OPEN ACCESS

University of BRISTOL

Peer reviewed version

Link to published version (if available):
10.1021/acschembio.6b00935

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Submitted to ACS Chemical Biology in consideration for publication as an ARTICLE

N@*a* and N@*d*: Oligomer- and Partner-specification by Asparagine in Coiled-coil Interfaces

Jordan M. Fletcher[1,*], Gail J. Bartlett[1], Aimee L. Boyle[1], Jonathan J. Danon[1], Laura E. Rush[1], Andrei N. Lupas[2] and Derek N. Woolfson[1,3,4]*

[1]School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

[2]Department of Protein Evolution, Max Planck Institute for Developmental Biology, 72076 Tübingen, Germany

[3]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK

[4]BrisSynBio, University of Bristol, Life Science Building, Tyndall Avenue, Bristol BS8 1TQ, UK

*To whom correspondence should be addressed: Jordan.Fletcher@bristol.ac.uk; D.N.Woolfson@bristol.ac.uk,

Key Words: Asparagine, Coiled Coil, Oligomeric State, Protein Design, Protein-protein Interactions

*Running Title:* Asparagine in Designed Coiled-coil Interfaces

ABSTRACT:

The α-helical coiled coil is one of the best-studied protein-protein interaction motifs. As a result, sequence-to-structure relationships are available for the prediction of natural coiled-coil sequences and the *de novo* design of new ones. However, coiled coils adopt a wide range of oligomeric states and topologies, and our understanding of the specification of these and the discrimination between them remains incomplete. Gaps in our knowledge assume more importance as coiled coils are used increasingly to construct biomimetic systems of higher complexity; for this, coiled-coil components need to be robust, orthogonal and transferable between contexts. Here we explore how the polar side chain asparagine (Asn, N) is tolerated within otherwise hydrophobic helix-helix interfaces of coiled coils. The long-held view is that Asn placed at certain sites of the coiled-coil sequence repeat selects one oligomer state over others, which is rationalized by the ability of the side chain to make hydrogen bonds, or interactions with chelated ions within the coiled-coil interior of the favored state. We test this with experiments on *de novo* peptide sequences traditionally considered as directing parallel dimers and trimers; and more widely through bioinformatics analysis of natural coiled-coil sequences and structures. We find that when located centrally, rather than near the termini of such coiled-coil sequences, Asn does exert the anticipated oligomer-specifying influence. However, outside of these bounds, Asn is observed less frequently in the natural sequences, and the synthetic peptides are hyperthermostable and lose oligomer-state specificity. These findings highlight that not all regions of coiled-coil repeat sequences are equivalent, and that care is needed when designing coiled-coil interfaces.

INTRODUCTION

The coiled coil is arguably the best understood of all protein folds.[1, 2] Coiled coils comprise two or more amphipathic α-helices that supercoil around a central hydrophobic core.[3] There are many potential coiled-coil oligomer states and topologies, ranging from simple parallel dimers and trimers, Figure 1A&B, through to multi-helix bundles with mixed parallel and antiparallel arrangements of the helices.[4, 5] Coiled coils can form discrete domains within proteins, or direct protein-protein interactions.[6, 7]
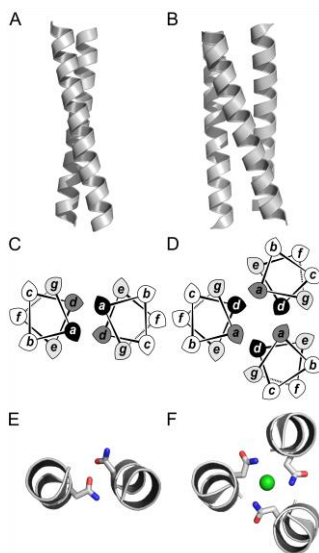
Figure 1. The α-helical coiled coil. A&B: Classical parallel dimeric and trimeric coiled coils formed by peptides CC-pIL-I17N (PDB code 4DZM) and CC-pII-I13N (4DZK), respectively.[8] C&D: Helical-wheel diagrams for coiled-coil dimers and trimers, showing how the heptad sequence repeat relates to the helical and oligomeric structures. E&F: Buried asparagine residues at the *a* and *d* sites of a parallel dimer (4DZM) and a parallel trimer (4DZK), respectively. The central chloride ion coordinated by the Asn residues in the trimer is shown in green.

Despite this diversity of structure, and with it function, coiled coils have common features. First, the adjacent helices pack via so-called "knobs-into-holes" (KIH) interactions [9], which are the signature of all coiled-coil structures.[10] Secondly, KIH packing is related to underlying repeats in coiled-coil sequences, which typically, though not always,[11, 12] exhibit a characteristic heptad repeat (***abcdefg***)$_n$. In these, hydrophobic residues are usually found at the ***a*** and ***d*** positions, and polar residues elsewhere, Figure 1C&B.[1, 2] On average, this spaces hydrophobic side chains 3.5 residues apart. Because the α helix has 3.6 residues per turn, this gives rise to a hydrophobic seam running the length of each constituent helix. In turn, these amphipathic helices provide the primary driving force for helix-helix association. However, because 3.5 and 3.6 are not the same, the helices wrap, or supercoil around each other, Figure 1. Developing this further, the precise geometry of the KIH interactions change with oligomer state. As a result, different side-chain combinations are preferred in the different KIH-packing regimes; put another way, heptad repeats with different sequences direct the assembly of different oligomer states.[2, 13, 14]

This direct and predictive relationship between coiled-coil sequence and structure was first articulated by Harbury and coworkers,[13] who showed that different combinations of isoleucine (Ile, I) and leucine (Leu, L) at the ***a*** and ***d*** sites of the GCN4 leucine zipper lead to

dimers, trimers and tetramers. This provides "rules of thumb" for oligomer-state prediction and design: a = Ile plus d = Leu favors dimer; a = d = Ile, trimer; and a = Leu plus d = Ile, tetramer. These rules have been validated by database studies,[10, 14] and are used as the cornerstone of much coiled-coil design.[2, 15-17]

In addition, charged residues at e and g, which flank the hydrophobic core, Figure 1C&D, often form salt bridges in natural systems.[2, 16, 18, 19] This is also exploited in coiled-coil design: judicious placement of complementary charges—*e.g.*, Lys-Glu pairs at e and g—can be used to pattern sequences to control homotypic or heterotypic assembly of peptides otherwise possessing the same core a & d residues.[19-22] Together, such heuristics have been used to design a suite of coiled-coil components,[8, 23-25] which have been used as modular building blocks in various synthetic-biology applications.[26-30]

The above said, the a/d interfaces are not the exclusive province of hydrophobic residues; indeed, ≈1/4 of residues at these interfaces of structurally defined coiled coils are polar.[2, 4, 10] In many instances, these residues are conserved through evolution,[31] and likely play vital roles in specifying coiled-coil oligomeric state and function.[2] Indeed, engineered and *de novo* designed coiled-coil peptides and proteins with exclusively hydrophobic cores tend to be hyperstable,[8, 13, 24, 25, 32] which, for natural sequences, may cause problems in folding, assembly, dynamics and overall function. Nature's solution here appears to use less-than-perfect repeats, including the incorporation of polar residues, to make even extremely long coiled-coil assemblies dynamic, and with thermally accessible unfolded states.[7, 33]

By far the most abundant of these polar inclusions is asparagine (Asn). For example, in the classic dimeric leucine zipper, GCN4-p1, Asn16 falls at an a position,[18, 34] referred to herein as N@a, Figure 1E. This is highly conserved across classical leucine zippers.[35] Moreover, whilst replacement of this residue with the canonical valine increases thermal stability, the mutant loses oligomer-state specificity.[13] From considerable work, the generally accepted view is that, at a sites in coiled-coil dimers, Asn residues make side chain-side chain hydrogen bonds to partly offset the cost of burying a polar group, but that this is not possible in other oligomers and topologies.[35-38] Studies have also been performed to quantify the contribution of these Asn-Asn pairs to dimer stability at least.[39-41]

To summarize a large body of literature, N@a is destabilising, but it specifies parallel dimer over alternative states such as antiparallel dimer and parallel trimer. Asn inclusions do occur in trimers, though less frequently than in dimers.[4] For instance, inspired by the observation of N@d in trimeric autotransporter adhesins, Hartmann and colleagues describe a number of model peptides bearing N@d residues and show that these are tolerated in the core.[42] However, they are accommodated differently from N@a in parallel dimers: an analysis of N@d residues in natural trimeric coiled-coil domains shows that typically they coordinate anions to form discrete, well-ordered layers in trimeric coiled coils,[42] Figure 1F.

