distribution at Val 9(***a***) (Figure 3).

In contrast to the ***d***-position polar residues, the ***a***-position serines and threonines are not packed with 3-fold symmetry and are in different rotamers. The serines in pV_SL have the most structural variability, with one hydroxyl group directed toward the exterior, another making an i to $i - 3$ hydrogen bond to the peptide carbonyl of Leu 13(***e***), and the third with the partial occupancy of two alternate rotamers, either directed to the core or pointing outward toward the bulk solvent (Figure 2D). Two coordinated water molecules are observed in the core. The hydroxyl groups of threonine in pV_TL make intrahelical peptide carbonyl contacts, but with some variation: all three make an $i - 3$ intrahelical contact, but one also makes the alternative $i - 4$ contact observed for ***d***-position threonines (Figure 2C).

Supercoil Parameters. The geometry of a regular coiled coil is defined by a simple set of parameters [denoted Crick supercoil parameters (CSP)] that include the radius (R_0 and R_1) and pitch (ω_0 and ω_1) for both the supercoil and α-helix, respectively,

and the ***a***-position C_α phase angle (ϕ) (Figure 1) (*24, 25*). To examine how single-residue substitutions affect the coiled-coil geometry, we determined the CSP values of each coiled coil (Table 2) using the central three heptads (residues 5–25) (*24, 25*). We fit only the central three heptads because the regions outside these residues are less structured and fit poorly to ideal supercoil parameters (Figure 4D). For each set of superhelical parameters, we determined the C_α root-mean-square deviation (RMSD) between the actual structure and a model coiled coil built using the CSP values. Three of the structures fit closely to the model coiled coil, with RMSDs less than 0.29 Å. pVL_S, however, deviates more substantially, with an RMSD of 0.41 Å (Table 2).

To investigate local effects of the buried polar residues, we utilized a heptad-specific, local Crick supercoil parameter (LCSP), determined by fitting only seven residues at a time. This local fitting of coiled-coil parameters—sliding the window in one-residue increments—allows detection of changes in structure not readily apparent by visual inspection or global fitting (Figure 4). Our method extends previously described local fitting methods (*26*) by simultaneously determining multiple parameters for the coiled coil. Polar residues in the ***a*** position seem to have a smaller effect on the LCSP than ***d***-position polar residues, as judged by the observation that LCSP values for the ***a***-position variants follow a common trend, while ***d***-position LCSP values differ more widely (Figure 4). As noted earlier, ***d***-position polar variants also vary more in stability than ***a***-position polar variants. Local structural variations include a decrease in R_0 centered around the buried serine residues in pVL_S, and variations in ϕ for both pVL_S and pVL_T. The observation that ***a***-position polar residues have a smaller effect on supercoil parameters is also supported by comparisons of GCN4-pV_QL (*10*) with the

structures shown here, and between the GCN4-pII (*27*) and GCN4-pI_QI (*28*) trimer structures (data not shown).

Cavity Volumes in the Core. An analysis of coiled-coil trimers reveals sizeable cavities in the cores (Figure 5). The cavities vary in volume, position, and structured water molecules, particularly around the buried polar positions. Generally, cavities are found in trimeric coiled coils between the ***a*** and ***d*** layers, although pV_SL also has a cavity that extends from the position 12(***d***) layer to the position 19(***d***) layer (Figure 5D). There are no cavities observed between the position 12(***d***) layer and the position 16(***a***) layer for either pVL_T or pVL_S (Figure 5A,B).

One might expect ***d***-position serine variants to have larger cavity volumes than the threonine variants. However, due to differences in supercoil radius and pitch (Table 2 and Figure 4), the cavities for pVL_S are in fact smaller than the corresponding cavities in pVL_T (Figure 5A,B). The largest cavity in pVL_T contains structured water molecules. The ***a***-position polar variants vary more in cavity size and position. Although pV_SL has two fewer cavities than pV_TL, the extended cavity in pV_SL accounts for space that contains two cavities in pV_TL (Figure 5C,D). This extended cavity is the largest in all four structures and contains two structured water molecules. The cavities of pV_TL are the smallest and most regular.

Interhelical Surface Contacts. Ionic surface interactions between helices are important in both dimeric and trimeric coiled coils and are determinants of coiled-coil specificity (*29, 30*). Nevertheless, it is commonly observed that not all possible interhelical interactions are formed. The four structures presented here all have identical surface residues, but different surface contact patterns (Table 3). Some seemingly

repulsive charge–charge interactions are allowed because of the presence of a third, mutually attractive, charged group forming a salt bridge triad (e.g., the A-chain Arg 1 to B-chain Lys-3 interaction in pVL_S is mediated by the B-chain Glu-6 carboxylate).

While the variation in surface residue contacts may be a result of crystal packing, all four trimers required different crystallization conditions and crystallized in different space groups (Table 2), suggesting that differences between the four trimers exist in solution. Differences in supercoil radius and pitch may partially account for variations in surface interhelical interactions, as subtle shifts in these parameters may be enough to bring potentially interacting residues (e.g., ***e*** and ***g*** positions) closer together or further apart. These differences illustrate that *ab initio* design and structure prediction of coiled coils may need to account for alternative ionic interactions among the surface residues.

DISCUSSION

There are several possible reasons for the presence of buried polar residues in natural coiled coils. Buried polar residues may help specify the oligomeric state in sequences that would otherwise exist as a mixture of dimers, trimers, or higher order structures (*6, 8, 10, 20*). We find that determining oligomeric specificity may be a significant role of only a few buried polar residues, as only an ***a***-position asparagine and a ***d***-position threonine are found to specify unique oligomeric states. In contrast, other polar residues do not specify a unique oligomeric preference. As hydrophobic core sequences are an important determinant of oligomeric state (*7*), the effects of polar residues on oligomeric specificity may be limited and depend on the identities of the other core residues.

Buried polar residues in coiled coils likely serve important roles in determining structural uniqueness. Interactions between polar atoms may compensate for the unfavorable desolvation energy of buried polar residues. These interactions are commonly observed in protein cores and have been proposed to be an important source of structural specificity (*8, 31*). In the case of coiled coils, the selection of polar–polar pairings could ensure that composite α-helices are aligned correctly in homotypic coiled coils (*8, 14*). In an analogous manner, polar–polar pairings may be a discriminating factor in partner selection in heterotypic coiled coils. A preference for polar–polar pairing has been confirmed for ***a***-position asparagines in a number of genetic selection experiments (*32–34*). In the structures presented here, the core-position threonines and serines satisfy their hydrogen bond requirements either with residues on adjacent strands (often meditated though water molecules), with the peptide backbone, with external

residues and solvent, or combinations of the above. This ability to satisfy hydrogen bond requirements through several means may allow these polar residues flexibility in forming heterotypic coiled coils as they may pair with both polar and hydrophobic residues. The potential dual nature of polar residues suggests an avenue of further research in determining factors involved in heterotypic coiled-coil assembly.

Finally, a comparison of the four closely related peptide structures presented here illustrates the plasticity of coiled coils in response to polar substitutions. These structures show two distinct mechanisms for accommodating polar residues in the coiled-coil core. The first method is the filling of cavities with structured water molecules, and the second involves changes in the supercoil parameters, particularly the supercoil radius. Additionally, single polar residue core substitutions lead to variation in surface contacts. These variations may need to be accounted for in computational methods that have been successful at predicting and designing coiled coils, but that used cores that lacked polar residues (*24, 35*). In addition, these computational methods have assumed constant supercoil parameters along the length of the coiled coil. Our results show that superhelical parameters can be fit using local Crick supercoil parameters (LCSP), even as they vary along the length of the coiled coil. Utilization of such local fitting methods could play an important role in future coiled-coil design efforts.

Acknowledgments

We thank Dr. Amy Keating for useful discussion and intellectual contributions, Chris Liu and other members of the Kim lab for comments on the manuscript, Dr. Craig Ogata and the staff at the X4A beamline at the National Synchrotron Light Source (Brookhaven National Laboratory, Upton, NY) for beamtime allocation and help, and Dr. Thomas Earnest for beamtime allocation and help at the 5.0.2 beamline at the Advanced Light Source (Berkeley, CA). We are grateful to Leslie Gaffney for assistance in preparation of the figures and manuscript, and Michael Burgess and James Pang for peptide synthesis.

References

1. Cohen, C., and Parry, D. (1994) *Science 263*, 488–489.

2. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Sci. 6*, 1179–1189.

3. Parry, D. A. (1982) *Biosci. Rep. 2*, 1017–1024.

4. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 8259–8263.

5. Lupas, A., van Dyke, M., and Stock, J. (1991) *Science 252*, 1162–1164.

6. Woolfson, D. N., and Alber, T. (1995) *Protein Sci. 4*, 1596–1607.

7. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) *Science 262*, 1401–1407.

8. Lumb, K. J., and Kim, P. S. (1995) *Biochemistry 34*, 8642–8648.

9. Oakley, M. G., and Kim, P. S. (1998) *Biochemistry 37*, 12603–12610.

10. Gonzalez , L., Jr., Woolfson, D. N., and Alber, T. (1996) *Nat. Struct. Biol. 3*, 1011–1018.

11. Wagschal, K., Tripet, B., Lavigne, P., Mant, C., and Hodges, R. (1999) *Protein Sci. 8*, 2312–2329.

12. Tripet, B., Wagschal, K., Lavigne, P., Mant, C., and Hodges, R. (2000) *J. Mol. Biol. 300*, 377–402.

13. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1989) *Science 243*, 538–542.

14. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) *Science 254*, 539–544.

15. Lockhart, D. J., and Kim, P. S. (1992) *Science 257*, 947–951.

16. Edelhoch, H. (1967) *Biochemistry 6*, 1948–1954.

17. Laue, T. M., Shah, B. D., Ridgeway, T. M., and Pelletier, S. L. (1992) in *Analytical Ultracentrifugation in Biochemistry and Polymer Science* (Harding, S. E., Rowe, A. J., and Horton, J. C., Eds.) pp 90-125. Royal Society of Chemistry, Cambridge, U.K.

18. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymology 276*, 307–326.

19. Collaborative Computational Project (1994) *Acta Crystallogr. 4*, 760–763.

20. Gonzalez, L., Jr., Brown, R. A., Richardson, D., and Alber, T. (1996) *Nat. Struct. Biol. 3*, 1002–1009.

21. Brunger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905–921.

22. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110–119.

