Published in final edited form as:
Nat Chem Biol. ; 7(12): 935–941. doi:10.1038/nchembio.692.

A *de novo* peptide hexamer with a mutable channel

Nathan R. Zaccai[1,†], **Bertie Chi**[1,2,†], **Andrew R. Thomson**[2], **Aimee L. Boyle**[2], **Gail J. Bartlett**[2], **Marc Bruning**[2], **Noah Linden**[3], **Richard B. Sessions**[1], **Paula J. Booth**[1], **R. Leo Brady**[1], and **Derek N. Woolfson**[1,2]

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

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

[3]Department of Mathematics, University of Bristol, University Walk, Bristol, BS8 1TW, UK

Abstract

The design of new proteins that expand the repertoire of natural protein structures represents a formidable challenge. Success in this area would increase understanding of protein structure, and present new scaffolds that could be exploited in biotechnology and synthetic biology. Here we describe the design, characterisation and X-ray crystal structure of a new coiled-coil protein. The *de novo* sequence forms a stand-alone, parallel, 6-helix bundle with a channel running through it. Although lined exclusively by hydrophobic leucine and isoleucine side chains, the 6 Å channel is permeable to water. One layer of leucine residues within the channel is mutable accepting polar aspartic acid (Asp) and histidine (His) side chains, and leading to subdivision and organization of solvent within the lumen. Moreover, these mutants can be combined to form a stable and unique $(Asp\text{-}His)_3$ heterohexamer. These new structures provide a basis for engineering *de novo* proteins with new functions.

Nature presents many beautiful and intriguing protein structures. However, these appear to have explored only a limited part of the potential structural space[1,2,3]. Thus, although natural proteins clearly provide excellent frameworks for the evolution of biological function and a basis for protein engineering, it is pertinent to explore past these boundaries and attempt to expand the repertoire of protein structures and assemblies synthetically[4]. Rational or *de novo* protein design offers one route towards this goal, but challenges remain. Indeed, except for a small number cases — for example, the Top7 protein[5] — successful protein designs have focused on reproducing or mimicking existing protein architectures. Other new protein structures can be generated *in silico*[6], but our current incomplete understanding of sequence-to-structure relationships in proteins leaves the major obstacle of translating these into real sequences for experimental testing. For certain assemblies of α-helices, known as

Correspondence and requests for materials should be addressed to D.N.Woolfson@bristol.ac.uk and L.Brady@bris.ac.uk..
†These authors contributed equally to this work.

Supplementary Methods for Bioinformatics and Modelling are given in the Supplementary Information.

Author contributions: ART, BC, GJB, and DNW designed the peptide sequences; BC and ART synthesized the peptides; BC, ALB and ART performed the solution-phase biophysics; NRZ, BC and RLB solved the peptide X-ray crystal structures; GJB performed the bioinformatics; MB, ART and DNW conducted the SOCKET and TWISTER analyses; RBS, NL and DNW conceived and conducted the computational modelling; PJB, RLB and DNW conceived and supervised the experimental program; DNW wrote the manuscript.

Competing financial interests: The authors declare no competing financial interests.

Additional information: Supplementary information is available online at http://www.nature.com/naturechemicalbiology/. Atomic coordinates and structure factors have been deposited with the Protein Data Bank with the following accession numbers: CC-Tet, 3R4A; CC-Tet-Φ_{22}, 3R4H; CC-Hex-Φ_{22}, 3R3K; CC-Hex-H_{24}, 3R47; CC-Hex-D_{24}, 3R46; CC-Hex-$D_{24}Y_{15}$:CC-Hex-H_{24} mixture, 3R48. Reprints and permissions information is available at http://npg.nature.com/reprintsandpermissions/.

Europe PMC Funders Author Manuscripts

coiled coils, there are good sequence-to-structure relationships[7,8]. Nevertheless, even for this relatively well-understood class of protein structures considerable gaps in our knowledge remain.

Our understanding of coiled-coil domains stems from a unique sequence-to-structure relationship in which patterns of hydrophobic (*H*) and polar (*P*) residues—usually as a heptad repeat, $(HPPHPPP)_n$, and often denoted ***abcdefg***, Fig. 1a—encode amphipathic helices that interact *via* their hydrophobic seams to form helical bundles. Importantly, the interaction of side chains between helices is intimate and highly specific, being referred to as knobs-into-holes (KIH) packing[9,10]. The heptad repeat and KIH packing are the accepted hallmarks of α-helical coiled-coil sequences and structures, and provide a firm basis for their rational design[11,12], including computationally aided designs[13,14,15,16,17]. However, the coiled-coil helices may associate in many ways to give bundles of different oligomeric states, and with the helices arranged in parallel, antiparallel or mixed topologies[18,19]. Although much progress has been made to understand sequence-to-structure relationships for parallel dimers, trimers and tetramers, our knowledge is far from complete; these states represent only part of potential coiled-coil structural space, leaving many structures poorly understood and untapped in design and engineering.

Natural, “classical” coiled coils with 2, 3, 4 and 5 helices cemented by symmetric cycles of KIH interactions are all known, in respect of having been structurally defined to atomic resolution[18,19], Fig. 1b. Beyond these, ring-like arrangements of helices are more complicated. For example, an engineered mutant of the GCN4-p1 leucine zipper forms a parallel 7-helix bundle[20]. As judged by the program SOCKET,[10] which indentifies KIH interfaces in protein structures, this comprises a ring of back-to-back coiled-coil dimers. Successive dimers are slipped relative to the long axis of the structure, ultimately leading to helices 1 and 7 being offset from one another by a whole heptad to give a spiralled structure, Supplementary Fig. 1. Similarly, part of the natural bacterial exporter TolC[21] is a 12-helix ring comprising six antiparallel dimers[22]. As yet there are no known coiled coils with rings of 6, 8, 9, 10 and 11 helices. Indeed, it has been proposed that a classical 6-helix coiled-coil should not be possible[23]. However, we note that others have considered the possibility of 6-helix coiled-coil-like assemblies, and indeed developed functional models and protein designs based on these[17,24,25]. As rings of more than five helices all have central channels, discovering or designing more such structures would provide a basis for the *de novo* design of tubular and ion-channel proteins[24,26]. Therefore, we sought to explore the possibilities for higher-order coiled-coil assemblies through protein design.

Here we describe the design and full structural characterisation of the first stand-alone coiled-coil hexamer, CC-Hex. Solution-phase characterisations were consistent with a fully α-helical and highly stable hexamer; and an X-ray crystal structure of the *de novo* peptide revealed a bundle of six, parallel α-helices. This represents a new protein fold. Moreover, the helices combine to form a well-defined central pore approximately 6 Å in diameter. For the parent peptide, this channel is lined by hydrophobic residues. However, the lumen is mutable accepting polar aspartic acid and histidine side chains, as well a symmetric $(AB)_3$-type combination of these.