The above underlines a good and clear understanding of Asn insertions in the hydrophobic interfaces of natural coiled-coil systems. Nonetheless, and with increasing use of coiled coils as components, or building blocks, in protein engineering and design in synthetic biology,[29] it

is important to see how this understanding translates to completely *de novo* systems. Here we examine the effects of Asn insertions into the hydrophobic cores of *de novo* parallel dimeric and trimeric coiled-coil peptides,[8] placing particular emphasis on the consequences of the positioning of the residue along the length of the coiled coil. We begin with a description of experiments in which a single Asn residue is "walked" through the *a* or *d* residues in each sequence.[37, 38] These experiments show that the precise positioning of the inclusion with respect to the whole sequence does influence coiled-coil folding, stability and preferred oligomer state, with the adoption of the parallel dimer state being particularly sensitive to this walk. In addition, we use this experimental system to examine how the positioning of Asn insertions along a coiled coil affects orthogonal, homotypic assembly. Finally, we present bioinformatics analyses of the protein structure and sequences databases. Results from these largely support our experimental findings.

EXPERIMENTAL SECTION:

Peptide Synthesis:

Rink amide ChemMatrix[TM] resin was obtained from PCAS Biomatrix Inc. (St-Jean-sur-Richelieu, Canada); Fmoc-L-amino acids and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were obtained from AGTC Bioproducts (Hessle, UK); all other reagents were obtained from Fisher (Loughborough, UK). Peptides were synthesised on a 0.1 mmol scale on Rink amide resin using a Liberty[TM] microwave peptide synthesizer (CEM; Mathews, NC, U.S.A.) employing Fmoc solid-phase techniques and systematically repeated steps of coupling and deprotection interspaced with washings (5 × 7 mL dimethylformamide (DMF)). Coupling: Fmoc-amino acid (5eq.), HBTU (4.5 eq.), diisopropylethylamine (10eq.), in DMF (7mL) for 5 min with 20W microwave irradiation at 75-. Deprotection: 20% piperidine in DMF for 5 min with 20W microwave irradiation at 75˚C. Following linear assembly, the peptide was acetylated (acetic anhydride (3 eq.), DIPEA (4.5 eq.) in DMF (7 mL) for 20 min), and then cleaved from the resin with concomitant removal of side-chain protecting groups by treatment with a cleavage cocktail (10 mL) consisting of trifluoracetic acid (TFA; 95%), triisopropylsilane (2.5%) and H_2O (2.5%) for 3 h at room temperature. Suspended resin was removed by filtration, the peptide precipitated in ice-cold diethyl ether, centrifuged, the pellet dissolved in 1:1 MeCN/H_2O, and freeze-dried. Purification was performed by RP-HPLC using a Kromatek (semi micro, 5 µm, 100 Å, 10 mm ID x 150 mm L) C18 reverse phase column. Eluents used were 0.1% TFA in H_2O (A) and 0.1% TFA in MeCN (B); the peptide was eluted by applying a linear gradient (at 3 mL/min) of 20% to 80 % B over 40 min. Fractions collected were examined by MALDI-TOF mass spectrometry and those found to contain exclusively the desired product were pooled and lyophilized. Analysis of the purified final product by RP-HPLC indicated a purity of >95%. Successful synthesis was confirmed by MALDI-TOF mass spectrometry. (See Supplementary Material)

Sedimentation-equilibrium experiments by Analytical Ultracentrifugation (AUC)

Sedimentation-equilibrium experiments were conducted at 20°C in a Beckman-Optima XL-I analytical ultracentrifuge using an An-60 Ti rotor. Solutions were prepared in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, and 10 mM phosphate buffer) at pH 7.4 with peptide concentrations in the range 75 – 400 µM and spun at speeds in the range 20,000 – 50,000 rpm. Datasets were initially fitted to a single, ideal species model using *Ultrascan.*[43] The partial specific volume for each of the various peptides and the solvent density was calculated using *Sednterp.*[44]

Circular Dichroism

CD spectra were obtained using a JASCO J-810 spectropolarimeter fitted with a Peltier temperature controller. Peptide concentrations were determined by UV absorption at 280 nm (ε(Trp) = 5690 mol[-1] cm[-1]; ε(Tyr) = 1280 mol[-1] cm[-1]).[45] Peptide solutions were prepared in PBS, and examined in 5 mm quartz cuvettes. Thermal denaturation experiments were performed by ramping temperature from 5°C to 90°C at a rate of 10°C/h. Full spectra were

recorded at 5°C intervals, whilst the CD signal at 222 nm was recorded at 1°C intervals (1 nm interval, 1 nm bandwidth, 16 s response time). All peptides in the present study were examined at 10 µM concentration. Raw data (in mdeg) were normalized for peptide concentration, pathlength of the cuvette, and number of amide bonds present to give mean residue ellipticity (MRE; deg cm^2 dmol-res^{-1}). CD mixing experiments to probe the orthogonal assembly of CC-pIL-I10N & CC-pIL-I17N and CC-pII-I13N & CC-pII-I20N pairs were performed by examining the thermal denaturation of 10 µM + 10 µM mixtures of the two peptides, and comparing it to the average MRE when each of the peptides was examined alone.

Disulfide exchange experiments

Analogues of peptides CC-pIL-I10N & CC-pIL-I17N and CC-pII-I13N & CC-pII-I20N bearing an *N*-terminal Ac-Cys^{StBu}-Gly-Gly— tag were prepared: The StBu Cysteine protection used so as to prevent premature disulfide-linked homodimer formation during sample workup and handling. Peptides (50 µM) were mixed in redox buffer (4 mM oxidized glutathione, 1 mM reduced glutathione in PBS, pH 7.8, 1mL) at room temperature. Reaction mixtures were analyzed by HPLC immediately and again after 7 days. Each peak was collected, freeze-dried and its identity established by MALDI-TOF mass spectrometry.

RESULTS AND DISCUSSION

Peptide design, mutagenesis and synthesis

As starting points for our experiments, we used two completely *de novo* peptides 4-heptad repeats in length, previously designed and fully characterized in solution and through to X-ray protein crystal structures.[8] These were CC-pIL-I17N, a parallel homodimer with Asn at the *a* position of the 3rd heptad; and CC-pII-I13N, a parallel homotrimer with Asn at the *d* position of the 2nd heptad, Table 1. *N.b.*, variants of these peptides with completely hydrophobic (Ile/Leu or Ile/Ile cores), CC-pIL and CC-pII, are hyperstable parallel homotrimers.[8]

Initially, six additional sequences were designed, essentially "walking" the Asn residue through each of the remaining *a* or *d* positions of the two sequences, respectively. To aid biophysical characterization, each peptide was equipped with a two-residue *C*-terminal tag to give it a unique mass and a UV chromophore. All peptides were prepared by Fmoc-based solid phase peptide synthesis, followed by reverse-phase HPLC, and confirmed by mass spectrometry (see Supporting Information Figures S1 & S2).

Table 1. Sequences of N@*a* series (top) and N@*d* series (middle), and Asn-swapped variants (bottom) designed for this study. †T_M values were obtained by monitoring CD signal at 222 nm (MRE_{222}) during thermal denaturation experiments (5 to 90°C) at 10 µM peptide concentration in PBS at pH 7.4. ‡ MRE_{222} values recorded at 20°C and 10 µM peptide concentration. *Oligomeric states were determined by analytical ultracentrifugation, and are expressed relative to monomer molecular mass. Due to high thermal stability, the T_M of peptides CC-pIL-I3N and CC-pII-I27N could not be determined. Peptide CC-pII-I10N_I17 was unfolded.