23. Brunger, A. T., Adams, P. D., and Rice, L. M. (1997) *Structure 5*, 325–336.

24. Harbury, P. B., Tidor, B., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 8408–8412.

25. Crick, F. H. C. (1953) *Acta Crystallogr. 6*, 689–697.

26. Seo, J., and Cohen, C. (1993) *Proteins: Struct., Funct., Genet. 15*, 223–234.

27. Harbury, P. B., Kim, P. S., and Alber, T. (1994) *Nature 371*, 80–83.

28. Eckert, D. M., Malashkevich, V. M., and Kim, P. S. (1998) *J. Mol. Biol. 284*, 859–865.

29. O'Shea, E. K., Rutkowski, R., and Kim, P. S. (1992) *Cell 68*, 699–708.

30. Nautiyal, S., Woolfson, D. N., King, D. S., and Alber, T. (1995) *Biochemistry 34*, 11645–11651.

31. Cordes, M. H. J., Davidson, A. R., and Sauer, R. T. (1996) *Curr. Opin. Struct. Biol. 6*, 3–10.

32. Zeng, X., Herndon, A. M., and Hu, J. C. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 3673–3678.

33. Arndt, K., Pelletier, J., Muller, K., Alber, T., Michnick, S., and Pluckthun, A. (2000) *J. Mol. Biol. 295*, 627–639.

34. Pelletier, J., Arndt, K., Pluckthun, A., and Michnick, S. (1999) *Nat. Biotechnol. 17*, 683–690.

35. Harbury, P. B., Plecs, J., Tidor, B., Alber, T., and Kim, P. S. (1998) *Science 282*, 1462–1467.

36. Nicholls, A., Bharadwaj, R., and Honig, B. (1003) *Biophys. J. 64*, A166.

Table 1: Circular Dichroism and Sedimentation Equilibrium Data for GCN4-pVL and Polar Variants

Peptide	Polar Substitution	$-[\theta]_{222}$ (10^3 deg cm^2dmol^{-1})	Tm (°C)	$MW_{OBS}/MW_{MONOMER}$[a] 20μM	50μM	100μM	Database Distributions[b] Dimer(%)	Trimer(%)
pVL	none	30.0	94	2.5	2.6	2.8	–	–
pV_NL*	Asn 16(*a*)	30.9	54	2.0	2.1	2.1	3.9	0.7
pV_TL	Thr 16(*a*)	30.3	68	2.4	2.4	2.6	0.7	5.6
pV_SL	Ser 16(*a*)	28.5	52	2.3	2.3	2.4	2.0	5.1
pV_QL	Gln 16(*a*)	30.7	56	2.3	2.3	2.4	0.9	2.4
pVL_N	Asn 12(*d*)	15.2	22	2.0	2.2	2.3	1.3	3.1
pVL_T	Thr 12(*d*)	30.5	62	3.0	3.1	3.2	3.6	8.2
pVL_S	Ser 12(*d*)	32.0	44	2.6	2.6	2.8	4.1	2.7
pVL_Q	Gln 12(*d*)	25.8	26	2.1	2.1	2.2	1.8	1.6

*pV_NL is equivalent to GCN4-p1

[a]$MW_{obs}/MW_{monomer}$ is the molecular weight determined from sedimentation equilibrium analysis divided by the expected mass of a monomer.

[b]Database distributions are the frequencies of the specified polar residue at a given heptad position in dimeric and trimeric coiled-coil databases, respectively (*2*).

Table 2. Data Collection, Refinement Statistics, and Crick Supercoil Parameters

Crystal	pVL_T Thr 12(*d*)	pVL_S Ser 12(*d*)	pV_TL Thr 16(*a*)	pV_SL Ser 16(*a*)
Space Group	$P2_1$	$P2_1$	$P2_12_12_1$	P1
a,b,c (Å)	25.7, 42.2, 42.7	35.5, 28.6, 44.5	35.8, 46.3, 51.4	25.6, 25.8, 34.4
α,β,γ (°)	90.0, 98.9, 90.0	90.0, 109.6, 90.0	90.0, 90.0, 90.0	86.5, 84.5, 79.4
Data Collection				
X-Ray Source	In-house	In-house	ALS	NSLS
Resolution (Å)	1.86	1.86	1.70	1.80
Total observations	80296	70770	63455	23291
Unique reflections	7316	7255	9768	7378
Completeness (%)	95.3 (94.4)*	94.4 (93.7)*	99.3 (99.2)*	92.6 (89.0)*
R_{merge} [a]	0.046 (0.306)*	0.065 (0.315)*	0.048 (0.283)*	0.059 (0.180)*
Refinement				
R_{cryst} [b]	0.227	0.211	0.243	0.220
R_{free} [c]	0.278	0.258	0.279	0.269
RMSD Bonds (Å)	0.010	0.014	0.014	0.009
RMSD Angles (°)	1.3	1.9	1.6	1.3
Crick Supercoil Parameters (CSP)[d]				
Alpha Carbon Phase (ϕ [°])	16.1	18.2	18.9	19.7
Supercoil Radius (R_0 [Å])	6.25	6.03	6.25	6.29
Supercoil Pitch (residues/turn)	101	88	98	99
Supercoil Deviation (RMSD [Å])[e]	0.292	0.408	0.256	0.265

*Values in parentheses correspond to the highest resolution shell: 1.94–1.86 Å for pVL_T and pVL_S, 1.76–1.70 Å for pV_TL and 1.86–1.80 Å for pV_SL.

[a] $R_{merge} = \Sigma\Sigma_j \left| I_j(hkl) - \langle I(hkl)\rangle \right| / \Sigma\Sigma_j \left| \langle I(hkl)\rangle \right|$, where I_j is the intensity measurement for reflection *(hkl)* and $\langle I\rangle$ is the mean intensity over *j* measurements.

[b] R_{cryst} (R_{free}) $= \Sigma \left| \left| F_{obs}(hkl) \right| - \left| F_{calc}(hkl) \right| \right| / \Sigma \left| F_{obs}(hkl) \right|$, where F_{obs} and F_{calc} are observed and calculated structure factors, respectively. No σ-cutoff was applied.

[c] Ten percent of the reflections were excluded from refinement and used to calculate R_{free}.

[d] The Crick supercoil parameters (CSP) were determined by fitting an ideal coiled coil to the central three heptads of the final structure using a user-defined CHARMM function (see Results) (*24, 25*).

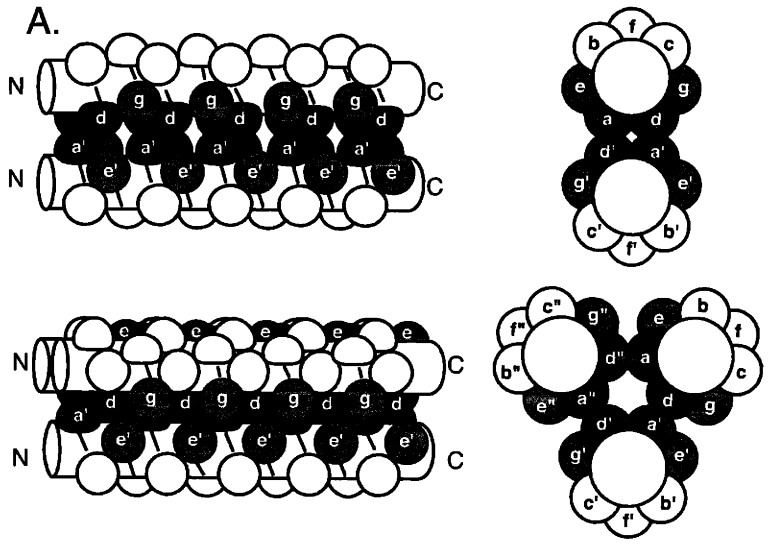
[e] Supercoil RMSD is between the C_α atoms of the model coiled coil and the observed structure.

Table 3: Interhelical Interactions Between Charged and Polar Groups

Inter-helical Contacts					
		Interaction Distances (Å)			
Type	Residues	pVL_T	pVL_S	pV_TL	pV_SL
g - e'	Arg A1($N_{\eta 1}$) – Glu B6 ($O_{\varepsilon 1}$)	2.8	3.6	3.0	
g - e'	Arg B1 ($N_{\eta 2}$) – Glu C6 ($O_{\varepsilon 1}$)	2.7	3.0	3.3	2.7
g - e'	Arg C1 ($N_{\eta 2}$) – Glu A6 ($O_{\varepsilon 2}$)	W		3.1	2.7
g - e'	Lys A15 (N_ζ) – Glu B20 ($O_{\varepsilon 1}$)	2.6			3.7
g - e'	Lys C15 – Glu A20	W	W		W
g - e'	Glu A22 ($O_{\varepsilon 1}$) – Lys B27 (N_ζ)	2.9	2.9	3.6	2.7
g - e'	Glu B22 ($O_{\varepsilon 2}$)– Lys C27 (N_ζ)	2.7	3.0	2.7	3.5
g - e'	Glu C22 ($O_{\varepsilon 1}$) – Lys A27 (N_ζ)	2.7	2.7	2.8	
g - b'	Arg A1 ($N_{\eta 2}$) – Lys B3 (N_ζ)		3.0		
g - b'	Lys A8 (N_ζ) – Glu B10 ($O_{\varepsilon 2}$)	2.9		2.9	
g - b'	Lys B8 (N_ζ) – Glu C10 ($O_{\varepsilon 2}$)	2.5	W		
g - b'	Lys C8 (N_ζ) – Glu A10 ($O_{\varepsilon 2}$)	3.7			
g - a'	Lys A15 (N_ζ) – Ser B16 (O_γ)				3.4
g - a'	Lys C15 (N_ζ) – Thr A16 ($O_{\gamma 1}$)			2.8	
e - g'	Glu A6 ($O_{\varepsilon 1}$) – Lys C8 (N_ζ)		3.8		
c - e'	Gln A4 ($N_{\varepsilon 2}$) – Glu B6 ($O_{\varepsilon 2}$)		3.9	W	
c - e'	Arg A25 ($N_{\eta 2}$) – Lys B27 (N_ζ)	2.8	W		
c - e'	Arg B25 ($N_{\eta 2}$) – Lys C27 (N_ζ)				2.6
c - b'	Glu A11 ($O_{\varepsilon 2}$) – Tyr B17 (O_η)		2.8		
Intra-helical contacts					
b - e	Lys B3 (N_ζ) – Glu B6 ($O_{\varepsilon 2}$)		3.0		
g - c	Glu A22 ($O_{\varepsilon 2}$) – Arg A25 ($N_{\eta 2}$)	2.9	W		

Interactions listed as ***g*** to ***e'*** indicate an interaction between the ***g***-position residue of one chain and the ***e***-position residue of an adjacent chain. Interacting residues are identified by amino acid type, chain designation (A, B, and C for the three chains) and residue number. Interaction distances are shown. A "W" indicates that a potential salt bridge interaction is mediated through a structured water molecule, but that the residues are too far apart to directly interact. The intrahelical contacts listed mediate an otherwise repulsive interhelical contact

Figure 1. (A) Schematic view of dimeric (top) and trimeric (bottom) coiled coils. The ***a*** and ***d*** core positions are indicated in red and the ***e*** and ***g*** positions in blue. (B) Coiled-coil schematic of GCN4-pVL peptide as a trimer. Residues 16(***a***) and 12(***d***), the positions of the polar substitutions, are boxed. Interhelical ***g*** to ***e'*** salt bridges, indicated with dashed lines, are from the GCN4-pI_QI structure(*28*). The Crick supercoil parameters (CSP) are indicated as: R_0 and R_1, the radius for the supercoil and α-helix, respectively; ω_0 and ω_1, the supercoil pitch and α-helix pitch, respectively; ϕ, the C_α phase angle, defined as the angle between vectors from the α-helix center to the supercoil center and to the C_α of the ***a***-position residue.