RESULTS

Design rationale

The design of novel enlarged coiled-coil assemblies is currently severely limited by a paucity of sequence and structural information for assemblies above tetramer. Moreover and related to this, although both sequence- and structure-based computational methods are improving[16,28], there are no approaches that allow high-order coiled-coil oligomers to be

predicted or designed *de novo* with confidence *in silico*. Therefore, we took the following semi-empirical approach to the problem. Our starting point was a parallel tetramer, the highest-order coiled coil for which reliable sequence-to-structure relationships are available to direct protein design[11,29]. First, we made a fully *de novo* sequence, CC-Tet, in which the position of each residue in a 32-residue, 4-heptad sequence was accounted for, Fig. 2a: the ***a*** and ***d*** positions were made leucine (Leu, L) and isoleucine (Ile, I), respectively, based on the tetrameric mutant of the natural leucine-zipper peptide, GCN4-p1[29]; the ***g*** and ***e*** positions were made complementary glutamic acid (Glu, E) and lysine (Lys, K), respectively, to provide inter-chain charge complementarity[30,31,32]; the ***b*** and ***c*** positions were made helix-favouring alanine (Ala, A); the ***f*** positions were varied for solubility, and to provide a chromophore; and the sequence was terminated by glycine (Gly, G) and Ala spacers. This gave a sequence based on the heptad repeat $(ELAAIKX)_4$, Table 1. Consistent with these design principles, an X-ray crystal structure of CC-Tet determined at 2.1 Å resolution revealed a parallel 4-helix bundle, Figs. 2b&c and Supplementary Table 1. This tested positive as a left-handed coiled coil by the programs SOCKET and TWISTER[10,33], which identify KIH packing and general coiled-coil structural features, respectively, Supplementary Table 2. In solution, CC-Tet was largely α-helical as judged by circular dichroism (CD) spectroscopy, had high thermal stability, and sedimented as a tetramer in analytical ultracentrifugation, Figs. 2d – f.

Next, we attempted to make larger oligomers in the CC-Tet framework by expanding the hydrophobic interfaces between its helices. This was partly inspired by the protein engineering study on GCN4-p1[20] that produced the aforementioned 7-helix bundle. However, our decision stemmed mainly from earlier observations that the hydrophobic interfaces between the helices of coiled coils become progressively wider with increasing oligomer state[29,34]. More specifically, the *core* KIH interactions, *i.e.* those centred on the ***a*** and ***d*** residues of the heptad repeat, are supplemented by *peripheral* KIH interactions involving the ***e***, ***g***, ***b*** and ***c*** positions[23] (this is discussed in more detail below). We hypothesised that extending the hydrophobic interface in this way would stabilise higher-order oligomer states, and hence open possibilities for different and indeed completely new folding arrangements[19,23].

With this rationale in mind, we exchanged all of the Lys residues at ***e*** positions in CC-Tet with Ala residues at ***b***, Fig. 2a and Table 1. In solution, the resulting peptide was predominantly α-helical and highly stable to heat, Figs. 2d&e. Intriguingly, AUC analysis clearly indicated a hexamer, Fig. 2f. The X-ray crystal structure was solved to 2.2 Å using experimental phasing from the anomalous signal of iodine introduced via 4-iodophenylalanine incorporated in place of tryptophan-22 (Trp-22), Table 1 and Supplementary Table 1. This revealed six helices arranged in a parallel bundle with C_6 symmetry, Figs. 3a&b. As we arrived at it semi-empirically and with a degree of serendipity, we named the peptide retrospectively as CC-Hex. The crystal structure of CC-Hex is surprising and interesting for a number of other reasons as outlined below.

CC-Hex has a new protein fold

As judged by both TWISTER and SOCKET CC-Hex formed a regular parallel left-handed hexameric coiled coil, Supplementary Table 2. As such, it is a new protein fold. To confirm this, we searched the RCSB Protein Data Bank of protein structures[35]. The closest matches were for the M2 segment of the acetylcholine receptor (PDB code 1eq8[36]); the cartilage oligomeric matrix protein (COMP, 1vdf[37]); an engineered tryptophan-zipper (1t8z[38]); the hexameric tyrosine-coordinated heme protein (HTHP, 2oyy[39]); and the aforementioned heptameric mutant of GCN4-p1 (2hy6[20]), Supplementary Fig. 2. However, M2, COMP and the tryptophan-zipper are all pentamers; and, as stated above, the GCN4 mutant is a

heptamer and a complex coiled coil. Part of HTHP does form a parallel 6-helix coiled coil, but this is a small part of a larger globular assembly: according to SOCKET analysis, the coiled-coil region has just two KIH layers; moreover, we made a synthetic peptide corresponding to this region and it did not fold, Supplementary Fig. 3. Therefore, as a free-standing, parallel, 6-helix bundle with complete and classical coiled-coil packing CC-Hex is unique.

CC-Hex is a classical coiled coil

On the basis that KIH packing between helices would be severely compromised, we had postulated that a true hexameric coiled coil, like CC-Hex, should not exist[23]. Indeed, SOCKET analysis of CC-Hex revealed that the Leu at ***a*** and Ile at ***d*** form KIH interactions at the extremes of the packing-angle distributions observed for natural coiled-coil dimers, trimers and tetramers, Fig. 4. Nonetheless, the hexamer is clearly accessible within the coiled-coil energy landscape. Also, the hydrophobic interfaces between the helices were expanded as anticipated in the design rationale: the KIH packing extended to the Ala and Glu residues at ***e*** and ***g***, respectively; effectively giving two complete and offset heptad repeats, LxxxAxx and xxxIxxE, within a conventional assignment, ***abcdefg***.