Peptide	Sequence *gabcdef gabcdef gabcdef gabcdef*				T_M (°C)†	MRE_{222}‡ (deg cm². dmol res⁻¹)	Oligomeric State*
CC-pIL-I3N	Ac-G E**N**AALKQ	EIAALKK	EIAALKW	EIAALKQ GYA–NH	> 90	-37,000	3.0
CC-pIL-I10N	Ac-G EIAALKQ	E**N**AALKK	EIAALKW	EIAALKQ GWG–NH	66	-36,000	2.6
CC-pIL-I17N	Ac-G EIAALKQ	EIAALKK	E**N**AALKW	EIAALKQ GYY–NH	70	-37,000	2.1
CC-pIL-I24N	Ac-G EIAALKQ	EIAALKK	EIAALKW	E**N**AALKQ GWW–NH	70	-32,000	2.4
CC-pII-I6N	Ac-G EIAA**N**KQ	EIAAIKK	EIAAIKW	EIAAIKQ GWW–NH	57	-28,000	3.5
CC-pII-I13N	Ac-G EIAAIKQ	EIAA**N**KK	EIAAIKW	EIAAIKQ GYG–NH	49	-29,000	3.1
CC-pII-I20N	Ac-G EIAAIKQ	EIAAIKK	EIAA**N**KW	EIAAIKQ GWA–NH	51	-26,000	2.9
CC-pII-I27N	Ac-G EIAAIKQ	EIAAIKK	EIAAIKW	EIAA**N**KQ GYW–NH	> 90	-27,000	3.2
CC-pIL-L13N	Ac-G EIAALKQ	EIAA**N**KK	EIAALKW	EIAALKQ GYY–NH	69	-35,000	3.7
CC-pII-I17N	Ac-G EIAAIKQ	EIAAIKK	E**N**AAIKW	EIAAIKQ GYG–NH	36	-15,000	2.7
CC-pII-I10N_I17N	Ac-G EIAAIKQ	E**N**AAIKK	E**N**AAIKW	EIAAIKQ GYG–NH	≪ 5	-3,000	NA

Biophysical characterization in solution

The secondary structures of the CC-pIL-N@*a* and CC-pII-N@*d* series were probed by circular dichroism (CD) spectroscopy, Figure 2. CD spectra recorded at 20°C for all eight peptides were similar with minima at 208 and 222 nm, and intensities typical of fully folded α-helical structures, Figures 2A&B. Thermal-denaturation experiments monitoring the loss of mean residue ellipticity at 222 nm (MRE_{222}), however, revealed differences, Figures

2C&D. For the four peptides with centrally located Asn residues – *i.e.*, in the 2nd or 3rd heptads of CC-pIL or CC-pII – the melting curves were sigmoidal, typical of cooperatively folded moieties, and had similar midpoints (T_M): the T_M values for CC-pIL-I10N and CC-pIL-I17N were within 3 – 4°C of each other, and those for CC-pII-I13N and CC-pII-I20N were within 2 – 3°C, Table 1. However, the unfolding behaviors for the four remaining peptides were less consistent, *i.e.* with the Asn residue in the terminal heptad repeats. The differences from the mean T_M values were most extreme for CC-pIL-I3N and CC-pII-I27N – that is, for the peptides in each series with Asn at the ***a*** or ***d*** site closest to one of the termini. These peptides were particularly stable, with only the beginnings of sigmoidal unfolding curves at high temperatures and T_M values > 90˚C. In both cases, the Asn residue lacks an adjacent core hydrophobic residue between it and the *N* or *C* terminus. Therefore, we posit that the high thermal stabilities are due to the destabilizing Asn residues being excluded from the core, effectively leaving more-stable three-and-a-half-heptad coiled coils with completely hydrophobic interfaces. Alternatively, and to account for the high helicities observed by CD spectroscopy, we speculate that the Asn residues could be excluded from the cores by transitions to 11-residue repeats[11, 12, 46, 47] near the termini; *i.e.*, 4-4-3 and 3-4-4 hydrophobic repeats to start CC-pIL-I3N and CC-pII-I27N, respectively, Figure S3. This is only one possibility of course and high-resolution structures will be needed in future to provide a definitive answer. In the absence of such structures, but with the CD data showing near-complete α helicity and hyperstability for these extreme cases, we believe that the Asn side chains must be excluded from the core somehow, and, as a result, that any influence that Asn imparts on oligomer-state specification is lost.
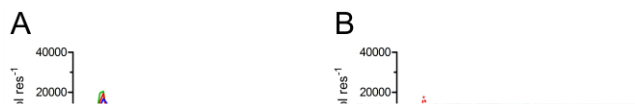
Figure 2: CD spectra and unfolding curves for the CC-pIL and CC-pII series. A&B: CD spectra of CC-pIL-N@***a*** (A) and CC-pII-N@***d*** (B) series. C&D: Thermal denaturation curves for CC-pIL-N@***a*** (C) and CC-pII-N@***d*** (D) series. Key: solid lines, N@***a*** series; broken lines N@***d*** series; with Asn residues in first (green), second (red), third (blue), and fourth (orange) heptads. All experiments were performed at 10 µM peptide concentration in PBS (pH 7.4); for A&B, spectra were recorded at 20°C; and for C&D temperature was ramped from 5°C to 90°C at 10°C/h using a Peltier thermal controller.

Next, we used sedimentation-equilibrium experiments in analytical ultracentrifugation (AUC) to examine the effect of the Asn walk on the oligomeric states of the peptides, Figures S4A-H and Table 1. For the CC-pIL series, it was immediately apparent that the simple presence of an N@***a*** was not sufficient to maintain the dimer state, and that oligomeric state was

extremely sensitive to the precise placement of the polar residue in the sequence. Indeed, only when located in the most central position along the coiled coil – *i.e.*, the parent peptide CC-pIL-I17N – was a dimer unambiguously observed. Shifting the Asn residue to the positions either side of this particular *a* site returned ambiguous molecular weights midway between those expected for dimer and trimer, Table 1. Our best explanation for these data is that the peptides exist in a dimer-trimer equilibrium, for which there are precedents in the literature with mutants of GCN4-p1.[35, 48] For the hyperstable variant, CC-pIL-I3N, the AUC data were again unambiguous, but the oligomer state returned was trimer. This result is consistent with the contention above that when placed adjacent to the terminus the Asn residue is not a *bona fide* "core" residue, and exerts little influence on coiled-coil architecture; and also with foregoing studies of a four-heptad peptide with a completely Ile/Leu core, CC-pIL, which is a hyperstable parallel homotrimer.[8]

By contrast, the oligomeric states determined by AUC for the CC-pII series were more consistent, and mostly unambiguous trimers, Table 1. Although CC-pII-I6N had an apparent molecular weight in solution of 3.5 x monomer, which we cannot explain, the remaining peptides of this series gave values in the range 2.9 – 3.2 x monomer. This is entirely consistent with the literature-wide view that *a* = *d* = Ile strongly directs trimerization irrespective of the inclusion of buried Asn residues.[8, 13]

The trends noted above are best illustrated by the composite plot of CD and AUC data against Asn position, Figure 3. Taken together, these data indicate that for design purposes the influences of N@*a* (to promote parallel dimer formation) and N@*d* (to ameliorate trimer hyperstability) are best exerted in the central heptads of coiled-coil peptide sequences. When placed nearer to the termini the folded states of the peptides become more thermally stable, but have less-well specified oligomer states. We contend that this is because in the latter cases the Asn is likely excluded from the core and has little influence over coiled-coil stability and structure.
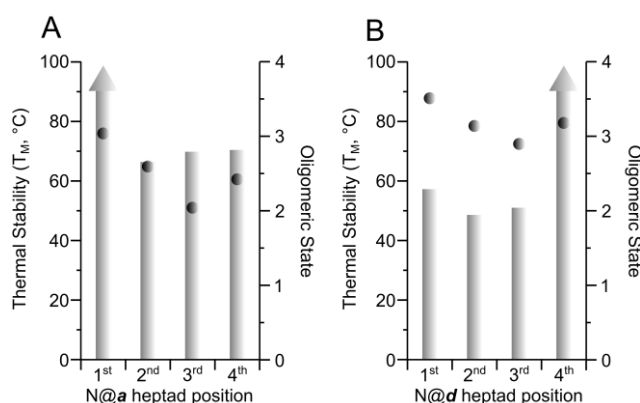
Figure 3: Summary and side-by-side comparison of the biophysical data for the CC-pIL-N@*a* (A) and CC-pII-N@*d* (B) series of peptides. Data are presented as double-y plots, with thermal stabilities as measured by T_M values on y1, and oligomeric state on y2. T_M values are depicted as bars, and oligomeric states as grey spheres. For the former, arrows indicate hyperstable peptides for which T_M could not be determined.