A.
N
C
N
C
N
C
N
C

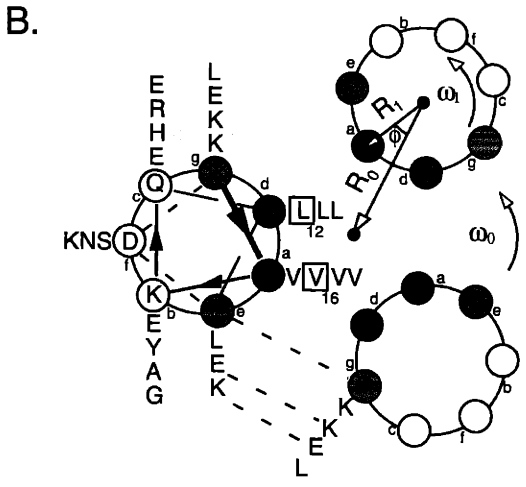
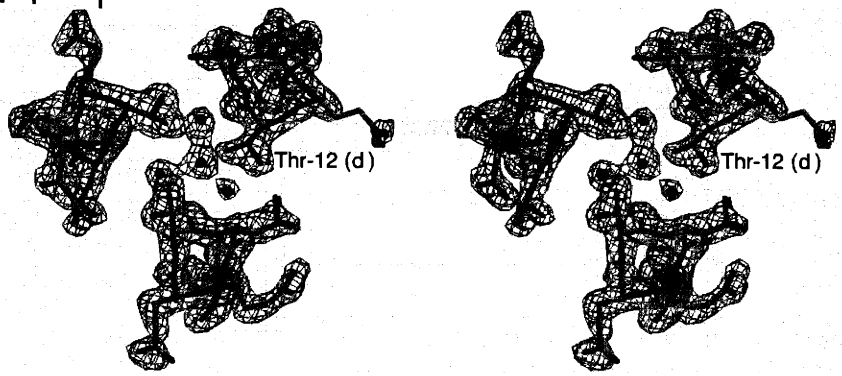
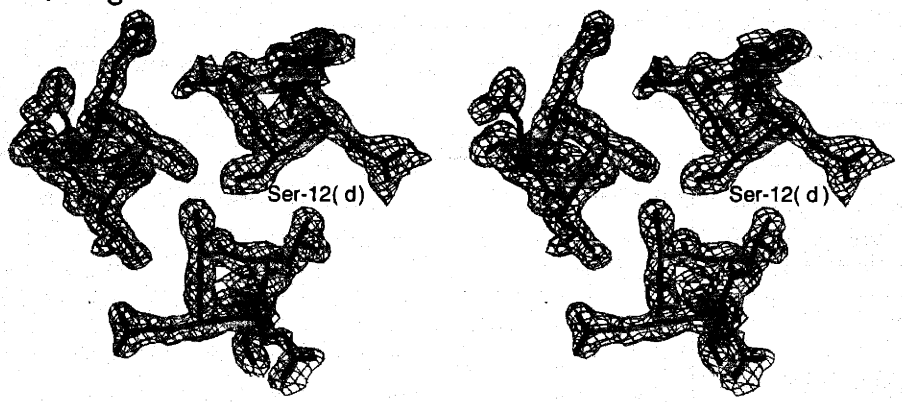

B.
ERHEQ
KNSD
KEYAG
LEKK
LEK
LLL
12
VVVV
16
R_1
ω_1
R_0
ω_0
KKEL

Figure 2. Stereodiagrams of selected area composite omit electron density maps superimposed onto the final refined structures. Maps are contoured at 1σ. Water molecules are represented as red spheres. (A) pVL_T; (B) pVL_S; (C) pV_TL, with alternate rotamers for the labeled Thr 16(***a***); and (D) pV_SL, with alternate rotamers for the labeled Ser 16(***a***).

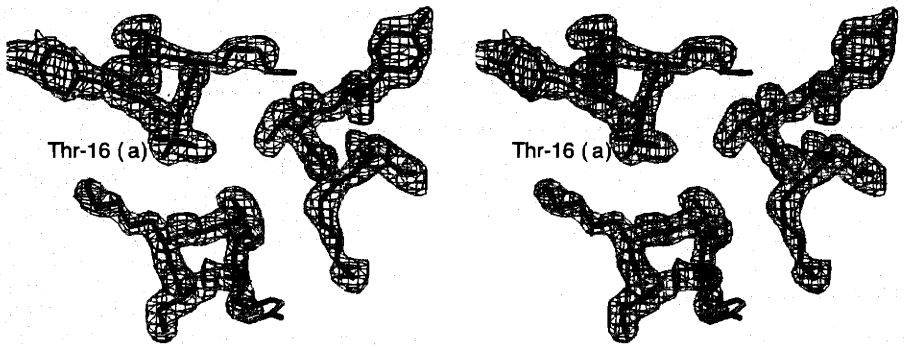
A. pVL_T



B. pVL_S



C. pV_TL



D. pV_SL

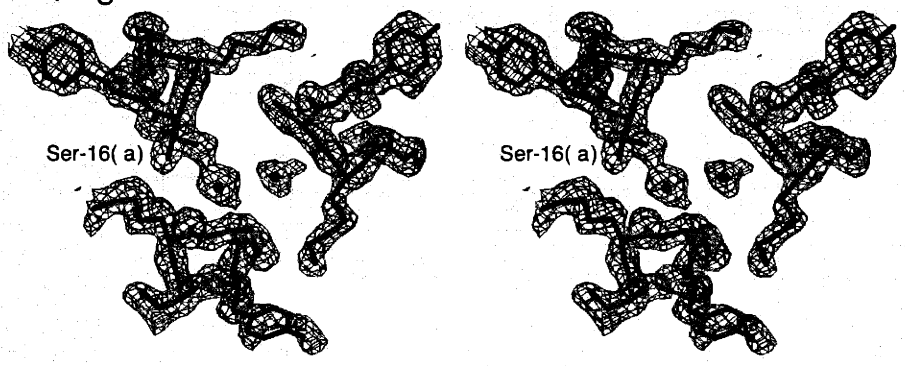


Figure 3. Stereodiagram of pVL_S (residues 8–12, purple backbone) superimposed on pVL_T (gray backbone) including the polar ***d*** position (residue 12). Val 9(***a***) and Thr 12(***d***) of pVL_T are shown as thicker rods. A hydrogen-bond network, shown in red, is observed between water molecules (red spheres), all Thr 12(***d***) residues, and the peptide carbonyl oxygens of residues 8 and 9 in the pVL_T structure. The decreased supercoil radius of pVL_S is apparent. The labeled Val 9(***a***) of pVL_S adopts a unique rotamer not seen for any other ***a***-position valines.

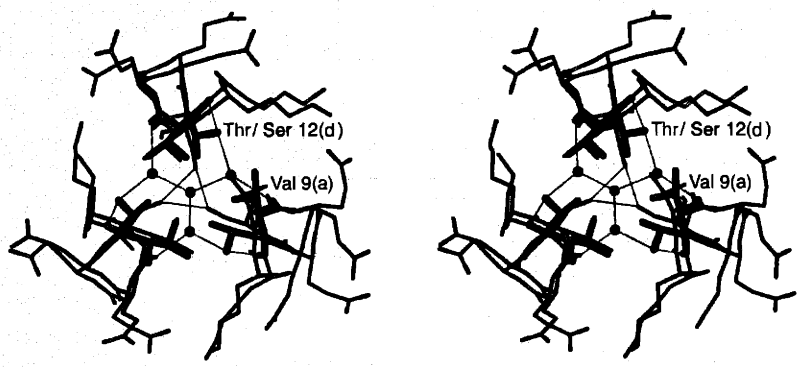

Thr/ Ser 12(d)
Val 9(a)
Thr/ Ser 12(d)
Val 9(a)

Figure 4. Local Crick supercoil parameters (LCSP), as defined in Figure 1, plotted for pVL_T (closed diamonds), pVL_S (closed triangles), pV_TL (open squares), and pV_SL (open circles). The indicated heptads were used to determine LCSP values. (A) Supercoil radius (R_0) in angstroms; (B) supercoil pitch (ω_0) in residues per turn; (C) ϕ in degrees; and (D) deviation from ideal supercoil.