As we had also presented previously[23], such *offset double heptad repeats* afford the opportunity of larger coiled-coil oligomers and complex coiled-coil arrangements. However, the number of possible coiled-coil assemblies is limited based on constraints that follow from superposition of the heptad repeat on helical geometry. In essence, the two heptad repeats each set up a hydrophobic seam on the surface of the α-helix: the LxxxAxx repeat is assigned as the ***a/e*** seam, and xxxIxxE as the ***d/g*** seam. When projected onto an idealised coiled-coil helical wheel, the calculated angle between the centres of these two seams is 103.5°, Supplementary Fig. S1. This would be the ideal helix-contact angle to maximise coiled-coil KIH contacts involving both of these seams. However, this is not possible geometrically. The closest oligomer packing consistent with 103.5° is a pentamer, which has geometrically idealised helix-contact angles of 108° (180° – 360°/5). Thus, for a pentamer, there would be a discrepancy of just 4.5° from the ideal helix-contact angles derived from considering the hydrophobic seams and simple geometry. However, the contact angle realised is also affected by steric constraints of the side chains at ***g***, ***a***, ***d*** and ***e***. For CC-Hex, the small Ala residues at ***g***, and the nature of the packing of the Glu residues at ***e***, Figs. 4a&b, appear to relax these constraints from the heptad repeats. In this case helix-contact angles of 120° are achieved, *i.e.* it is a symmetric hexamer; with the discrepancy of 17° from the ideal angle based on hydrophobic packing being the largest observed in classical coiled-coil structures to date.

Extending this analysis to the 7-helix variant of GCN4[20] is illuminating also. This structure also assembles *via* ***a/e*** and ***d/g*** interfaces, but the presence of Ala residues at both ***e*** and ***g*** causes the angle between adjacent helices to be wider still. The measured helix-contact angles are ~132°, which fall between the geometrically ideal values of 128.6° and 135° for heptamer and octamer, respectively, Supplementary Fig. 1. This conflict is resolved by the aforementioned spiralling arrangement of 7 helices; as a result, the helix-contact angles reduce to 128.6°, *i.e.* for the heptamer, when projected onto the plane perpendicular to the long axis of the assembly.

There are two other combinations of offset double heptad repeats[23], which lead to different ideal coiled-coil helix-contact angles, Supplementary Fig. 1. Therefore, there are three potential discrepancy frames to consider between these and the geometrically ideal structures. In the coiled-coil repeats and structures observed to date, the frame that minimises this discrepancy is observed. This applies to all cases except the heptamer, which falls between two of the angles dictated by the heptad repeats as noted above. Again, the

sequence and structure are conflicted, and we suggest that the heptamer is the tipping point between classical and complex coiled-coil assemblies.

These analyses and observations immediately suggest new computational routes towards the design of higher-order oligomer states with 8 or more helices, which we will explore elsewhere.

A defined and mutable channel

The helices of CC-Hex form a ring and create a central linear channel that extends through the whole molecule, Fig. 3c&d. This is lined exclusively by the methyl groups from the Leu and Ile residues at ***a*** and ***d***. The channel is open at both the *N*- and *C*-terminal ends, and, from the X-ray data, electron density is present within it. This density was best modelled as a chain of water molecules, with discontinuities at each Leu layer where the channel constricts slightly.

Intrigued by the nature and potential applications of the channel, we synthesized two non-conservative mutants in which the hydrophobic position Leu-24 — the last ***a*** site of the coiled-coil repeat — was replaced by the polar residues aspartic acid (Asp, D), a near-isosteric change, and to histidine (His, H). These peptides, L24D and L24H, were crystallized and their X-ray structures determined by molecular replacement using the CC-Hex structure. The *C*-termini of the L24H peptides were slightly uncoiled, but otherwise both structures were parallel hexameric coiled coils. The 1.75 Å structure of L24D revealed well-ordered solvent molecules within the channel, Fig. 5a, notably: two clusters of four water molecules, one above and one below the ring of Asp side chains; these linked to a chain of water molecules along the channel, which was again disrupted at the remaining Leu layers. In both structures, the rings of polar residues and associated solvent constricted the channel, effectively partitioning it into a main chamber to the *N*-terminal side, and an antechamber to the other, Fig. 5a&c.

The proximity of six polar and potentially charged residues at the heart of the structure is expected to destabilize both assemblies. This is for electrostatic reasons in L24D, where the mutation is nearly isosteric, Fig. 5b; and possibly for both electrostatic and steric reasons in L24H; Fig. 5c. Indeed, under similar conditions to those used for the parent—100 μM peptide, pH 7.4, 20 °C—L24D and L24H were compromised in both their helical folding, stability and association in solution, Supplementary Figs. 4&5. For instance, the helicity of L24D monitored as a function of pH showed a transition from largely unfolded (> pH 6.5) to fully folded (< pH 4).

These data fitted to an apparent pK_a of 5.2, Supplementary Fig. 4, which is perturbed from the intrinsic values for Asp (3.9) and Glu (4.3). Similar data for L24H were flatter, Supplementary Fig. 4, and could not be analyzed to give a pK_a value. Consistent with these observations, in the two crystal structures the carboxylate and imidazole side chains tilt away from each other, Figs. 5d&e; though we cannot discount the possibility that the shifted pK_a arises in part from the glutamic acid residues in the sequences, as these are at ***g*** sites and also involved in the helix-helix interfaces, Fig. 4a.

A new $(AB)_3$-type heterohexamer

Together, the above data on the mutant homohexamers show that six Asp or six His side chains are accommodated within the channel, but with some frustration. To test if these residues could complement one another within the core to relieve this strain, the two peptides were mixed in a 1:1 ratio. The pH-dependences of both the helicities (Supplementary Fig. 4), and the stabilities (Supplementary Fig. 5) of the experimental

mixtures showed them to be more helical and more stable than those predicted by the simple averages of the L24D and L24H data; that is, except at pH 3.4 where the Asp side chains are expected to be neutral. This suggested that the two peptides form a complex above pH 3.4. Indeed, a 1:1 mixture of L24D and L24H sedimented as a hexamer in solution at pH 4.4, Supplementary Fig. 6. Moreover, crystallisation from such a mixture at pH 6.5, followed by X-ray structure determination revealed an $(AB)_3$-type heterohexamer with a clear and intriguing alternation of L24D and L24H chains, Fig. 5f. In this case, the Asp and His side chains tilt into plane and towards each other making hydrogen-bonded contacts.

Whilst the alternating $(Asp\text{-}His)_3$ arrangement is intuitively reasonable, it is not obvious: there are 13 distinct possible combinations of the two peptides in a parallel hexamer, Supplementary Fig. 7. To begin exploring the energy landscape of this peptide-assembly system, we constructed the 13 models based on the L24D:L24H crystal structure. In each case, 11.4 million conformers were generated in which the χ_2 torsion angles of the Asp and His side chains were systematically incremented. The relative energies of all of these were compared using the consistent-valence forcefield in Discover. Of the three possible non-equivalent 1:1 combinations of Asp and His—*i.e.*, Asp_3His_3, Asp_2-His-Asp-His_2 and $(Asp\text{-}His)_3$—the alternating combination had the lowest energy; indeed for the case where both the Asp and His side chains are fully charged, this had the lowest modelled energy of all 13 combinations, Supplementary Fig. 7. Whilst calculations of this simplicity cannot distinguish between possible alternative oligomeric states—indeed, the development of computational methods to do this is still in its infancy[16,28]—their combination with experimentally derived structures as described here suggests that the CC-Hex system may be controlled and directed to give a prescribed hetero-hexameric assembly, which holds considerable promise for future rational designs using the system.