Mixed messages: putting polar residues and hydrophobic combinations in the core in conflict

To probe further the impact of buried, centrally located Asn residue on coiled-coil stability and oligomer state, we prepared two more peptides, CC-pIL-I13N and CC-pII-I17N, Table 1. In the first, an N@*d*, which is tolerated in trimers, was placed in the *a* = Ile plus *d* = Leu background, traditionally associated with dimeric coiled coils. For the second peptide, a dimer-favoring N@*a* was placed in the trimer-specifying *a* = *d* =Ile background. In other words, for these peptides the hydrophobic *a/d* backgrounds and the buried Asn residues were effectively in conflict.

The differences in the helicity and thermal stabilities of these two peptides were marked, Figure S5. CC-pIL-I13N was highly helical and stable, and, in these respects, largely indistinguishable from three of the four other CC-pIL variants, Table 1. However, this variant gave another oligomer state for this series – the AUC data were most consistent with a tetramer, or a trimer-tetramer equilibrium (Figure S4I). Therefore, the emerging picture for this series of peptides with *a* = Ile plus *d* = Leu is that this background is plastic and compatible with multiple oligomeric states. In other words, the energy landscape for sequences based on *a* = Ile plus *d* = Leu is rugged with multiple minima that lie close in energy. This is consistent with literature reports of GCN4-based sequences, which have *a* = Val plus *d* = Leu and show a wide range of oligomer states,[13, 48-50] and with other design studies.[8, 51, 52]

By contrast, CC-pII-I17N was markedly less folded and less thermally stable than any of the other peptides from either series, Table 1: it was only ≈50% helical at 10 µM concentration and 5°C; and the T_M was 36°C, Figure S5B. Nonetheless, AUC data were still most compatible with a trimeric assembly, Figure S4J and Table 1; albeit with lowered apparent oligomer state, presumably because of the reduced T_M. Again, we interpret this as follows: the *a* = *d* = Ile background is highly trimer specifying, and this overrides the dimer-specifying residue N@*a*; however, N@*a* is poorly accommodated within this background, and is therefore destabilizing. Consistent with this idea, an additional peptide in this series with a second N@*a*, CC-pII-I10N_I17N, was fully unfolded at 20°C, Figure S5C and Table 1.

Autonomous and orthogonal folding of peptides with central Asn residues
In addition to their role in influencing oligomeric state and thermal stability, we tested if buried Asn residues could direct orthogonal coiled-coil assembly in our two systems. This was done with two mixing experiments using CC-pIL-I10N plus CC-pIL-I17N, and CC-pII-I13N plus CC-pII-I20N. The reasoning was that homo-oligomerisation should maximize hydrophobic-hydrophobic and Asn-Asn contacts in the core. This was inspired by work from Vinson and colleagues,[39] which shows that Ile-Asn pairings at *a-a'* sites in a heterodimeric coiled coil are destabilizing by 4.3 – 4.9 kcal/mol, compared with favorable coupling energies for Ile-Ile and Asn-Asn of -0.5 and -0.5 kcal/mol, respectively. Thus, if successful, in addition to the use of complementary charged residues at *e* and *g* in specifying multiple, orthogonal, heterodimeric pairs,[23, 24, 53, 54] the addition of Asn residues at *a* and *d* could help expand the repertoire of orthogonal coiled coils for use in protein design and synthetic biology.

We examined mixtures of CC-pIL-I10N plus CC-pIL-I17N, and of CC-pII-I13N plus CC-pII-I20N in two ways. First, thermal unfolding curves following the MRE_{222} were recorded for the individual peptides and for the two mixtures, Figures 4A&B. The former were then averaged to give two theoretical curves for CC-pIL-I10N plus CC-pIL-I17N, and for CC-pII-I13N plus CC-pII-I20N. These curves are what would be expected in the absence of any interaction within each pair; *i.e.*, with homo-oligomerization only. We found that curves for the experimental mixtures and the theoretical curves were indeed coincident, Figures 4A&B. This is consistent with our hypothesis that the peptides of each pair fold independently to obligate homo-oligomers. However, as the individual and composite unfolding curves for each pair were close, to be sure on this point we conducted a further experiment using disulfide exchange.[34]
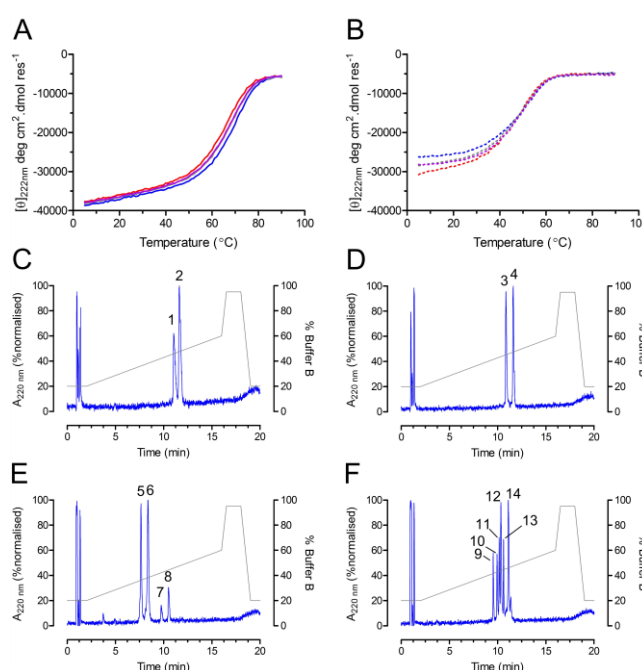
Figure 4: Asparagine-mediated homotypic association. (A&B): Partner specificity followed by CD spectroscopy for the CC-pIL-I10N & CC-pIL-I17N (A) and CC-pII-I13N & CC-pII-I20N (B) systems. Thermal unfolding transitions for the individual peptides (red and blue lines, same key as for Figure 2) and peptide mixtures (purple lines) were followed by MRE_{222}. Theoretical transitions (gray lines) were calculated by averaging the red and blue lines. Any deviation between the purple and gray curves would have indicated heteromeric association. Conditions: all individual peptide concentrations were 10 µM; the buffers were PBS at pH 7.4; and temperatures were ramped from 5 °C to 90 °C at 10°C/h in. (C-F): Partner specificity followed by disulfide-exchange experiments, which were monitored by HPLC and mass spectrometry. C: Mixture of $C^{StBu}GG$-CC-pIL-I10N and $C^{StBu}GG$-CC-pIL-I17N at t = 0 min. D: Mixture of $C^{StBu}GG$-CC-pII-I13N and $C^{StBu}GG$-CC-pII-I20N at t=0min. E: Reaction mixture from Panel C after 7 days of incubation at 20˚C. F: Reaction mixture from Panel D after 7 days of incubation at 20˚C. Conditions: all experiments were performed using 50 µM of each peptide; in PBS (pH 7.8); and with 4 mM of oxidized (GSSG) and 1 mM of reduced glutathione (GSH). HPLC was performed using a C18 reverse phase column (Phenomenex Kinetex, 5 µm particle size, 100 Å pore size, 100 × 4.6 mm) and gradients of Buffer A (0.1 % TFA in H_2O) and Buffer B (0.1 % TFA in MeCN) depicted in gray. Fractions corresponding to each peak were collected, analyzed by MALDI-TOF mass spectrometry and identified as: 1, $C^{StBu}GG$-CC-pIL-I17N; 2, $C^{StBu}GG$-CC-pIL-I10N; 3, $C^{StBu}GG$-CC-pII-I13N; 4, $C^{StBu}GG$-CC-pII-I20N; 5, CGG-CC-pIL-I17N *homodimer*; 6, CGG-CC-pIL-I10N *homodimer*; 7, CGG-CC-pIL-I17N/GS *heterodimer*; 8, CGG-CC-pIL-I10N/GS *heterodimer*; 9, CGG-CC-pII-I13N/GS *heterodimer*; 10, GG-CC-pII-I13N; 11, GG-CC-pII-I20N/GS *heterodimer*; 12, GG-CC-pII-I13N *homodimer*;

13, GG-CC-pII-I20N; and 14, CGG-CC-pII-I20N *homodimer*. Masses and mass spectra are provided in Table S2 and Figures S1e & S1f.