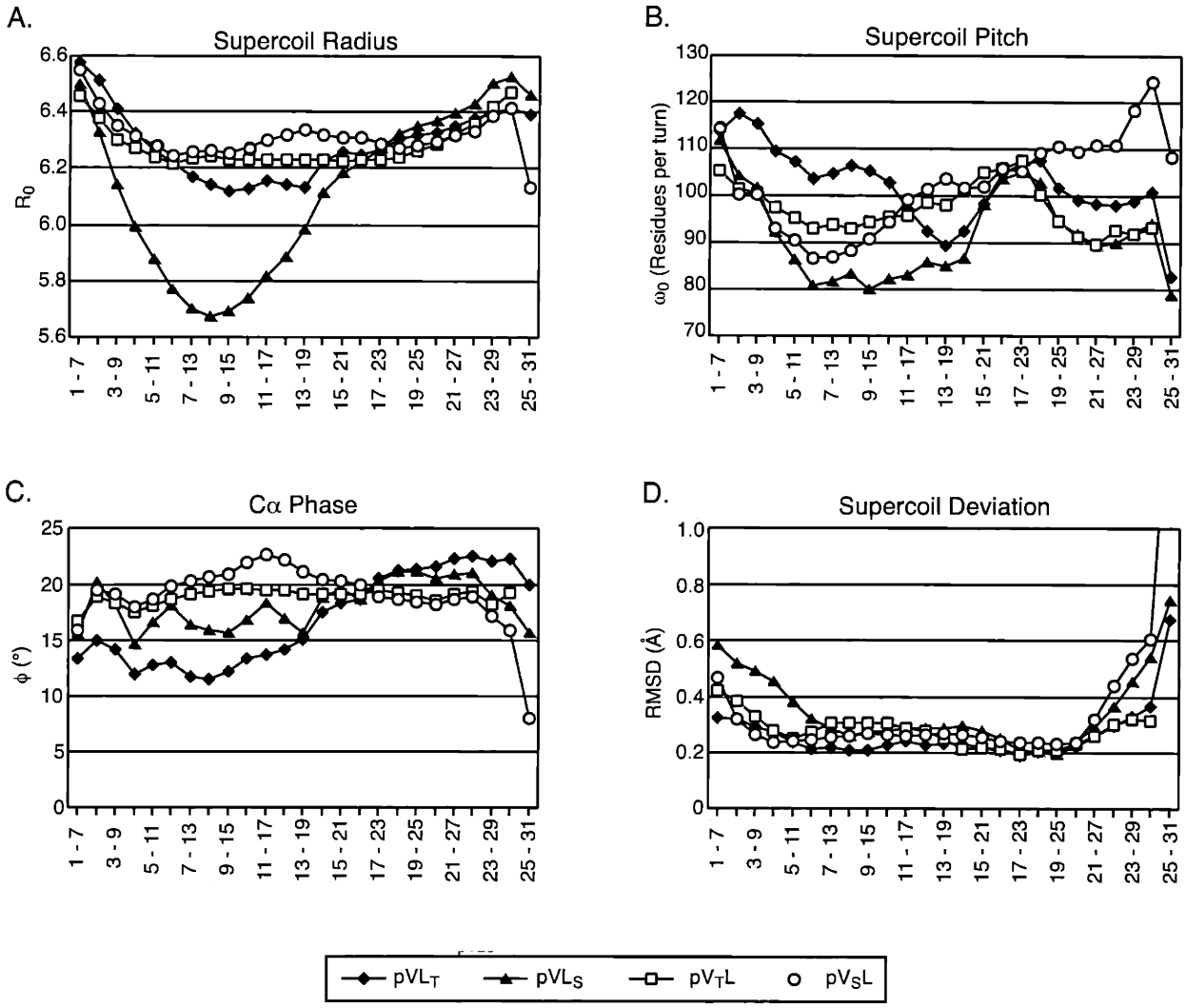
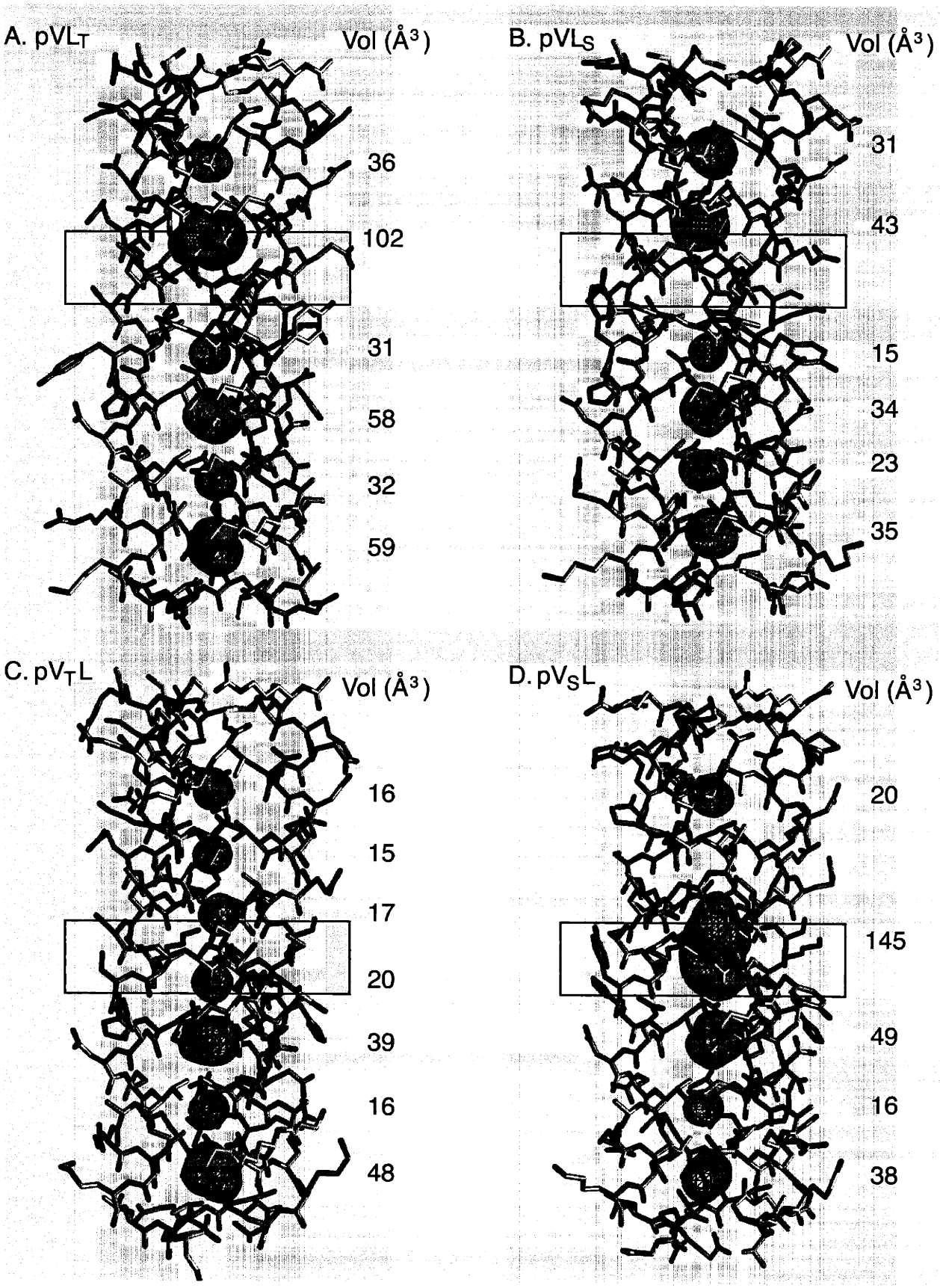

A.
Supercoil Radius
R_0
6.6
6.4
6.2
6.0
5.8
5.6
B.
Supercoil Pitch
ω_0 (Residues per turn)
130
120
110
100
90
80
70
C.
Cα Phase
ϕ (°)
25
20
15
10
5
0
D.
Supercoil Deviation
RMSD (Å)
1.0
0.8
0.6
0.4
0.2
0
1 - 7
3 - 9
5 - 11
7 - 13
9 - 15
11 - 17
13 - 19
15 - 21
17 - 23
19 - 25
21 - 27
23 - 29
25 - 31
pVL_T
pVL_S
pV_TL
pV_SL

Figure 5. Cavities in the cores of coiled-coil trimers: (A) pVL_T, (B) pVL_S, (C) pV_TL, and (D) pV_SL. Vol, cavity volume ($Å^3$). Cavities were from molecular surface calculated with the program GRASP (*36*). Water molecules were omitted for these calculations. The cross-sections containing the core polar residues (see also Figure 2) are indicated with boxes. There are four water molecules (red spheres) in the cavity of pVL_T (A), two in pV_TL (C), and two in pV_SL (D).


A. pVL$_T$
Vol (Å^3)
36
102
31
58
32
59
B. pVL$_S$
Vol (Å^3)
31
43
15
34
23
35
C. pV$_T$L
Vol (Å^3)
16
15
17
20
39
16
48
D. pV$_S$L
Vol (Å^3)
20
145
49
16
38

Chapter 3

An Examination of Predicted Heterotypic Coiled Coils

ABSTRACT: We describe an attempt to establish limited proteolysis as a method for high throughput analysis of coiled-coil interactions. A subset of heterotypic coiled-coil interactions from *Sacharromyces cerevisiae*, predicted by large-scale two-hybrid screen, were used as a test set to evaluate the potential of proteolysis for determination of interacting coiled-coils and identification of the register of interaction. Six coiled-coil domains from four proteins were predicted to form a nexus centered around the KEL1 gene product. The KEL1 protein contains three coiled-coil domains, each predicted to form a single heterotypic interaction in this nexus and accounting for three of the four predicted interactions. CD data confirm only one weak interaction of the four predicted: between the second coiled-coil domain of KEL1 and the coiled-coil domain of PAC1. However, proteolysis conditions that appear to confirm the one weak interaction are prone to error as they also appear to confirm interactions that are clearly seen to not exist by circular dichroism. If proteolysis is developed as a useful probe for heterotypic coiled-coil interactions, a different subset of coiled-coils may be necessary, as the subset used does not appear to have any readily observed strong interactions.

One of the challenges in the post-genomic era is to assign putative function to new protein sequences. This challenge is the primary goal in the developing field of proteomics. A common tactic in assigning function to novel proteins is the identification of interacting proteins—the "guilt by association" approach. An interaction between a novel protein and a protein of known function would implicate the novel protein in the pathway or function of the known protein.

The difficulty in discovering new protein–protein interaction lies with the ever-increasing number of potential interactions to be tested. There are a number of possible approaches, both experimental and computational, to predict protein–protein interactions. Identification of all two-protein interactions requires methods that can be scaled with the square of the number of interacting proteins. For experimental work on the scale of most genomes, this requires high-throughput screening techniques. Recently, several groups published large-scale two-hybrid screens of yeast proteins to identify novel interactions (*1, 2*). Traditional two-hybrid screens have several disadvantages, including uncontrolled expression levels and a dimeric DNA binding site that tends to mask homotypic interactions and leads to a high rate of false positives and false negatives.