DISCUSSION

This report of the first example of a parallel 6-helix coiled-coil structure, CC-Hex, not only presents a new protein fold, but it also changes and extends our understanding of coiled-coil assemblies. Specifically, it pushes the boundaries for formation of *classical coiled coils*, which had previously encompassed only dimers, trimers, tetramers and pentamers. CC-Hex adds to this group, and represents the largest coiled coil with a classical cycle of KIH interactions[23]. All currently observed assemblies above this are *complex coiled coils*, comprising multiple coiled-coil interfaces.[23] The heptamer, for which there is currently only a single example from a mutant protein, appears to be the tipping point between these two classes. We posit that all structures above a heptamer will necessarily be complex coiled coils. This is based on the comparison of geometrically ideal helix-helix contact angles and those predicted and constrained through KIH interactions, which arise from heptad sequence repeats and helical geometry, Supplementary Fig. 1[23]. This is not to say that higher-order structures, such as octamer and above, could not be achieved; it is just that these would require altogether different sequence patterns, and likely judicious placement of residues at the helical interfaces[22,23].

Coiled-coil and coiled-coil-like helical oligomers above pentamer all have central channels that increase in diameter with increasing oligomer state. These present opportunities for the rational design of tubular proteins, membrane-spanning channels and other targets[17,26]. Coiled coils observed thus far that potentially add to this effort include those with 5, 6 and 7 helices, which provide channel diameters ranging from 5 – 7 Å, Supplementary Fig. 1. Other natural α-helical pore-containing assemblies exist, but these are large, membrane-associated proteins and not stand-alone coiled coils[21,40,41]. In these respects, the CC-Hex structure is particularly interesting and appealing for engineering and design purposes because it is a small, defined peptide, and it has a narrow and well-defined channel approximately ~6 Å in

diameter running uninterrupted through the centre of the helical bundle. Though hydrophobic, the channel of CC-Hex is of similar dimensions to water molecules, and in the crystal structures it is occupied by solvent. It is also robust to mutation: the introduction of polar aspartic acid and histidine residues within the channel gives rise to mutant homohexamers, and an alternating $(Asp\text{-}His)_3$ heterohexamer. In each case, the side chains introduced by these mutations form active constrictions that subdivide the channels and alter the organisation of bound solvent. On this basis, we put forward the CC-Hex as a new scaffold for rational, structure-based protein design and engineering, particularly for tubular proteins and membrane-spanning channels with defined lumens.

METHODS

Peptide synthesis

Peptide synthesis was carried out according to standard Fmoc SPPS protocols on a CEM Liberty microwave-assisted automated peptide synthesiser (CEM Corporation, Buckingham, UK). All peptides were synthesised on a Rink amide ChemMatrix resin (PCAS Biomatrix, Quebec, Canada), making use of HBTU activation. Following automated synthesis the peptides were acetylated using acetic anhydride and pyridine in DMF. Cleavage of the crude peptide from the resin was achieved by treatment with a 95:2.5:2.5 mix of trifluoroacetic acid:water:triisopropylsilane for 2 h at room temperature. The cleavage mix was dripped into ten times the volume of cold (0 °C) diethyl ether and the resultant precipitate was isolated by centrifugation. The solid was redissolved in a 1:1 mix of water and acetonitrile and freeze-dried to give the crude peptide. This material was purified using reversed-phase HPLC using a C18 column (10 × 150 mm, Kromatek, UK) and running a gradient from 20 % to 80 % acetonitrile in water over 30 minutes.

Circular dichroism spectroscopy

CD measurements were made using a JASCO J-815 spectropolarimeter fitted with a Peltier temperature controller (Jasco UK, Great Dunmow, UK). Peptide samples were made up as 1, 10 and 100 μM solutions in phosphate buffered saline (PBS) (8.2 mM sodium phosphate, 1.8 mM potassium phosphate, 137 mM sodium chloride, 2.7 mM potassium chloride, pH 7.4). CD spectra were recorded in 10, 5, and 1 mm path length quartz cuvettes at 20°C for the 1, 10 and 100 μM peptide concentrations, respectively. The instrument was set with a scan rate of 50 nm/min, 1 nm interval, 1 nm bandwidth and a response time of 1 s. Melting and cooling data were acquired at 222 nm, between 5 and 95°C, with settings as above and a ramping rate of 40°C/hr. pH Titration experiments were conducted at 100 μM total peptide concentrations for the CC-Hex-D_{24} and CC-Hex-H_{24} mutants, and at and 200 μM for the mixtures of the two peptides. The peptide samples were made up in saline phosphate buffer (20 mM sodium mono-hydrogen phosphate, 137 mM sodium chloride) pH adjusted to desired values with a solution of 0.1 M citric acid containing 137 mM sodium chloride. Baselines recorded using the same buffer, cuvettes and parameters were subtracted from the data. The spectra were converted from ellipticities (deg) to molar ellipticities (deg cm^2 dmol res^{-1}) by normalizing for concentration of peptide bonds and cell path length.

Analytical ultracentrifugation

Sedimentation-equilibrium experiments were conducted at 20°C in a Beckman-Optima XL-I analytical ultracentrifuge using an An-60 Ti rotor (Beckman-Coulter, High Wycombe, UK). Solutions of CC-Tet and CC-Hex were prepared in PBS (pH 7.4) to give initial absorbances of 0.2, 0.4 and 0.6. Solutions of CC-Hex-D_{24}, CC-Hex-H_{24} and a 1:1 mix of CC-Hex-D_{24}:CC-Hex-H_{24} were prepared in the same way but using saline phosphate buffer adjusted to pH 4.4. The peptides were centrifuged at speeds in the range 16,000 – 35,000 rpm. Datasets were fitted to a single, ideal species model using *Ultrascan*[42]. The partial specific

volume for each of the peptides (CC-Tet, 0.7740 $cm^3\ g^{-1}$; CC-Hex, 0.7740 $cm^3\ g^{-1}$; CC-Hex-D_{24}, 0.7651 $cm^3\ g^{-1}$; CC Hex-H_{24}, 0.7654 $cm^3\ g^{-1}$; CC-Hex-D_{24}:CC-Hex-H_{24}, 0.7653 $cm^3 g^{-1}$) and the buffer density (PBS, pH 7.4, 1.0054 g cm^{-3}; phosphate-buffered saline, pH 4.4, 1.0149 g cm^{-3}) were calculated using *Sednterp* (http://www.jphilo.mailway.com/).