Four additional peptides were prepared, each with a Cys^{StBu}-Gly-Gly *N*-terminal tag to the four sequences described above, Table 1. In separate experiments, the two pairs (C^{StBu}GG-CC-pIL-I10N plus C^{StBu}GG-CC-pIL-I17N, and C^{StBu}GG-CC-pII-I13N plus C^{StBu}GG-CC-pII-I20N) were mixed in redox buffer, and disulfide-exchange was monitored by HPLC at t ≈ 0 min and after one week. Peaks were collected and assigned as monomeric peptide, covalent dimers, or peptide-glutathione adducts by mass spectrometry, (Figures 4C-F, Table S2 and Figures S1e &S1f). These experiments showed clearly that only peptide homodimers were formed, with no evidence for promiscuous heterotypic assemblies. Thus, in this designed peptide system, buried Asn residues appear to confer orthogonality on coiled-coil assembly.

Analysis of natural coiled-coil sequences and structures

To compare our experimental findings with the occurrence of N@*a* and N@*d* in all known coiled-coil structures, we searched the August 2015 release of the CC+ coiled-coil database.[4] CC+ generated from the Protein Data Bank, identifying 367 coiled coils on the basis of knobs-into-holes interactions.[10] These featured Asn as follows: on a per-coiled-coil basis, approximately half of all parallel dimers had N@*a*; approximately half of all the trimers had N@*d*; and Asn occurred less frequently in tetramers, with 15% having N@*a* and 30% N@*d*.

Table 2a. Contribution of asparagine to sequence composition at *a* and *d* sites in a set of non-redundant (≤ 50% sequence identity, > 14 residues in length), parallel dimeric (251), trimeric (89) and tetrameric (27) coiled-coils. Raw numbers were converted to propensities (shown in brackets) by normalizing for the proportion of asparagine in the whole dataset.

Oligomeric State	x=a				x=d			
	All	2	3	4	All	2	3	4
Frequency of N@*x**	139 (1.27)	128 (1.97)	7 (0.23)	4 (0.28)	69 (0.64)	19 (0.30)	42 (1.38)	8 (0.55)
All-Asn layers**	96	91 (6.39)	5 (81)	0	40	8 (0.57)	32 (508)	1

*Calculated on a per-chain basis; *i.e.* each unique chain is counted only once.

***i.e.*, for dimers this means an Asn-Asn pair, for trimers Asn-Asn-Asn trios etc.

Table 2b. Number of asparagine pairs/triplets at *a* and *d* positions within individual coiled-coil sequences, culled from the same dataset as Table 2a. Raw numbers represent the number of coiled-coil sequences containing the number of N@*x* residues stated in the far-left column. Expected numbers of each are given in brackets; these assume random distributions of Asn at the *a* and *d* positions.

# N@*x* residues in each sequence	x=a				x=d			
	Dimer		Trimer		Dimer		Trimer	
	Homo	Hetero	Homo	Hetero	Homo	Hetero	Homo	Hetero
0	105	71	75	10	162	81	58	10
1	51 (17)	9 (1.5)	3 (1)	0	8 (1.7)	0	15 (7.3)	0
2	13 (2)	1 (0.03)	1 (0.1)	0	0	0	3 (0.7)	0
3	1 (0.15)	0	0	0	0	0	1 (0.06)	0
4	0	0	0	0	0	0	2 (0.005)	0
Total Sequences	170	81	79	10	170	81	79	10

Overall, the propensity for N@*a* (1.27) was higher than that for N@*d* (0.64), Table 2a. More specifically, the propensity for N@*a* in dimers (1.97) was higher than in trimers (0.23); with this trend reversed for N@*d*, at 0.30 and 1.38, respectively. Asn occurred less often than

expected by chance in tetramers at both the *a* and *d* positions, with propensities of 0.28 and 0.55, respectively. These findings, for dimers and trimers at least, concur with the preferences for Asn occurring at the two sites determined from the experiments above and those presented in the literature.[8, 13]

Next, we considered all-Asn layers, *i.e.* combinations of Asn found at equivalent *a* or *d* positions of different helices in coiled-coil interfaces and that should be in proximal in the structures, Table 2b. Unsurprisingly, pairs of N@*a* were more prevalent in dimers than trimers, while trios of N@*d* were more prevalent in trimers. Only one asparagine 'quad' was identified at the *d* positions of a homotetramer: namely, the multimerization domain of Nipah virus phosphoprotein.[55] The majority of the pairs and trios were in homomers rather than heteromers, Table 2b. Moreover, the proportion of coiled-coil interfaces with "incomplete" Asn layers—*i.e.*, a single Asn in a dimeric interface, or two Asn at one layer of a trimer—was low at 24% over all hetero-oligomers. For dimers, the majority of these examples had just one Asn pair. Again, these findings accord with the experimental studies presented above and elsewhere: buried Asn residues are destabilizing, and more so if they are not partnered with other Asn or polar side chains.

Given that our experiments indicate that Asn exerts most influence on coiled-coil structure and stability when centrally located, we identified where the examples of N@*a* lie in coiled-coil sequences, by segmenting these sequences into three heptad 'types': the *N-* and the *C-*terminal heptads, and all central heptads, Figure 5. For comparison, we determined the distribution of Asn in all natural sequences of the same lengths as the selected coiled coils, and regardless of secondary structure adopted. For the selected coiled coils, Asn occurred in central heptads ≈1.5x more often than expected from all other sequences, whilst in the *N-* and *C-*terminal heptads Asn occurred at rates expected by chance. Again, this fits our experimental observation that centrally located Asn residues bestow some functional benefit on the coiled-coil structure. A similar analysis for N@*d* in trimers revealed no bias in position of these residues; *i.e.*, the distributions of these at the three types of heptad mirrored those found across all natural sequences. This concurs with our experimental findings that whilst the position of N@*d* in a coiled-coil peptide affects the stability of the folded state, it does not influence oligomer state to the same degree as N@*a.* Our dataset does include examples of coiled coils that form part of larger assemblies, and for these, the end effects that we observe may be mitigated by the presence of adjacent polypeptide structure.
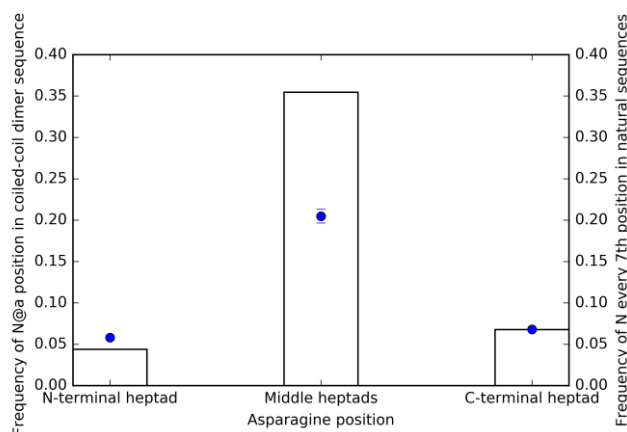
Figure 5: The distribution of N@*a* positions in dimeric coiled-coil sequences. Proportions of Asn residues in different heptad types of dimeric coiled coils (bars), compared with the expected rates of occurrence of Asn determined from all sequences in Uniref50 (blue points, with standard deviation on error bars, n=3, random picks of a dataset ten times the size of the dimeric coiled-coil set, with the same sequence length distribution).

Building on the point about incomplete layers, and following the experimental work of Vinson,[39] we also analyzed the partnering of 33 Asn residues found in the 81 heterodimeric coiled coils of CC+, Table 2b. In these, one third (11) of the examples with N@*a* made Asn-Asn pairs; approximately one quarter (9) did not make complete, reciprocated knobs-into-holes interactions; and the remainder (13) paired with a range of other and predominantly polar residues ([Q,K,R,L] > [A,S,T,E,V]). These data demonstrate further that Asn-Asn pairs are preferred over pairings with other residues. There were not enough examples of structurally defined heterotrimers for us to conduct a similar analysis meaningfully.