Computational methods can limit the number of interactions tested. Several computational methods were devised to predict protein–protein interactions on a genome scale. These include the "Rossetta Stone" and phylogenetic profile methods, which have the potential to limit the number of experiments by directing screens toward interaction s that are more likely to occur (*3, 4*). Recently, Newman et al. improved on the traditional two-hybrid screen by using a directed screen (*5*). This work focused on coiled-coil domains from the open reading frames from the yeast *Sacharromyces cerevisiae*. This approach has several advantages. By working only with coiled-coil domains, the number of proteins screened is limited. Coiled coils are known to fold autonomously and thus should interact in the absence of other protein domains. By focusing strictly on the oligomerization motif, Newman and colleagues increased the percentage of positive interactions and limited

"shielding" effects, where a different domain may interact with the coiled-coil region and obstruct it from interacting with other coiled coils.

One observation of the directed coiled-coil screen is the existence of a number of nexuses in which several proteins are related by virtue of interactions with one or more common central proteins. One such nexus contains the KEL1 protein as the central component (Figure 1). The KEL1 protein has been implicated in cell mating and establishment of cell polarity (*6*). KEL1 contains three coiled-coil domains, which we designate KEL1-CC1, KEL1-CC2, and KEL1-CC3, respectively, each separated by a short linker region. In the original two-hybrid screen, both the entire coiled-coil region and each individual coiled-coil region were included. Interestingly, the PAC1 (perish in absence of CIN8) and DYN1 coiled-coil domains, which interact separately with separate KEL1 coiled coils, have some common functionality. The deletion of either DYN1 or PAC1 is synthetic lethal with the CIN8 gene (*7*). (PAC1 was originally identified in a screen for genes that were synthetic lethal with CIN8.) In addition, both PAC1 and DYN1 exhibit defects in alignment of the spindle pole body during mitosis. The observation that they both interact with KEL1 through separate coiled-coil domains leads to the theory that they are involved in a ternary complex with KEL1 during mitosis, and potentially indicates a new function for the KEL1 protein.

We attempt to both confirm a limited number of the results of the two-hybrid assay, and establish proteolysis as a general method for the identification of heterotypic coiled-coil interactions. Six predicted coiled-coil regions predicted in the KEL1 nexus were either expressed recombinantly or synthesized and purified. These six coiled coils were predicted by the two-hybrid assay to form four heterotypic interactions: KEL1-CC1 – DYN1, KEL1-CC2 – PAC1, KEL1-CC3 – YAL011W, and PAC1 – YAL011W. In addition, KEL1-CC1 was predicted to form a homotypic coiled coil.

The predicted interactions were checked by both circular dichroism (CD) and proteolysis. If two proteins interact through a heterotypic coiled coil, the complex is more stable than the average of the individual homotypic coiled coils. Thus, a mixture of the two proteins should result in an increase in helical content as compared to the sum of the individual spectra. Similarly, there should be a change in the thermal denaturation profile of the mixed proteins, as measured by monitoring CD signal as a function of temperature. To evaluate the potential of proteolysis as a general method of identifying heterotypic coiled coils and to look for potential unidentified interactions among our set of coiled coils, we performed proteolysis assays on the all 15 heterotypic coiled-coil pairs. As the complex in a heterotypic pair should be more stable than the individual peptides, there should be protection against proteolysis in the complex. This effect can be particularly dramatic for peptides which are incapable of forming homotypic coiled coils and thus would be completely degraded in the absence of partner strands.

MATERIALS AND METHODS

Protein Expression and Purification. Reading frames for expression of the coiled-coil regions KEL1-CC1, KEL1-CC2, DYN1, PAC1, and YAL011W were amplified by PCR from a genomic DNA (S288C). The resulting linear constructs were ligated into the pAED4 expression vector (*8*). The cloning was verified by DNA sequencing of the expression plasmids. The sequences for the PCR primers are as follows:

KEL1-CC1: 5'-GGAATTCCATATGACCGAAGTCGTAGATAGGGCA

5'-CGGGATCCTCATTAGGGTTCACCGACAGTTTTCT

KEL1-CC2: 5'-GGAATTCCATATGCAGAAAACTGTCGGTGAACCCT

5'- CGGAATTCTCATTAGACGCCATTTTTGAACTTCA

PAC1: 5'-GGAATTCCATATGCTACCGAAGAAGTGGAACT

5'-CGGAATTCTCATTATTTCAAAACATTATGTGCGCTG

DYN1: 5'-GGAATTCCATATGGCTGAAGAAATGACACAGGATTT

5'-CGGAATTCTCATTATGCTTTGTAGTGTTCAGCCACC

YAL011W: 5'-GGAATTCCATATGAAATGGCAAGATGAGCAAGA

5'-CGGAATTCTCATTATCCCTGTTTATTTTTATTTTGTTCC

The final peptide sequence are as follows:

KEL1-CC1: MTEVVDRALFEKLRSELQSLKELTHEKALEAGAHIKELETELWQLKS QKNSGTTKEIDELDSVRLQSKCEILEADNHSLEDKVNELEELVNSKFLDIE NLNEVIQFQNEKIKSLELEPNYKEKLEELQIEHENLSRENERLKNESKQHN EDIINNVANYSSQLGSLISHWKENRANSSFLESSSSLISVSDENGEKTVGEP

KEL1-CC2: MQKTVGEPYGDQSRHHRVVINKLTNRLDDLLERSQELTISKEKLSSE YHALKMEHSSLSQDVLVKENEIKKIQNDYKESISSMDSASKALMVSQREL EKYKSLNKKLIDELDELKFKNGV

KEL1-CC3: MTLKAELFITNQERDDLKSEVLELKKRLLNLENNTKQVNEDADSDW

PAC1: MLPKKWNSIVRLQKKIIELEQNTETLVSQIKDLNTQVSELAQFKPTTSNGT SAHNVLK

DYN1: MAEEMTQDLEASIEVSKRKYSLLIRDVEAIKTEMSNVQANLDRSISLVKSL TFEKERWLNTTKHNENSSMISDPAANKARKEAELAAATAEQ

YAL011W: MKWQDEQELKKKEKELKRKNDAEAKRLRMEERKRQQMQKKIAK EQKLQLQKENKAKQKLEQEALKLKRKEEMKKLKEQNKNKQG

All peptides, except KEL1-CC3, were expressed in *E. coli* BL21 pLysS cells. KEL1-CC3 was toxic to *E. coli* and therefore made by solid-phase peptide synthesis. For all expressed peptides, cells were lysed by a freeze/thaw cycle followed by sonication at 4°C in lysis buffer (50mM Tris [pH 8.7], 15% glycerol, 5 mM EDTA, 1 mM DTT). Protease inhibitors (PMSF 1 mM, leupeptin 1 μg/mL, pepstatin 1 μg/mL) were added. During sonication, DNAse 2 μg/mL and MgCl 8mM were included, followed by a further addition of EDTA to 10 mM final concentration. KEL1-CC2 and DYN1 peptides formed inclusion bodies, which were washed with 1% TX-100 in 20 mM Tris (pH 8.0), 5 mM EDTA, and rinsed twice in 20 mM Tris (pH 8.0), 5 mM EDTA. The inclusion bodies were

resuspended in 6 M Urea, 25 mM Tris (pH 8.0), 1 mM EDTA, 1 mM DTT, and initially purified over a DEAE sepharose column with a 0–0.5 M NaCl gradient. KEL1-CC2 came out in the flow through, and DYN1 in the gradient. The peak fractions were pooled, dialyzed against 5% acetic acid and further purified by reverse-phase high-performance liquid chromatography (HPLC) with a 0.1%/min acetonitrile gradient at 10 mL/min.

The PAC1 and YAL011W peptides were purified by DEAE-sepaharose in 25 mM Tris (pH 8.5), 1 mM EDTA, 1 mM DTT, where they came out in the flow-through, followed by SP-sepharose in the same buffer with a 0–0.5 M NaCl gradient over 200 mL. The peak fractions were pooled, acetic acid added to 5%, and purified by reverse-phase HPLC as above. The KEL1-CC1 peptide was purified over DEAE sepharose with 25 mM Tris (pH 8.0), 1 mM EDTA, 1 mM DTT with a 0–0.5 M NaCl gradient. The peak fractions were pooled, diluted 1:2 in the same loading buffer, and further purified over Q sepharose and reverse-phase HPLC. After solid-phase synthesis, the KEL1-CC3 peptide was purified directly by reverse-phase HPLC. The purity of all peptides was checked by analytical HPLC and the masses verified by MALDI-TOF mass spectrometry. The N-terminal methionine of KEL1-CC1 was cleaved *in vivo* as determined by mass spectrometry.

Circular Dichroism. Circular dichroism (CD) measurements were preformed in 150 mM NaCl, 50 mM Na_2HPO_4 (pH 7.0) at 4 °C with an AVIV 62A or AVIV 62A/DS CD spectrometer. Peptide concentrations were determined from the tyrosine and tryptophan absorption maxima in 6 M guanidine-hydrochloride. Thermal melts were performed using the same buffer conditions and peptide concentrations. Data for the thermal melts were acquired by averaging the CD signal over 30 s with a 90-s equilibration time after each 2°C step

Proteolysis. Proteolysis reactions were done by incubating each peptide singly, and pairwise mixtures of peptides in 100 mM Tris (pH 8.5), 2 mM EDTA. The reactions were run with peptide concentrations at both 5 μM and 50 μM per peptide. The proteases used were trypsin, subtilisin, Proteinase K, and thermolysin at approximmately 1:1000

weight:weight ratio. After a 30-minute incubation at room temperature, the reactions were stopped by the addition of PMSF and EDTA. The unmixed digests were then combined and run in parallel on SDS-PAGE with the mixed digests. Gels were stained by silver staining for the 5 μM peptide concentration reactions, or by Commassie stain for the 50 μM reactions.

RESULTS

Circular Dichroism. Our results indicate that of the six coiled coils, KEL1-CC1 and KEL1-CC3 form stable, α-helical, homotypic complexes, consistent with the notion that they are coiled coils (figure 2). Only KEL1-CC1, however, was predicted by the two-hybrid assay to form a homotypic interaction. KEL1-CC2 and YAL011W had moderate α-helix content, but a low T_m for KEL1-CC2 and a broad, noncooperative melt for YAL011W indicate that at physiological conditions they are not stable, homotypic coiled coils (Figure 3). The homotypic CD data for four of the six coiled-coils indicate that if these peptides are capable of forming coiled coils, they would do so through heterotypic interactions.