Crystal-structure determination

Freeze-dried peptides were resuspended directly in deionised water at concentrations of ~10 mg/ml for vapour diffusion crystallization trials using standard commercial screens at 18°C with 0.2 μl peptide equilibrated with 0.2 μl of reservoir solution. Crystals of CC-Tet were obtained with reservoir buffer containing 0.2 M lithium sulfate, 0.1 M Tris pH 8.5 and 1.26 M ammonium sulphate. CC-Tet-Φ_{22} crystallized with 0.1 M Tris pH 7.5 and 3 M sodium formate. Crystals of CC-Hex-Φ_{22} were obtained with 20 mM sodium L-glutamate; 20 mM alanine (racemic); 20 mM glycine; 20 mM lysine hydrochloride (racemic); 20 mM serine (racemic), 50 mM sodium HEPES; 50 mM MOPS (acid) pH 7.5; 20 % ethylene glycol and 10 % PEG 8K. Crystals of CC-Hex-D_{24} were obtained from 1.5 M sodium chloride and 10 % v/v ethanol. CC-Hex-H_{24} crystallized with buffer 30mM sodium fluoride; 30 mM sodium bromide; 30 mM sodium iodide, 50 mM sodium HEPES; 50 mM MOPS (acid) pH 7.5, 20 % PEG-MME 550 and 10 % PEG 20K. The CC-Hex-$D_{24}Y_{15}$ and CC-Hex-H_{24} 1:1 mixture formed crystals with 0.2 M sodium chloride, 0.1 M sodium/potassium phosphate pH 6.5, and 25% w/v PEG 1000.

Prior to data collection, for cryo-protection, crystals were soaked in the respective crystallization solution supplemented with 30 - 40% glycerol. X-ray diffraction data were collected at the Diamond Light Source on stations IO2 and IO4 using radiation of variable wavelength from 0.95 Å to 1.7 Å. Data from CC-Tet-Φ_{22} and CC-Hex-Φ_{22} crystals were processed with MOSFLM and SCALA implemented as part of the CCP4 suite of programs[43]. Data for the other structures were processed with HKL2000.

The structures of CC-Tet-Φ_{22} and CC-Hex-Φ_{22} were solved with SHELX[44] by experimental phasing using single anomalous diffraction from the iodine atoms of the iodophenylalanine side chains. For CC-Tet-Φ_{22} six α-helices could be identified in the resulting electron density map, whilst for CC-Hex-Φ_{22} 3 α-helices were immediately clear. The models were automatically built using ARP/wARP[45]. The final refined structure was obtained by subsequent iterative model building with the program COOT[46], and refinement with REFMAC[47] and Phenix[48].

Thereafter, the non-iodinated structures of CC-Tet, CC-Hex-H_{24} and CC-Hex-D_{24} were solved by molecular replacement using as model the parent tetramer and hexamer as a search model, but with the iodophenyl groups substituted by tyrosine. The structure of CC-Hex-$D_{24}Y_{15}$ - CC-Hex-H_{24} 1:1 mixture was solved using the CC-Hex-D_{24} hexamer structure. Although these peptides had crystallized in different space groups, unambiguous molecular replacement solutions were found for each data set using the program PHASER[49]. The qualities of the resulting models were assessed using the CCP4 software suite. Ramachandran plots indicated that no residues fell outside the allowed regions of backbone conformational space.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank members of the Woolfson group for discussions, in particular Mr Craig Armstrong and Dr Jordan Fletcher; Prof Hagan Bayley, Prof Andrei Lupas and Dr Jim Spencer for comments on the manuscript; and we are grateful to the BBSRC of the UK for grants to DNW & PJB (BB/G008833/1) and to RLB & DNW (BB/F007256/1).

REFERENCES

1. Liu XS, Fan K, Wang W. The number of protein folds and their distribution over families in nature. Proteins. 2004; 54:491–499. [PubMed: 14747997]
2. Levitt M. Growth of novel protein structural data. Proc. Natl. Acad. Sci. U. S. A. 2007; 104:3183–3188. [PubMed: 17360626]
3. Grainger B, Sadowski MI, Taylor WR. Re-Evaluating the “Rules” of Protein Topology. J. Comput. Biol. 2010; 17:1253–1266.
4. Taylor WR, Chelliah V, Hollup SM, MacDonald JT, Jonassen I. Probing the “Dark Matter” of Protein Fold Space. Structure. 2009; 17:1244–1252. [PubMed: 19748345]
5. Kuhlman B, et al. Design of a novel globular protein fold with atomic-level accuracy. Science. 2003; 302:1364–1368. [PubMed: 14631033]
6. MacDonald JT, Maksimiak K, Sadowski MI, Taylor WR. De novo backbone scaffolds for protein design. Proteins. 2010; 78:1311–1325. [PubMed: 20017215]
7. Mason JM, Arndt KM. Coiled coil domains: Stability, specificity, and biological implications. Chembiochem. 2004; 5:170–176. [PubMed: 14760737]
8. Parry DAD, Fraser RDB, Squire JM. Fifty years of coiled-coils and alpha-helical bundles: A close relationship between sequence and structure. J. Struct. Biol. 2008; 163:258–269. [PubMed: 18342539]
9. Crick FHC. The Packing of Alpha-Helices - Simple Coiled-Coils. Acta Crystallogr. 1953; 6:689–697.
10. Walshaw J, Woolfson DN. SOCKET: A program for identifying and analysing coiled-coil motifs within protein structures. J. Mol. Biol. 2001; 307:1427–1450. [PubMed: 11292353]
11. Woolfson DN. The design of coiled-coil structures and assemblies. Adv. Prot. Chem. 2005; 70:79–112.
12. Grigoryan G, Keating AE. Structural specificity in coiled-coil interactions. Curr. Opin. Struct. Biol. 2008; 18:477–483. [PubMed: 18555680]
13. Harbury PB, Plecs JJ, Tidor B, Alber T, Kim PS. High-resolution protein design with backbone freedom. Science. 1998; 282:1462–1467. [PubMed: 9822371]
14. Nautiyal S, Woolfson DN, King DS, Alber T. A Designed Heterotrimeric Coiled-Coil. Biochemistry. 1995; 34:11645–11651. [PubMed: 7547896]
15. Grigoryan G, Reinke AW, Keating AE. Design of protein-interaction specificity gives selective bZIP-binding peptides. Nature. 2009; 458:859–U2. [PubMed: 19370028]
16. Grigoryan G, DeGrado WF. Probing Designability via a Generalized Model of Helical Bundle Geometry. J. Mol. Biol. 2011; 405:1079–1100. [PubMed: 20932976]
17. Grigoryan G, et al. Computational Design of Virus-Like Protein Assemblies on Carbon Nanotube Surfaces. Science. 2011; 332:1071–1076. [PubMed: 21617073]
18. Lupas AN, Gruber M. The structure of alpha-helical coiled coils. 2005; 70:37–78.
19. Moutevelis E, Woolfson DN. A Periodic Table of Coiled-Coil Protein Structures. J. Mol. Biol. 2009; 385:726–732. [PubMed: 19059267]
20. Liu J, et al. A seven-helix coiled coil. Proc. Natl. Acad. Sci. U. S. A. 2006; 103:15457–15462. [PubMed: 17030805]
21. Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature. 2000; 405:914–919. [PubMed: 10879525]
22. Walshaw J, Woolfson DN. Open-and-shut cases in coiled-coil assembly: alpha-sheets and alpha-cylinders. Protein Sci. 2001; 10:668–673. [PubMed: 11344336]