These analyses demonstrate the over-representation of Asn residues in the core of parallel coiled-coil assemblies, specifically at *a* positions in dimers and at *d* positions in trimers. These buried Asn residues typically form part of a complete 'layer', *i.e.* they are stabilized by partnering Asn residues at equivalent positions on partner helices. Additionally, these partnerships are more likely to be centrally located in the sequence.

The above aside, the number of structurally defined coiled coils with N@*a* and/or N@*d* is small. Therefore, to glean more about the importance of these inclusions, we examined the more-plentiful protein-sequence databases. As seeds for our searches, we took three classes of sequences that are parallel homo-oligomers from CC+; namely, dimers with 1 x N@*a*, dimers with 1 x N@*d*, and trimers with 1 x N@*d*. (There were only two trimeric sequences with 1 x N@*a*, which we judged an insufficient basis for searches.) We examined the residue distributions at the N@*a* and N@*d* positions of homologous sequences identified from Uniref50[56] using BLAST,[57] Table S3A-C. For each individual search, the level of conservation was calculated as the percentage of Asn maintained at the specified position, Table S3A-C. This gave a series of percentages for each of the three classes, which were then used to calculate the following means and standard deviations, from which some overall trends were clear: N@*d* in trimers was the most conserved at 74% ± 30%; and in dimers N@*a* was more conserved than N@*d*, at 64% ± 34% and 52% ± 22%, respectively. The conservations of the main aliphatic residues found in parallel coiled-coil interfaces are listed in Table S3E for comparison.

For heterodimeric sequences, we performed a similar search using a CC+-derived heterodimer set with a single N@*a* pair, Table S3D. For these, N@*a* was highly conserved at 98% ± 2.7%. By comparison, the conservation of N@*a* in sequences related to those from CC+ with an unpaired N@*a* residues, Table S3F, was low at 12% ± 14%. *n.b.,* this search was necessarily limited to same-chain "heterodimers" to maintain the connection between helix pairings in the search.

Taken together, the results from the bioinformatics analysis show that buried Asn residues play vital roles in directing and maintaining coiled-coil structural states: (1) they are found preferentially at certain sites depending on oligomer state, *i.e.*, N@*a* in dimers and N@*d* in

trimers; (2) at these sites, they are conserved across many sequences; (3) conservation is particularly high in heterodimeric systems, suggesting roles in maintaining specificity and orthogonality of coiled-coil assemblies; and (4) their distribution along the lengths of coiled coils is skewed, with Asn occurring more often than expected in central rather than terminal heptads.

DISCUSSION

We have explored the impact on coiled-coil structure and stability of moving the polar residue asparagine (Asn) through the nominal hydrophobic-core positions (*a* and *d*) in two *de novo* coiled-coil systems: one with core residues traditionally accepted to favor parallel dimers, *i.e. a* = Ile plus *d* = Leu; and the other with *a* = *d* = Ile, which favors parallel trimers. In addition, we have examined the occurrence and conservation of Asn at *a* and *d* (N@*a* and N@*d*) sites of natural coiled coils in the structural and sequence databases. Our results can be summarized as follows:

(1) We confirm that N@*a* generally favors parallel dimer formation, and that N@*d* is tolerated in the trimer state.

(2) However, the impact of these polar inclusions is context dependent in two respects. The position of the inclusion along the coiled-coil sequence is critical. Centrally located inclusions have the most impact. Here they lower coiled-coil stability, but exert more influence over oligomer-state specificity. Whereas, when placed nearer the termini of the coiled-coil sequences they have less influence, leading to raised thermal stabilities, but reduced oligomer-state specification. The likely explanation of this is that when placed nearer to the termini polar Asn residues may be excluded from the hydrophobic cores. The resulting core combinations of only hydrophobic residues then determine the oligomer state adopted. In this respect, *a* = Ile, *d* = Leu, which is traditionally taken as the standard "dimeric" hydrophobic core, is more sensitive than the trimer-forming background, *a* = *d* = Ile.

(3) Developing this last point, *a* = Ile, *d* = Leu and related backgrounds appear to be more plastic, and compatible with multiple oligomer states. Whereas, *a* = *d* = Ile more robustly specifies a trimer. Indeed, for the latter N@*a* inclusions can be tolerated without compromising oligomer state, although overall stability is dramatically reduced.

(4) In natural sequences, Asn inclusions in the hydrophobic cores are highly conserved, particularly N@*a* in dimers and N@*d* in trimers, and especially complementary pairs of N@*a* in heterodimers. Moreover, and in accord with our experiments, Asn is found more often in the central regions of natural coiled-coil sequences.

Together, these experiments and analyses provide an improved basis for rational design and engineering of coiled-coil sequences, and for the application of these in protein engineering and synthetic biology. In this context, our study confirms CC-pIL-I17N (aka CC-Di[8]) as a robust parallel dimeric coiled-coil building block, and further delivers two stable, but not hyperstable, and orthogonal parallel trimeric building blocks, CC-pII-I13N and CC-pII-I20N.

Our findings must be placed in context of a large body of foregoing literature.[13, 18, 20, 35-37, 39-42, 58, 59] Indeed, it is well established that buried Asn residues have marked effects on coiled-coil structure and behavior, and that in these natural systems the extent to which oligomerization state-specifying motifs affect coiled-coil architecture can vary along the length of the coiled coil.[60] For instance, the well-known, and aforementioned example of the N@*a* residue Asn16 in the leucine-zipper region of the transcriptional activator GCN4, which appears vital to maintaining a parallel dimeric oligomeric state.[35] Regarding this natural GCN4 background, our observations appear to be at odds with the studies of Hu and colleagues,[37, 38] who mutated the *a* positions of GCN4-p1 to combinations of Ile and Asn and concluded: "*the effect of an Ile to Asn mutation on the free energy of unfolding is largest at the N terminus of the peptide and decreases almost twofold as we move the substitution from the N to C-terminal heptads*". There are, however, several differences between our systems and GCN4, and between the two studies. First, GCN4 is a natural sequence. As a result, each *a* site is part of a unique knobs-into-hole interaction, so the different *a* sites through GCN4-p1 are not strictly analogous as they are in our systems. Secondly, the region of GCN4 used is 35 residues in length, with potential *a* sites at positions 4, 11, 18, 25, and 32. However, only those at 11, 18, 25, and 32 are examined by Hu *et al.*. Thus, the "*most N-terminal residues*" are in fact located more centrally than those "*towards the C-terminus*"; indeed, residue 18 is the exact midpoint of the coiled-coil sequence.

In addition, Hu and colleagues report that the oligomer state of these various Ile/Asn mutants do not vary from the wild-type dimer, despite the all-hydrophobic *a* = Ile, *d* = Leu core.[38] Although consistent with the data of Harbury and colleagues,[13] it is surprising given our findings[8] and recent studies from Horne and colleagues[48] that show considerable plasticity in oligomer states formed by similar GCN4-p1 core mutants; and indeed more broadly from the studies of Lu, which indicate that mutants of this peptide offer up a wide range of oligomers.[49, 50]

The GCN4 leucine-zipper sequence and structure has been an incredibly rich, useful and productive source of experiments and data for studying and understanding coiled-coil folding and assembly, and, with good reason, this will likely continue. However, given its structural plasticity in response to small mutations, perhaps it is time to move towards using less-complicated and arguably more-robust *de novo* systems for certain synthetic-biology applications.

A different study examining the effects of Asn mutagenesis on self-interactions of membrane-spanning oligo-leucine peptides,[59] does concur with our finding that there is a "*stronger impact of asparagine on self-assembly when located at the center of the oligo-leucine sequence than at the termini*". We suggest that in both the aqueous environment of our studies, and the lipid/detergent system examined by Ruan et al.,[59] the reason for this is straightforward: when positioned adjacent to the termini of the peptide, end fraying, which would itself be increased by proximal destabilizing polar amino acids, effectively removes the Asn as a *bona fide* component of the coiled-coil core, and free to form hydrogen bonds (or other interactions) with the surrounding environment.