CD analysis of the four predicted heterotypic interactions only indicate a weak interaction between KEL1-CC2 and PAC1. At 20°C, the observed combined spectra shows a larger signal at both 208 and 222 nm than the sum of the two individual spectra, indicating that the mixture has a higher helical content than the individual peptides (figure 2). Additionally, the combined melt is different from the sum of the two individual melts. However, the interaction does not seem to be very strong, as the combined melting curve is not very cooperative, and the absolute value of the helical signal at 20°C is not very large (only –16,000 at 208 as compared to –30,000 to –35,000 for fully helical coiled coils). No interaction was seen for the other three predicted heterotypic pairs, KEL1-CC and DYN1, KEL1-CC3 and YAL011W, or PAC1 and YAL011W (figures 2,3).

Proteolysis. To confirm that the interaction observed between KEL1-CC2 and PAC1 is specific, we used a proteolysis assay. Proteolysis has been successful in

identifying interacting regions involving coiled coils (*9, 10*) and should be particularly useful in these experiments as the potentially interacting coiled-coil regions are of different lengths. As there is a significant size difference between KEL1-CC2 and PAC1 (120 residues for KEL1-CC2 and 58 residues for PAC1) the PAC1 peptide can only potentially protect a portion of the KEL1-CC2 peptide from proteolysis. A specific interaction of these peptides should occur with a defined register, thus the smaller PAC1 peptide should protect a defined fragment of the KEL1-CC2 peptide from proteolysis. Conversely, if the interaction is nonspecific, then no defined register should exist, and proteolysis should not result in any unique fragments. A variety of proteases with different site preferences were used in the protease assay (see Materials and Methods). Initially, a range of protease concentrations were tested to identify optimal conditions for identification of unique bands. It was determined that an approximate 1:1000 weight:weight peptide to protease ratio resulted in the identification of unique bands in reactions which included both peptides when the protease mixtures were run on SDS-PAGE (figure 4). (For simplicity, a value of 10 kD was used as the molar peptide weight for protease assays.) MALDI-TOF mass spectrometry and N-terminal sequencing identified the putatively protected bands as the N-terminal 83 and 90 amino acids of KEL1-CC2. Although this limits the size of a potentially interacting region of KEL1-CC2, the register is not uniquely defined, as there is a 25-residue difference in the size of the two peptides.

Although the CD data indicated that the other three predicted interactions were not direct, the potential for utilizing proteolysis as a large-scale screen for coiled-coil interactions provided motivation for scaling up the screen. The 15 mixed reactions and 6 isolated peptides were each probed with 4 proteases for a total of 84 reactions, which were readily accommodated by a 96-well plate format. As with the KEL1-CC2 and PAC1 pair, the reactions were run with peptide concentrations of 50 μM and 5μM each, and the entire screen was repeated to test for reproducible results. Although the results with the KEL1-CC2–PAC1 pair were encouraging, expanding the proteolysis screen to include all 15

possible interactions among the 6 peptides shows that the protease conditions used for this pair results in false positives. In these protease conditions, unique bands were observed for the KEL1-CC1–DYN1 pair, which were clearly seen not to interact by CD (figure 4).

DISCUSSION

The results presented here indicate that the interactions detected by the two-hybrid screen are not present when peptides corresponding to the probed regions are examined in isolation. CD only detects one of four predicted interactions, and that interaction is weak as seen by the low helical content and noncooperative melt of the mixed peptides. Attempts to confirm this interaction by proteolysis, while initially promising, were inconclusive because the protease conditions identified that seemed to confirm the KEL1-CC2–PAC1 interaction, also gave positive results for interactions that were seen not to exist by CD.

This study brings up two issues that need to be resolved. Why were interactions observed in the two-hybrid screen not seen *in vitro*? Why did the proteolysis reactions give false positive signals? There are several possible reasons the *in vitro* results failed to recapitulate the *in vivo* data. In the two-hybrid experiments, protein expression levels are not controlled. As seen in the case of KEL1-CC2 and PAC1, weak interactions can give rise to positive signals. It is possible that in one or more of the observed interactions, the expression of one or the other peptide was high enough to stimulate transcription of the reporter gene, with an even weaker interaction than seen for KEL1-CC2–PAC1.

Alternatively, the interactions observed may exist, but as part of a ternary complex. The peptides tested were all derived from yeast, and tested in a yeast two-hybrid experiment. Thus all necessary components for a ternary interaction are present in the two-hybrid experiments, but without further experiments we cannot know what those components may be. Finally, there may be a requirement for post-translational modification for these interactions to exist. It is known that disassembly of the coiled-coil regions of nuclear lamins is regulated by phosphorylation. Although there are no confirmed instances where

coiled-coil assembly is controlled by post-translation modifications, it is a possibility, and again as we are testing native yeast proteins, any required modification machinery would be present.

The failure of proteolysis to identify interacting coiled coils may simply be due to the lack of a good positive control to establish reaction conditions. At optimal reaction conditions, the protease concentrations used (1:1000 w:w) should be high enough to cut each peptide on the order of 100 times, based on published activity data (*11*). (The protease activities are: Subtilisin, 5-20 U/mg; Trypsin, 1000-13000 U/mg; Proteinase K, 10-300 U/mg; and Thermolysin, 40-100 U/mg. At 1:1000 w:w ratio, each unit represents 0.3 cuts on a given peptide strand over 30 m.) However, to allow the experiments to be run in parallel, the reaction conditions used were chosen as a reasonable compromise of the optimal reaction conditions for each protease. Thus, the actual activity could be low enough that minor differences in substrate quality result in preferential cleavage of one substrate over the other. These differences in substrate quality could be the stability of the homotypic coiled coil, where the more stable peptide would be cleaved only after full cleavage of the second peptide, or due to sequence differences at or near the cleavage sites resulting in similar biases. As the conditions were optimized for a weak positive test case, it is possible that they were too sensitive to small differences in substrate quality. An ideal positive test case would consist of a soluble heterotypic coiled coil with good stability, with strands long enough that protease fragments are still stable.

Although proteolysis may not be useful for detection of unknown heterotypic coiled coils, it should have utility for identification of coiled-coil register. This may be particularly important if a large number of heterotypic coiled coils are between predicted coiled-coil regions of different size. As many of the predicted interactions in the Newman two-hybrid (*5*) screen were between coiled coils of different lengths, there may be a need for a rapid method for determining interaction register.

References

1. Uetz, P., Giot, L., Cagney, G., Mansfield, T., Judson, R., Knight, J., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., Qureshi-Emili, A., Li, Y., Godwin, B., Conover, D., Kalbfleisch, T., Vijayadamodar, G., Yang, M., Johnston, M., Fields, S., and Rothberg, J. (2000) *Nature 403*, p623-7.

2. Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., and Sakaki, Y. (2000) *Proc Natl Acad Sci U S A 97*, p1143-7.

3. Pellegrini, M., Marcotte, E., Thompson, M., Eisenberg, D., and Yeates, T. (1999) *Proc Natl Acad Sci U S A 96*, p4285-8.

4. Marcotte, E., Pellegrini, M., Ng, H., Rice, D., Yeates, T., and Eisenberg, D. (1999) *Science 285*, p751-3.

5. Newman, J. R., Wolf, E., and Kim, P. S. (2000) *Proc Natl Acad Sci U S A 97*, 13203-13208.

6. Philips, J., and Herskowitz, I. (1998) *J Cell Biol 143*, p375-89.

7. Geiser, J., Schott, E., Kingsbury, T., Cole, N., Totis, L., Bhattacharyya, G., He, L., and Hoyt, M. (1997) *Mol Biol Cell 8*, p1035-50.

8. Studier, F., Rosenberg, A., Dunn, J., and Dubendorff, J. (1990) *Methods Enzymol 185*, 60-89.

9. Lu, M., Blacklow, S., and Kim, P. (1995) *Nat Struct Biol 2*, p1075-82.

10. Blacklow, S., Lu, M., and Kim, P. (1995) *Biochemistry 34*, p14955-62.

11. *Proteolytic Enzymes - a practical approach*, IRL Press, Oxford.

Figure 1. Interactions involving the KEL1 protein predicted by the yeast two-hybrid screen. KEL1 has three coiled-coil domains (CC1, CC2, CC3). The arrows indicate the observed interactions, with a loop indicating the protein is predicted to form a homotypic interaction. The shaded regions are included in the current study.