23. Walshaw J, Woolfson DN. Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. J. Struct. Biol. 2003; 144:349–361. [PubMed: 14643203]
24. Ghirlanda G, Lear JD, Ogihara NL, Eisenberg D, DeGrado WF. A hierarchic approach to the design of hexameric helical barrels. J. Mol. Biol. 2002; 319:243–253. [PubMed: 12051949]
25. Donald JE, et al. Transmembrane orientation and possible role of the fusogenic peptide from parainfluenza virus 5 (PIV5) in promoting fusion. Proc. Natl. Acad. Sci. U. S. A. 2011; 108:3958–3963. [PubMed: 21321234]
26. North B, Summa CM, Ghirlanda G, DeGrado WF. D-n-symmetrical tertiary templates for the design of tubular proteins. J. Mol. Biol. 2001; 311:1081–1090. [PubMed: 11531341]
27. Testa OD, Moutevelis E, Woolfson DN. CC plus : a relational database of coiled-coil structures. Nucleic Acids Res. 2009; 37:D315–D322. [PubMed: 18842638]
28. Armstrong CT, Vincent TL, Green PJ, Woolfson DN. SCORER 2.0: an algorithm for distinguishing parallel dimeric and trimeric coiled-coil sequences. Bioinformatics. 2011; 27:1908–1914. [PubMed: 21576179]
29. Harbury PB, Zhang T, Kim PS, Alber T. A Switch between 2-Stranded, 3-Stranded and 4-Stranded Coiled Coils in Gcn4 Leucine-Zipper Mutants. Science. 1993; 262:1401–1407. [PubMed: 8248779]
30. O'Shea EK, Lumb KJ, Kim PS. Peptide Velcro - Design of a Heterodimeric Coiled-Coil. Curr. Biol. 1993; 3:658–667. [PubMed: 15335856]
31. Bromley EHC, Sessions RB, Thomson AR, Woolfson DN. Designed alpha-Helical Tectons for Constructing Multicomponent Synthetic Biological Systems. J. Am. Chem. Soc. 2009; 131:928–930. [PubMed: 19115943]
32. Reinke AW, Grant RA, Keating AE. A Synthetic Coiled-Coil Interactome Provides Heterospecific Modules for Molecular Engineering. J. Am. Chem. Soc. 2010; 132:6025–6031. [PubMed: 20387835]
33. Strelkov SV, Burkhard P. Analysis of alpha-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. J. Struct. Biol. 2002; 137:54–64. [PubMed: 12064933]
34. Woolfson DN, Alber T. Predicting Oligomerization States of Coiled Coils. Protein Sci. 1995; 4:1596–1607. [PubMed: 8520486]
35. Berman H, Henrick K, Nakamura H, Markley JL. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data. Nucleic Acids Res. 2007; 35:D301–D303. [PubMed: 17142228]
36. Opella SJ, et al. Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy. Nat. Struct. Biol. 1999; 6:374–379. [PubMed: 10201407]
37. Malashkevich VN, Kammerer RA, Efimov VP, Schulthess T, Engel J. The Crystal Structure of a Five-Stranded Coiled Coil in COMP: A Prototype Ion Channel? Science. 1996; 274:761–765. [PubMed: 8864111]
38. Liu J, Yong W, Deng YQ, Kallenbach NR, Lu M. Atomic structure of a tryptophan-zipper pentamer. Proc. Natl. Acad. Sci. U. S. A. 2004; 101:16156–16161. [PubMed: 15520380]
39. Jeoung JH, Pippig DA, Martins BM, Wagener N, Dobbek H. HTHP: A novel class of hexameric, tyrosine-coordinated heme proteins. J. Mol. Biol. 2007; 368:1122–1131. [PubMed: 17395199]
40. Dong CJ, et al. Wza the translocon for E-coli capsular polysaccharides defines a new class of membrane protein. Nature. 2006; 444:226–229. [PubMed: 17086202]
41. Mueller M, Grauschopf U, Maier T, Glockshuber R, Ban N. The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. Nature. 2009; 459:726–730. [PubMed: 19421192]
42. Demeler, B. UltraScan A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments. In: Scott, DJ.; Harding, SE., editors. Modern Analytical Ultracentrifugation: Techniques and Methods. Royal Society of Chemistry; London: 2005. p. 210-229.
43. Bailey S. The Ccp4 Suite - Programs for Protein Crystallography. Acta Crystallogr. Sect. D-Biol. Crystallogr. 1994; 50:760–763. [PubMed: 15299374]

44. Sheldrick GM. A short history of SHELX. Acta Crystallogr. Sect. A. 2008; 64:112–122. [PubMed: 18156677]
45. Perrakis A, Harkiolaki M, Wilson KS, Lamzin VS. ARP/wARP and molecular replacement. Acta Crystallogr. Sect. D-Biol. Crystallogr. 2001; 57:1445–1450. [PubMed: 11567158]
46. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr. Sect. D-Biol. Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]
47. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. Sect. D-Biol. Crystallogr. 1997; 53:240–255. [PubMed: 15299926]
48. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. Sect. D-Biol. Crystallogr. 2010; 66:213–221. [PubMed: 20124702]
49. McCoy AJ, et al. Phaser crystallographic software. J. Appl. Crystallogr. 2007; 40:658–674. [PubMed: 19461840]