What is clear from many studies is that N@*a* is much better accommodated in parallel dimers than it is in trimeric coiled coils, which is apparent in our work from the dramatic loss of thermal stability of the single and double N@*a* mutants in the CC-pII background. The accepted wisdom is that in dimers N@*a* can form inter-chain side-chain hydrogen bonds to offset the penalty of including polar residues in the hydrophobic interface; but these interactions are not possible in trimers.[35] Understanding, the virtual exclusion of N@*d* in dimers is readily explained by the enforced perpendicular packing at these sites,[10, 13] which both precludes side-chain hydrogen bonding, and inhibits escape of the amide groups to solvent. At first site, the mode of action N@*d* in trimers appears less easy to explain: it has been shown from multiple structural examples that the Asn side chains do not form inter-side-chain hydrogen bonds; however, and presumably to help satisfy their hydrogen-bonding potential, the side chains do coordinate ions.[42] We posit that N@*d* are simply *tolerated* and that they may be conserved through evolution as they moderate the stabilities and increase the dynamics of what otherwise would potentially be hyperstable structures.

Pairwise energies for Asn-Asn side-chain interactions have been measured experimentally and found to be favorable compared with "unpaired" Asn in coiled-coil interfaces.[39-41] In a comprehensive study, Vinson and colleagues determined the thermodynamic stabilities of 100 heterodimeric coiled-coil mutants varying ***a-a'*** residue pairs. They state: "*the most extreme example of two amino acids regulating dimer preference has to do with Ile and Asn*". The coupling energies for homotypic Ile-Ile and Asn-Asn ***a-a'*** pairing are favorable at -0.9 and -0.5 kcal/mol, respectively. By contrast, those for the heterotypic pairings, Ile-Asn and Asn-Ile, are unfavorable at +4.3 and +4.9 kcal/mol, respectively. Thus, the relative energies between the homo- and heterotypic pairs are significant, corresponding to changes in binding constants of several orders of magnitude, and likely provide the driving force for the orthogonal, homotypic associations that we have shown herein for our *de novo* systems. On this theme, Gradišar and Jerala present a series of orthogonal coiled-coil pairs in which two asparagine residues per peptide are incorporated at ***a*** positions of four-heptad designed sequences. These Asn incorporations likely play a role alongside the designed charge patterns at ***e*** and ***g*** sites to drive assembly of and discrimination between the heteromeric helical pairings.[61]

Somewhat related to this, Boyken, Baker and colleagues have shown that intricate hydrogen-bonded networks involving multiple polar residues can be designed into α-helical bundles using computational design.[62] In these cases, and although the structures are hyper-thermostable, these networks appear to define the intended structures over aggregated alternatives formed by proteins with exclusively hydrophobic helix-helix interfaces.

In conclusion, we have extensively examined the effects of placing polar Asn residues in actual and potential hydrophobic-core sites of coiled coils on oligomer-state and partner selection and thermal stability. Whilst much remains to be understood about the intricate details of coiled-coil energy landscapes, it becomes ever clearer that polar residues such as Asn have tremendous utility in coiled-coil design and specification in several respects: namely, in designing a single peptide for a given task; in discriminating between alternate oligomer states and topologies; and for generating orthogonal sets of peptides able to

associate in a prescribed and non-promiscuous way in the presence of other coiled coils. Importantly for all of these endeavors, not only is the *inclusion* of a polar Asn residues in the hydrophobic core a design tool, but also their *positioning* along the coiled-coil length must be taken into account and offers additional scope in design.

REFERENCES

1. Lupas, A. N., and Gruber, M. (2005) The structure of alpha-helical coiled coils, *Adv. Prot. Chem. 70*, 37-78.
2. Woolfson, D. N. (2005) The design of coiled-coil structures and assemblies, *Adv. Prot. Chem. 70*, 79-112.
3. Gruber, M., and Lupas, A. N. (2003) Historical review: another 50th anniversary--new periodicities in coiled coils, *Trends Biochem. Sci. 28*, 679-685.
4. Testa, O. D., Moutevelis, E., and Woolfson, D. N. (2009) CC+: a relational database of coiled-coil structures, *Nucleic Acids Res. 37*, D315-322.
5. Moutevelis, E., and Woolfson, D. N. (2009) A periodic table of coiled-coil protein structures, *J. Mol. Biol. 385*, 726-732.
6. Mason, J. M., and Arndt, K. M. (2004) Coiled coil domains: Stability, specificity, and biological implications, *Chembiochem 5*, 170-176.
7. Rose, A., and Meier, I. (2004) Scaffolds, levers, rods and springs: diverse cellular functions of long coiled-coil proteins, *Cell. Mol. Life Sci. 61*, 1996-2009.
8. Fletcher, J. M., Boyle, A. L., Bruning, M., Bartlett, G. J., Vincent, T. L., Zaccai, N. R., Armstrong, C. T., Bromley, E. H., Booth, P. J., Brady, R. L., Thomson, A. R., and Woolfson, D. N. (2012) A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology, *ACS Synth. Biol. 1*, 240-250.
9. Crick, F. H. C. (1953) The Packing of Alpha-Helices - Simple Coiled-Coils, *ACTA Crystallogr. 6*, 689-697.
10. Walshaw, J., and Woolfson, D. N. (2001) SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures, *J. Mol. Biol 307*, 1427-1450.
11. Brown, J. H., Cohen, C., and Parry, D. A. (1996) Heptad breaks in alpha-helical coiled coils: stutters and stammers, *Proteins 26*, 134-145.
12. Hicks, M. R., Holberton, D. V., Kowalczyk, C., and Woolfson, D. N. (1997) Coiled-coil assembly by peptides with non-heptad sequence motifs, *Fold. Des. 2*, 149-158.
13. Harbury, P. B., Zhang, T., Kim, P. S., and Alber, T. (1993) A Switch between 2-Stranded, 3-Stranded and 4-Stranded Coiled Coils in Gcn4 Leucine-Zipper Mutants, *Science 262*, 1401-1407.
14. Woolfson, D. N., and Alber, T. (1995) Predicting oligomerization states of coiled coils, *Prot. Sci. 4*, 1596-1607.
15. Woolfson, D. N., Bartlett, G. J., Bruning, M., and Thomson, A. R. (2012) New currency for old rope: from coiled-coil assemblies to alpha-helical barrels, *Curr. Opin. Struc. Biol. 22*, 432-441.
16. Litowski, J. R., and Hodges, R. S. (2002) Designing heterodimeric two-stranded alpha-helical coiled-coils. Effects of hydrophobicity and alpha-helical propensity on protein folding, stability, and specificity, *J. Biol. Chem. 277*, 37272-37279.
17. Grigoryan, G., and DeGrado, W. F. (2011) Probing Designability via a Generalized Model of Helical Bundle Geometry, *J. Mol. Biol. 405*, 1079-1100.

18. O'Shea, E. K., Klemm, J. D., Kim, P. S., and Alber, T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil, *Science 254*, 539-544.

19. Lavigne, P., Sonnichsen, F. D., Kay, C. M., and Hodges, R. S. (1996) Interhelical salt bridges, coiled-coil stability, and specificity of dimerization, *Science 271*, 1136-1138.

20. O'Shea, E. K., Lumb, K. J., and Kim, P. S. (1993) Peptide 'Velcro': design of a heterodimeric coiled coil, *Curr. Biol. 3*, 658-667.

21. Nautiyal, S., and Alber, T. (1999) Crystal structure of a designed, thermostable, heterotrimeric coiled coil, *Prot. Sci. 8*, 84-90.

22. Nautiyal, S., Woolfson, D. N., King, D. S., and Alber, T. (1995) A designed heterotrimeric coiled coil, *Biochemistry 34*, 11645-11651.

23. Bromley, E. H., Sessions, R. B., Thomson, A. R., and Woolfson, D. N. (2009) Designed alpha-helical tectons for constructing multicomponent synthetic biological systems, *J. Am. Chem. Soc. 131*, 928-930.

24. Thomas, F., Boyle, A. L., Burton, A. J., and Woolfson, D. N. (2013) A set of de novo designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime, *J. Am. Chem. Soc. 135*, 5161-5166.

25. Thomson, A. R., Wood, C. W., Burton, A. J., Bartlett, G. J., Sessions, R. B., Brady, R. L., and Woolfson, D. N. (2014) Computational design of water-soluble alpha-helical barrels, *Science 346*, 485-488.