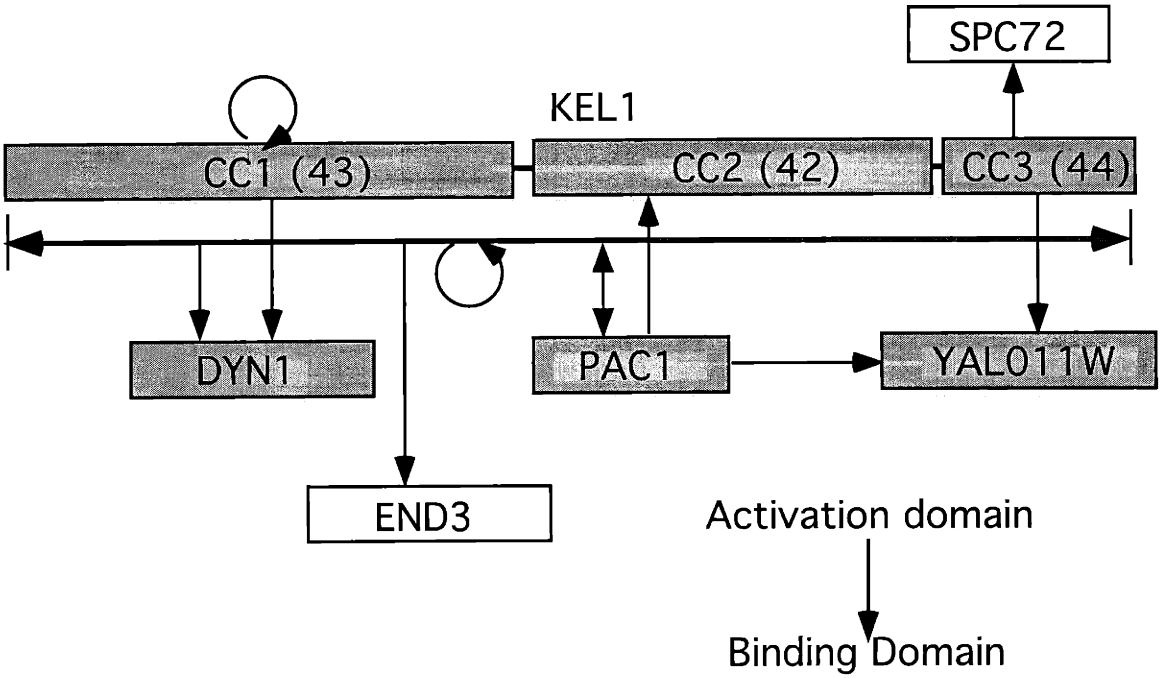
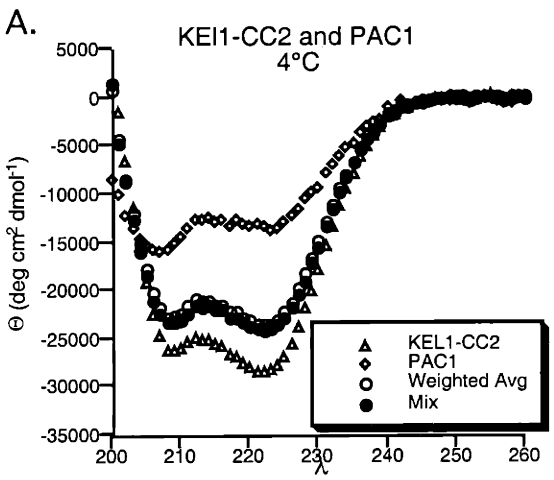
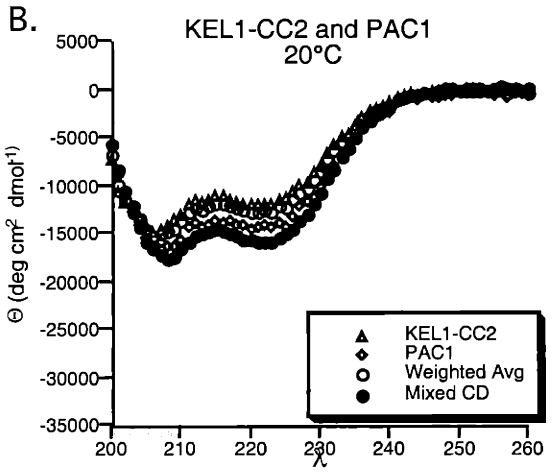
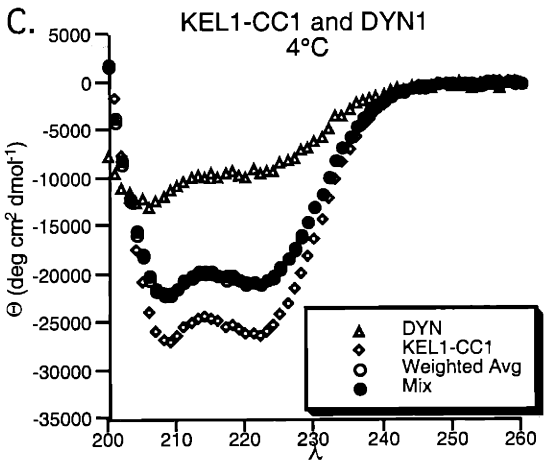
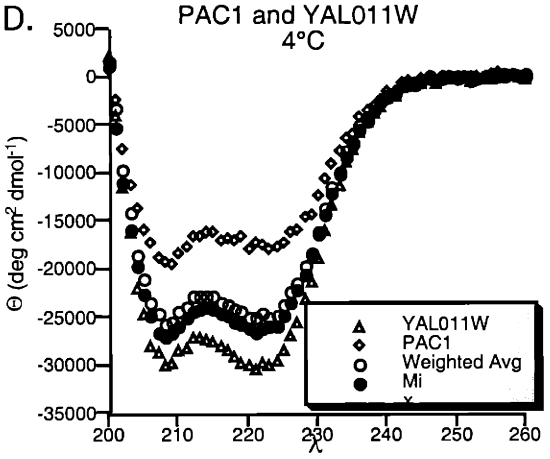

SPC72
KEL1
CC1 (43)
CC2 (42)
CC3 (44)
DYN1
PAC1
YAL011W
END3
Activation domain
Binding Domain

Figure 2. CD spectra of predicted heterotypic coiled coils. Spectra are reported for the individual peptides, the weighted average sum of the individual spectra, and the measured spectra from an equimolar mix of the peptides. (A) KEL1-CC2 and PAC1, 4°C. (B) KEL1-CC2 and PAC1, 20°C. (C) KEL1-CC1 and DYN1, 4°C. (D) YAL011W and PAC1, 4°C. (E) YAL011W and KEL1-CC3, 4°C .


A.
KEI1-CC2 and PAC1
4°C
Θ (deg cm² dmol⁻¹)
λ
KEL1-CC2
PAC1
Weighted Avg
Mix


B.
KEL1-CC2 and PAC1
20°C
Θ (deg cm² dmol⁻¹)
λ
KEL1-CC2
PAC1
Weighted Avg
Mixed CD


C.
KEL1-CC1 and DYN1
4°C
Θ (deg cm² dmol⁻¹)
λ
DYN
KEL1-CC1
Weighted Avg
Mix


D.
PAC1 and YAL011W
4°C
Θ (deg cm² dmol⁻¹)
λ
YAL011W
PAC1
Weighted Avg
Mi

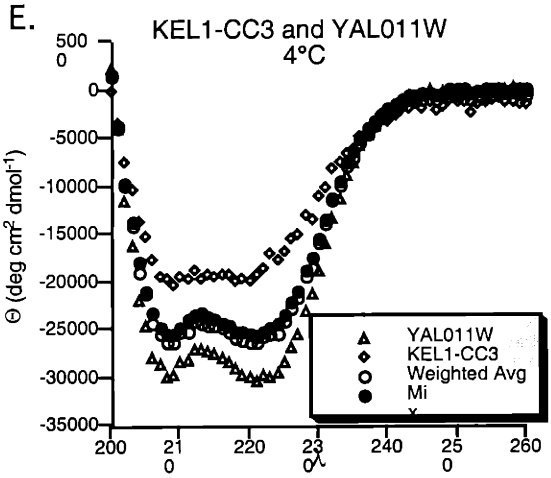

E.
KEL1-CC3 and YAL011W
4°C
Θ (deg cm² dmol⁻¹)
λ
YAL011W
KEL1-CC3
Weighted Avg
Mi

Figure 3. CD thermal melts of predicted heterotypic coiled coils. Spectra reported as for figure 2. (A) KEL1-CC2 and PAC1. (B) KEL1-CC1 and DYN1. (C) YAL011W and PAC1. (D) YAL011W and KEL1-CC3.

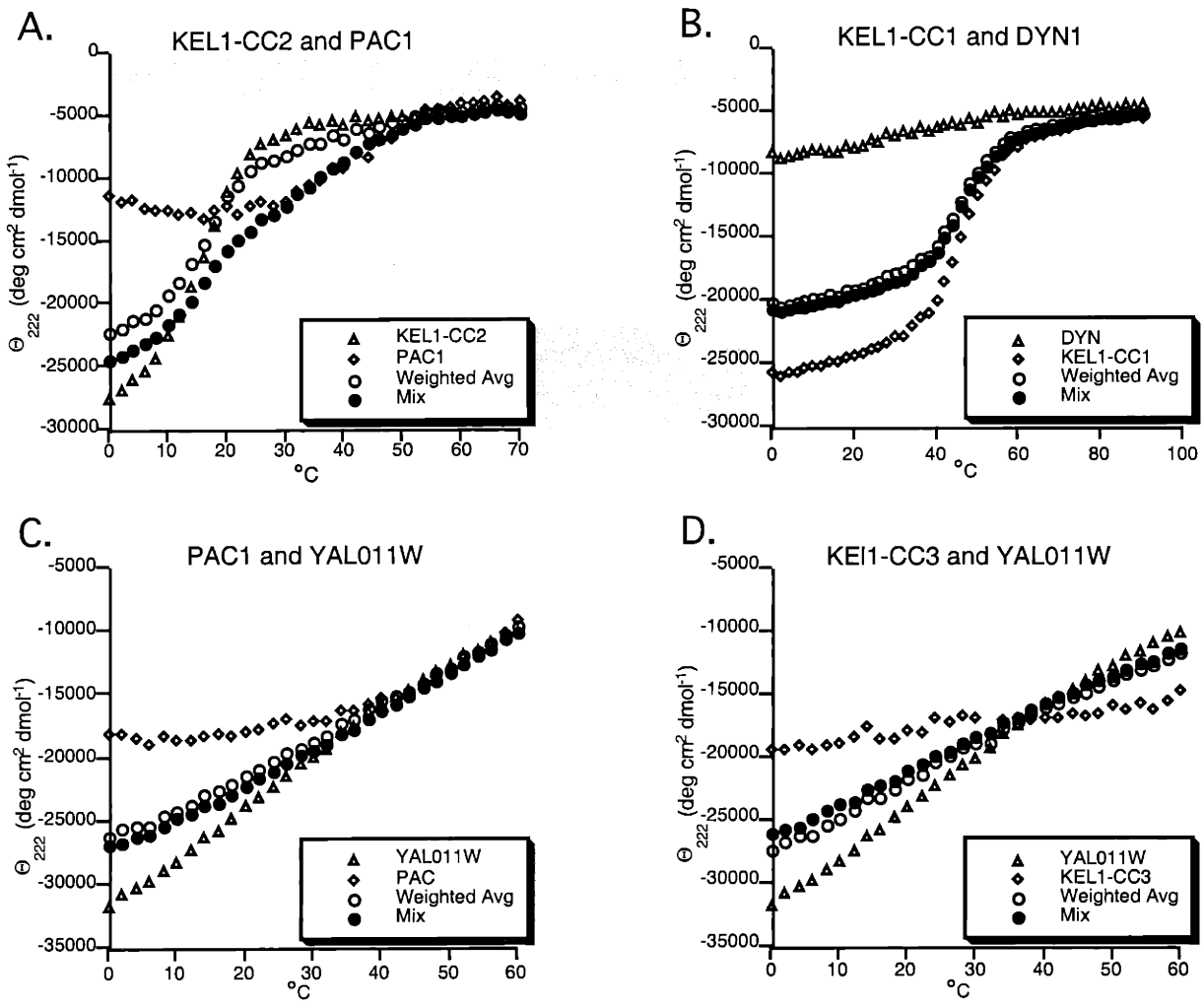
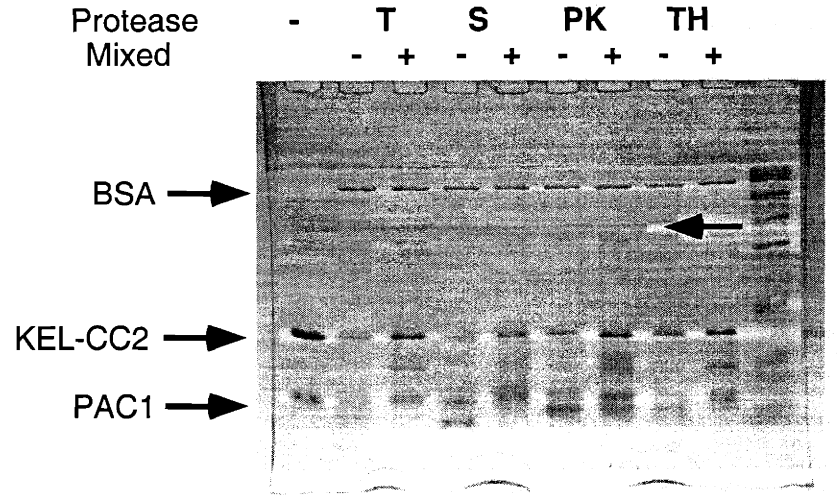