Europe PMC Funders Author Manuscripts

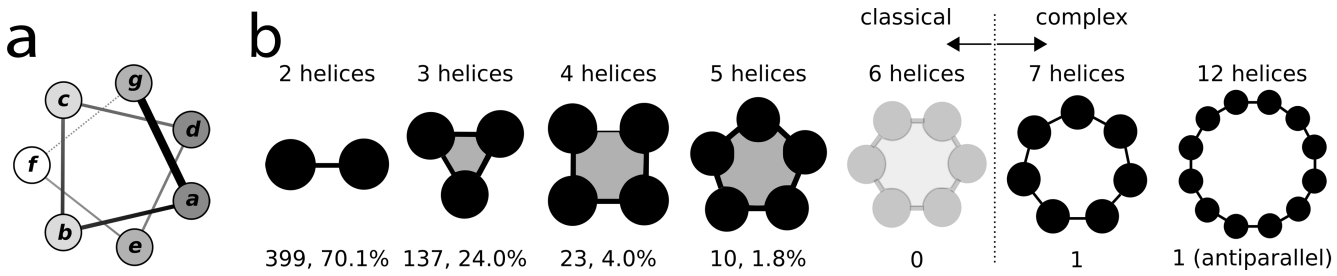

Figure 1. A coiled-coil helical wheel and parallel coiled-coil assemblies

(**a**) Helical wheel showing how the residues of a heptad repeat, ***abcdefg***, are displaced around an α-helix. Residues that fall closer to the helix-helix interfaces are shaded progressively darker. (**b**) Parallel coiled-coil architectures found in the CC+ structural database (http://www.coiledcoils.chm.bris.ac.uk)[19,27]. Circles represent helices; lines, knobs-into-holes interactions; and grey shaded regions, consolidated hydrophobic cores (cyclic rings of KIH interactions) found in *classical* coiled coils. The numbers and percentages are for non-identical structures observed for each class. At present, there are no stand-alone, parallel hexameric coiled coils with a consolidated hydrophobic core. The higher-order architectures shown with 7 and 12 helices have dimer-like interfaces between neighbouring helices and these do not contribute to cycles of KIH interactions; therefore, they are designated *complex* coiled coils.

Europe PMC Funders Author Manuscripts

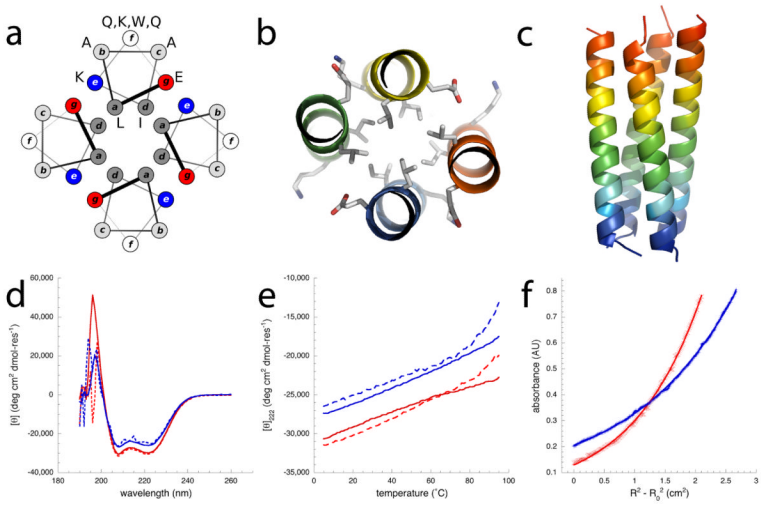

Figure 2. Helicity and oligomerization of CC-Tet and CC-Hex

(**a**) Helical wheels configured as a parallel 4-helix coiled coil. The uppermost wheel carries the sequence used as the basis for the *de novo* coiled coil CC-Tet. (**b**) A 1-heptad slab through the 2.1 Å X-ray crystal structure of CC-Tet. The side chains at positions ***g, a, d*** and ***e*** are shown as sticks. Panels **a** and **b** are juxtaposed to ease comparison. **(c**) Rainbow-coloured ribbon diagram of the CC-Tet structure; blue signifies the *N*-termini. (**d**) Circular dichroism (CD) spectra for CC-Tet (blue lines) and CC-Hex (red). Conditions: 10 μM (broken lines) and 100 μM (solid lines) peptide concentrations, phosphate-buffered saline (PBS), pH 7.4, 20°C. (**e**) Temperature dependence of the CD signal at 222 nm for CC-Tet and CC-Hex. Key: as for panel **d**. (**f**) Representative sedimentation-equilibrium curves from analytical ultracentrifugation for CC-Tet (red) and CC-Hex (blue). The CC-Tet data fitted to a single ideal species of 13,310 Da (95% confidence limits +135 and −91 Da); the CC-Hex data fitted to a single ideal species of 20,319 Da (95% confidence limits +119 and −111 Da); in both cases the monomer molecular mass is 3,375 Da. Conditions: 60 μM peptide concentrations, PBS, pH 7.4, 20°C; rotor speed 30,000 rpm. Panels **b** and **c** were rendered in PyMol (www.pymol.org).

Europe PMC Funders Author Manuscripts

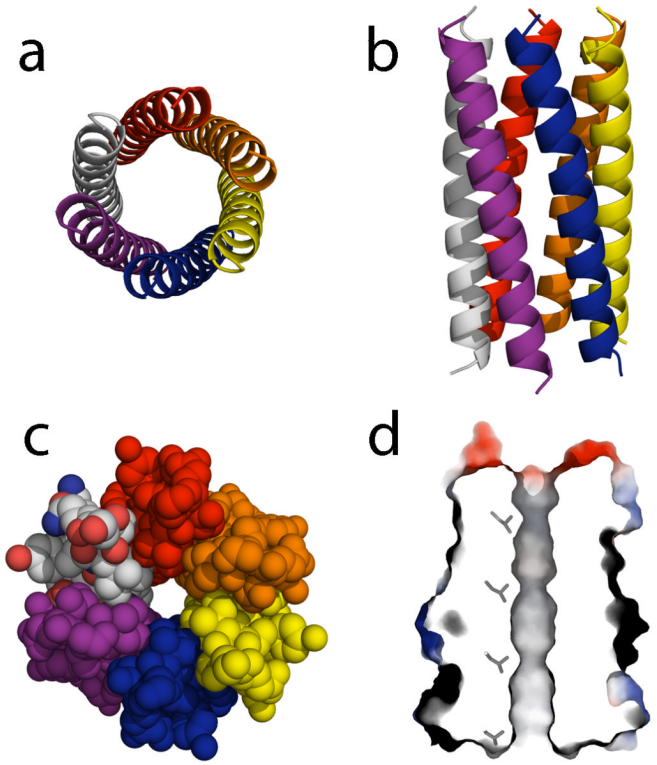

Figure 3. Structure of a parallel coiled-coil hexamer

(**a** and **b**) Orthogonal ribbon diagrams of the 2.2 Å X-ray crystal structure of CC-Hex. (**c**) The same view as in **a** with the assembly represented by space-filling models to reveal the central channel. (**d**) Similar view to **b**, giving a slice through the electrostatic surface, and with the Leu side chains at the ***a*** sites of one protomer rendered as sticks. In **a** and **c** the structures are viewed from their *C*-termini; in **b** and **d**, they are viewed with the *C*-termini at the top; and in **a – c** the colouring is by chain. Images created with PyMol (www.pymol.org).