26. Yoshizumi, A., Fletcher, J. M., Yu, Z., Persikov, A. V., Bartlett, G. J., Boyle, A. L., Vincent, T. L., Woolfson, D. N., and Brodsky, B. (2011) Designed coiled coils promote folding of a recombinant bacterial collagen, *J. Biol. Chem. 286*, 17512-17520.

27. Boyle, A. L., Bromley, E. H., Bartlett, G. J., Sessions, R. B., Sharp, T. H., Williams, C. L., Curmi, P. M., Forde, N. R., Linke, H., and Woolfson, D. N. (2012) Squaring the circle in peptide assembly: from fibers to discrete nanostructures by de novo design, *J. Am. Chem. Soc. 134*, 15457-15467.

28. Fletcher, J. M., Harniman, R. L., Barnes, F. R., Boyle, A. L., Collins, A., Mantell, J., Sharp, T. H., Antognozzi, M., Booth, P. J., Linden, N., Miles, M. J., Sessions, R. B., Verkade, P., and Woolfson, D. N. (2013) Self-assembling cages from coiled-coil peptide modules, *Science 340*, 595-599.

29. Robson Marsden, H., and Kros, A. (2010) Self-assembly of coiled coils in synthetic biology: inspiration and progress, *Angew. Chem. Int. Ed. Engl. 49*, 2988-3005.

30. Boyle, A. L., and Woolfson, D. N. (2011) De novo designed peptides for biological applications, *Chem. Soc. Rev. 40*, 4295-4306.

31. Rackham, O. J., Madera, M., Armstrong, C. T., Vincent, T. L., Woolfson, D. N., and Gough, J. (2010) The evolution and structure prediction of coiled coils across all genomes, *J. Mol. Biol. 403*, 480-493.

32. Huang, P. S., Oberdorfer, G., Xu, C., Pei, X. Y., Nannenga, B. L., Rogers, J. M., DiMaio, F., Gonen, T., Luisi, B., and Baker, D. (2014) High thermodynamic stability of parametrically designed helical bundles, *Science 346*, 481-485.

33. Gillingham, A. K., and Munro, S. (2003) Long coiled-coil proteins and membrane traffic, *Bba-Mol. Cell Res. 1641*, 71-85.

34. O'Shea, E. K., Rutkowski, R., Stafford, W. F., 3rd, and Kim, P. S. (1989) Preferential heterodimer formation by isolated leucine zippers from fos and jun, *Science 245*, 646-648.

35. Lumb, K. J., and Kim, P. S. (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil, *Biochemistry 34*, 8642-8648.

36. Gonzalez, L., Jr., Woolfson, D. N., and Alber, T. (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper, *Nat. Struct. Biol. 3*, 1011-1018.

37. Zeng, X., Herndon, A. M., and Hu, J. C. (1997) Buried asparagines determine the dimerization specificities of leucine zipper mutants, *Proc. Natl. Acad. Sci. U.S.A. 94*, 3673-3678.

38. Zhu, H., Celinski, S. A., Scholtz, J. M., and Hu, J. C. (2000) The contribution of buried polar groups to the conformational stability of the GCN4 coiled coil, *J. Mol. Biol. 300*, 1377-1387.

39. Acharya, A., Rishi, V., and Vinson, C. (2006) Stability of 100 homo and heterotypic coiled-coil a-a' pairs for ten amino acids (A, L, I, V, N, K, S, T, E, and R), *Biochemistry 45*, 11324-11332.

40. Steinkruger, J. D., Woolfson, D. N., and Gellman, S. H. (2010) Side-Chain Pairing Preferences in the Parallel Coiled-Coil Dimer Motif: Insight on Ion Pairing between Core and Flanking Sites, *J. Am. Chem. Soc. 132*, 7586-+.

41. Steinkruger, J. D., Bartlett, G. J., Hadley, E. B., Fay, L., Woolfson, D. N., and Gellman, S. H. (2012) The d '-d-d ' Vertical Triad Is Less Discriminating Than the a '-a-a ' Vertical Triad in the Antiparallel Coiled-Coil Dimer Motif, *J. Am. Chem. Soc. 134*, 2626-2633.

42. Hartmann, M. D., Ridderbusch, O., Zeth, K., Albrecht, R., Testa, O., Woolfson, D. N., Sauer, G., Dunin-Horkawicz, S., Lupas, A. N., and Alvarez, B. H. (2009) A coiled-coil motif that sequesters ions to the hydrophobic core, *Proc. Natl. Acad. Sci. U.S.A. 106*, 16950-16955.

43. Gorbet, G., Devlin, T., Hernandez Uribe, B. I., Demeler, A. K., Lindsey, Z. L., Ganji, S., Breton, S., Weise-Cross, L., Lafer, E. M., Brookes, E. H., and Demeler, B. (2014) A parametrically constrained optimization method for fitting sedimentation velocity experiments, *Biophys. J. 106*, 1741-1750.

44. http://sednterp.unh.edu/

45. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure, *Nat. Protoc. 1*, 2876-2890.

46. Hicks, M. R., Walshaw, J., and Woolfson, D. N. (2002) Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts, *J. Struct. Biol. 137*, 73-81.

47. Lupas, A. (1996) Coiled coils: new structures and new functions, *Trends Biochem. Sci. 21*, 375-382.

48. Oshaben, K. M., Salari, R., McCaslin, D. R., Chong, L. T., and Horne, W. S. (2012) The native GCN4 leucine-zipper domain does not uniquely specify a dimeric oligomerization state, *Biochemistry 51*, 9581-9591.

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

50. Liu, J., Zheng, Q., Deng, Y., Cheng, C. S., Kallenbach, N. R., and Lu, M. (2006) A seven-helix coiled coil, *Proc. Natl. Acad. Sci. U.S.A. 103*, 15457-15462.

51. Ogihara, N. L., Weiss, M. S., Degrado, W. F., and Eisenberg, D. (1997) The crystal structure of the designed trimeric coiled coil coil-V(a)L(d): Implications for engineering crystals and supramolecular assemblies, *Protein Sci. 6*, 80-88.

52. Schneider, J. P., Lombardi, A., and DeGrado, W. F. (1998) Analysis and design of three-stranded coiled coils and three-helix bundles, *Folding & Design 3*, R29-R40.

53. Thompson, K. E., Bashor, C. J., Lim, W. A., and Keating, A. E. (2012) SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains, *ACS Synth. Biol. 1*, 118-129.

54. Negron, C., and Keating, A. E. (2014) A set of computationally designed orthogonal antiparallel homodimers that expands the synthetic coiled-coil toolkit, *J. Am. Chem. Soc. 136*, 16544-16556.

55. Bruhn, J. F., Barnett, K. C., Bibby, J., Thomas, J. M., Keegan, R. M., Rigden, D. J., Bornholdt, Z. A., and Saphire, E. O. (2014) Crystal structure of the nipah virus phosphoprotein tetramerization domain, *J. Virol. 88*, 758-762.

56. Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., Wu, C. H., and UniProt, C. (2015) UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches, *Bioinformatics 31*, 926-932.

57. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. (2009) BLAST+: architecture and applications, *BMC Bioinformatics 10*, 421.

58. Choma, C., Gratkowski, H., Lear, J. D., and DeGrado, W. F. (2000) Asparagine-mediated self-association of a model transmembrane helix, *Nat. Struct. Biol. 7*, 161-166.

59. Ruan, W., Lindner, E., and Langosch, D. (2004) The interface of a membrane-spanning leucine zipper mapped by asparagine-scanning mutagenesis, *Prot. Sci. 13*, 555-559.

60. Ciani, B., Bjelic, S., Honnappa, S., Jawhari, H., Jaussi, R., Payapilly, A., Jowitt, T., Steinmetz, M. O., and Kammerer, R. A. (2010) Molecular basis of coiled-coil oligomerization-state specificity, *Proc. Natl. Acad. Sci. U.S.A. 107*, 19850-19855.

61. Gradisar, H., and Jerala, R. (2011) De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers, *J. Pept. Sci. 17*, 100-106.

62. Boyken, S. E., Chen, Z., Groves, B., Langan, R. A., Oberdorfer, G., Ford, A., Gilmore, J. M., Xu, C., DiMaio, F., Pereira, J. H., Sankaran, B., Seelig, G., Zwart, P. H., and Baker, D. (2016) De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity, *Science 352*, 680-687.