A.
KEL1-CC2 and PAC1
KEL1-CC2
PAC1
Weighted Avg
Mix
B.
KEL1-CC1 and DYN1
DYN
KEL1-CC1
Weighted Avg
Mix
C.
PAC1 and YAL011W
YAL011W
PAC
Weighted Avg
Mix
D.
KEl1-CC3 and YAL011W
YAL011W
KEL1-CC3
Weighted Avg
Mix
°C
(deg cm2 dmol-1)

Figure 4. (A) Protease reaction from a mixture of KEL1-CC2 and PAC1coiled-coil domains. Four proteases were used: T , Trypsin; S, Subtilisin; PK, Proteinase K; TH, Thermolysin. The proteins were digested separately and mixed after the reaction was stopped (digest lanes marked -) or digested as a mixture (digest lanes marked +). BSA was included in the loading buffer as a loading control. The arrow on the right of the gel indicates the position of the band(s) that were only seen when the proteins were digested together. (B) Protease digests for KEL1-CC1 and DYN1.

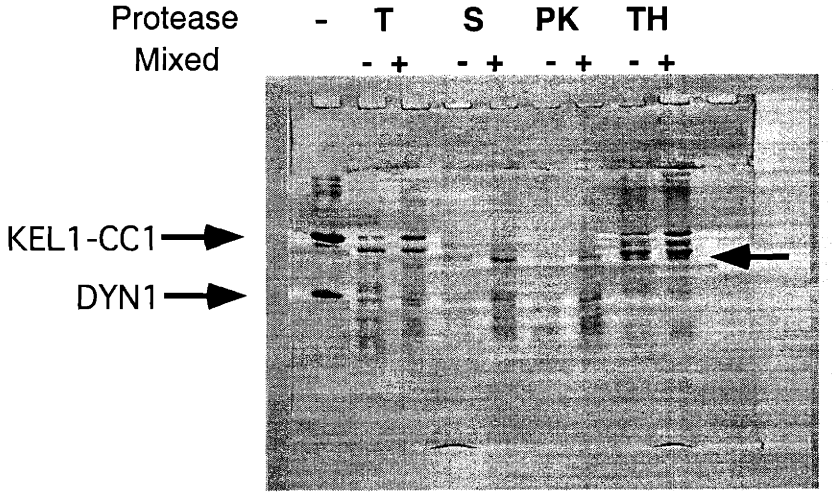
A.

KEL1-CC2 - PAC1 protease digests



B.

KEL1-CC1 - DYN1 protease digests



Chapter 4

Future directions in Coiled-Coil Research

A major goal coiled-coil research is to be able to reliably predict coiled-coil interactions. This problem can be broken down into several distinct parts: prediction of coiled-coil regions in proteins, prediction of oligomeric state of coiled coils, prediction of heterotypic coiled coils, prediction of partners of heterotypic coiled-coil strands, and prediction of strand orientation. The first two goals have been addressed by the coiled-coil prediction programs COILS, PairCoils, and MultiCoil (*1-3*). Accomplishment of the remaining goals will likely require a merging of biophysical experiments, large scale proteomic screens, statistical analysis, and molecular dynamics approaches.

Perhaps the most intriguing of these problems is that of predicting heterotypic coiled coils and their partner strands. We do not currently have any reliable estimates as to what percentage off coiled-coil proteins will form heterotypic structures. The best known examples of heterotypic coiled coils are the keratins, in which the type I keratins pair exclisively with the type II keratins, and the bZIP transcription factors. Unfortunately, these examples do not provide enough sequence data to build statistically significant residue distributions, and thus the approaches taken to create the COILS, PairCoil, and MultiCoil programs cannot easily be applied.

There are several methods to overcome the problem of small data set size for heterotypic interactions. An approach that has recently been developed is to use the interactions in all dimeric coiled coils as a model for heterotypic coiled coils (Mona Singh, unpublished results). This approach takes advantage of the fact that many of the important interactions between strands in coiled coils are not symmetric. Intrastrand interactions can be subdivided into two categories, the symmetric ***a*** - ***a'*** and ***d*** - ***d'*** interactions and the asymmetric ***g*** - ***e'***, ***g*** - ***a'***, ***d*** - ***e'***, ***a*** - ***d'***, and ***d*** - ***a'*** interactions.

Theoretically, the asymmetric interactions should be equivalent in both homotypic and heterotypic coiled coils, and thus a database of these asymmetric interactions generated from both homotypic and heterotypic coiled coils should be useful in predicting heterotypic interactions. One shortcoming to this approach is that the omitted ***a*** - ***a'*** and ***d*** - ***d'*** symmetric interactions are known to be important interaction determinants, especially in regards to the alignment of complementary polar residues at these positions. This problem can be addressed if a proper weighting scheme can be developed that optimizes the statistically derived asymmetric interaction weights, generated from the coiled-coil databases, with weights for symmetric interactions, derived either from biophysical data or from molecular modeling simulations.

One major unknown in biological systems is the extent to which heterotypic interactions are important. Determining this number is likely to be complicated by added variable of oligomeric state. As there is no comprehensive data on the frequency of heterotypic coiled coils, there is also no data on the oligomeric state of these coiled coils. While the best known heterotypic coiled coils, the keratins and bZIP transcription factors, are dimers, the example of the SNAP-SNARE coiled-coil bundle highlights the importance of higher order heterotypic complexes (*4*). While research by Newman, et al., begins to addresses identity of heterotypic coiled coils it cannot address the frequency of heterotypic interactions (*5*). As there are inherent limits as to the utility of predicting homotypic interactions via the two-hybrid system (e.g., the GCN4 homodimer is not found in the two-hybrid system) this method is not useful for determining heterotypic or homotypic interaction frequencies.

With the recent advent of protein chip technology, there is hope for using direct methods for identifying both homotypic and heterotypic coiled coils. While these methods will be useful in the detection of homotypic interactions and of two component heterotypic interactions, there is still the problem of identifying higher order heterotypic interactions. Ironically, one of the weaknesses of the two hybrid screen, that you cannot be certain detected interactions are direct interactions between two proteins, may be a strength in providing leads for the detection of higher order complexes. Two protein complexes, identified by the two hybrid screen, which do not appear to form direct physical interactions, may be useful as probes on protein chips for the identification of other strands in a complex. This approach, however, may be risky in that it assumes the only problem with recapitulating two-hybrid results is the lack of a third partner. As mentioned in the previous chapter, there are other reasons why the two-hybrid assay may not be detecting direct interactions.

Biophysical experiments can shed light on the determinants of heterotypic interactions. There is potential for the use of the acid-base "Peptide Velcro" designed heterodimer systems as a probe for partnering determinants (*6*). In particular, although it has clearly been shown that buried ***a***-position asparagines will preferentially pair with properly aligned asparagines on partner strands, it is unclear if observation will extend to other polar residues and other positions (*7-9*). As the structures presented in this thesis and elsewhere show different methods for accommodating buried polar groups, it is likely that there will be examples of "partner indifference" in certain polar residues, where a given residue may pair equally well with both polar and non-polar residues on the opposite strand.

The combination of molecular dynamics with Crick parameterization has been very successful in coiled-coil design (*10, 11*). Extension of these methods to include heterogeneous cores should prove useful in directing experiments to probe the partner specificity question. However, these methods may need to be modified to allow for variations in supercoil parameters along the length of the coiled coil. Ideally, supercoil parameterization can be used in computational screening of potential heterotypic interactions. Additionally, the data gained from modeling all the possible combinations from local interactions (i.e. one or two heptads) may be useful for deriving weights that can be used to score potential interactions.

While there are many questions to be addressed before we can say that we fully understand coiled-coil interactions, we also have many tools at our disposal to address these questions. A major challenge will be to learn how to integrate the knowledge gained from disparate experimental methods into one comprehensive system that can reliably be used to direct biological research. Additionally the knowledge gained from studying coiled coils has been and will continue to be useful in the analysis of other, more complex, protein–protein interactions.

References

1. Lupas, A., van Dyke, M., and Stock, J. (1991) *Science 252*, 1162-1164.

2. Wolf, E., Kim, P. S., and Berger, B. (1997) *Protein Sci. 6*, 1179-89.

3. Berger, B., Wilson, D. B., Wolf, E., Tonchev, T., Milla, M., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. USA 92*, 8259-63.

4. Sutton, R., Fasshauer, D., Jahn, R., and Brunger, A. (1998) *Nature 395*, p347-53.

5. Newman, J. R., Wolf, E., and Kim, P. S. (2000) *Proc Natl Acad Sci U S A 97*, 13203-13208.

6. O'Shea, E.K., L., K.J., and Kim, P. S. (1993) *Current Biology 3*, 658-667.

7. Arndt, K., Pelletier, J., Muller, K., Alber, T., Michnick, S., and Pluckthun, A. (2000) *J. Mol. Biol. 295*, 627-39.

8. Zeng, X., Herndon, A. M., and Hu, J. C. (1997) *Proc. Natl. Acad. Sci. USA 94*, 3673-8.

9. Oakley, M. G., and Kim, P. S. (1998) *Biochemistry 37*, 12603-12610.

10. Harbury, P. B., Plecs, J., Tidor, B., Alber, T., and Kim, P. S. (1998) *Science 282*, 1462-7.

11. Harbury, P. B., Tidor, B., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. USA 92*, 8408-12.