Europe PMC Funders Author Manuscripts

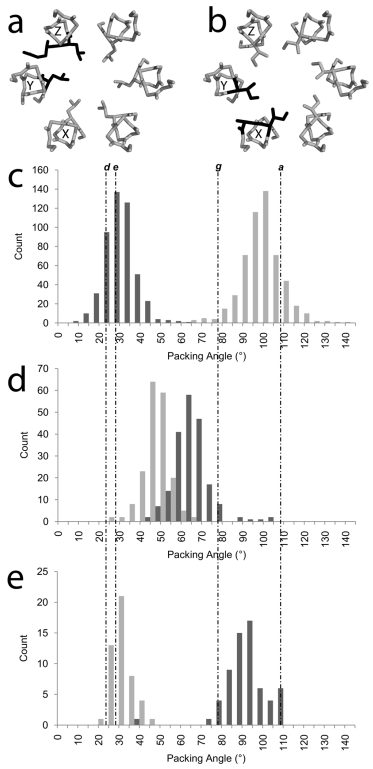

Figure 4. Knobs-Into-Holes (KIH) interactions in the CC-Hex structure

(**a** & **b**) Layers of cyclic KIH interactions. (**a**) Leucine residues at an ***a***-layer, with an $\boldsymbol{a_Y}$-$\boldsymbol{g_Z}$-$\boldsymbol{a_Z}$. (Leu-Glu-Leu), KIH interaction shaded black. (**b**) Isoleucine residues at a ***d***-layer, with a $\boldsymbol{d_Y}$-$\boldsymbol{d_X}$-$\boldsymbol{e_X}$. (Ile-Ala-Ile), KIH interaction shaded black. The structures are viewed from the *C*-terminal ends of the helices. Images created with PyMol (www.pymol.org). (**c – e**) Histograms showing the distributions of side-chain packing angles measured for side chains at ***a*** (darker shading) and ***d*** (lighter shading) of the known parallel dimers (**c**), trimers (**d**) and tetramers (**e**). The statistics for these distributions are: dimer ***a*** sites, number (n) 484, mean (μ) 29.1 and standard deviation (σ) 7.4; dimer ***d*** sites, n = 530, μ = 95.5, σ = 10.0; trimer ***a*** sites, n = 200, μ = 63.8, σ = 8.8; trimer ***d*** sites, n = 185, μ = 44.4, σ = 6.3; tetramer ***a*** sites, n = 63, μ = 90.3, σ = 10.7; tetramer ***d*** sites, n = 48, μ = 27.2, σ = 5.3. These angles were generated by SOCKET[10], and measured as the C_α-C_β bond vector of the knob residues to the C_α-C_α vector of the hole residues on the partnering helix. The average angles for the knob residues at register positions ***g, a, d,*** and ***e*** of CC-Hex are indicated by broken vertical lines.

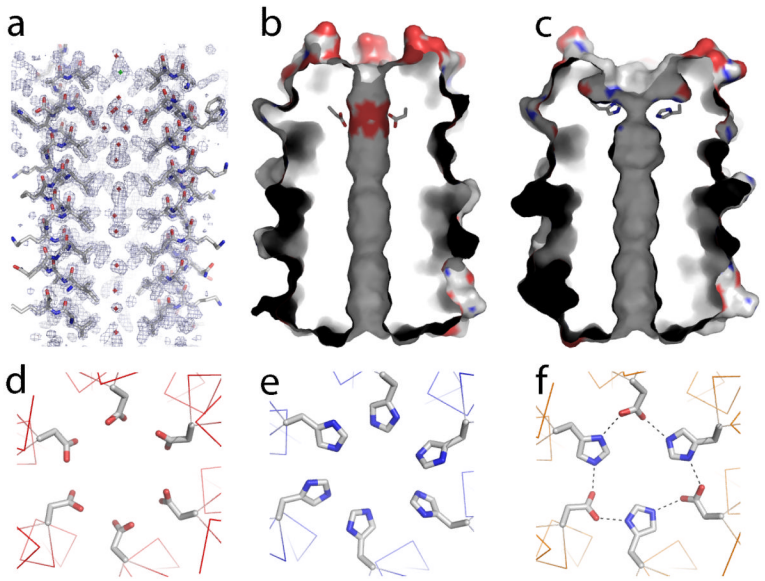

Figure 5. X-ray crystal structures of mutant CC-Hex channels

(**a**) Fitted electron density for the L24D structure showing the water molecules (red crosses) of the pore. The 2Fo - Fc map is rendered at the 2σ level, and the final refined model is shown as sticks for two protomers. (**b & c**) Slices through the vacuum electrostatic surfaces of the L24D and L24H structures. Some of the side chains at position 24 are shown, and aligned with Asp-24 in panel **a** to ease comparisons. (**d – f**) The layers of residues at positions 24 of the X-ray crystal structures for L24D, L24H and L24D:L24H, respectively. Sub-3 Å distances are shown by dotted lines. In **a – c** the *C*-termini of the peptides are topmost; and in **d – f** the helices are viewed from their *C*-terminal ends. Images created with PyMol (www.pymol.org).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Sequences of the *de novo* designed CC-Tet and CC-Hex peptides

Φ= 4-iodophenylalanine. Calculated masses are for the protonated monoisotopic species in each case. *N.b.*, In the main text of manuscript the abbreviations L24D and L24H are used for the mutants more-formally name here as CC-Hex-D_{24} (and CC-Hex-$D_{24}Y_{15}$), and CC-Hex-H_{24}, respectively.

Name	Peptide Sequences and Heptad Register	Mass (Da) calc/obs
	gabcdefgabcdefgabcdefgabcdef	
CC-Tet	AC-GE**L**AA**I**KQE**L**AA**I**KKE**L**AA**I**KWE**L**AA**I**KQGAG-NH_2	3373.9/3373.5
CC-Tet-Φ_{22}	AC-GE**L**AA**I**KQE**L**AA**I**KKE**L**AA**I**KΦE**L**AA**I**KQGAG-NH_2	3460.8/3460.3
CC-Hex	AC-GE**L**KA**I**AQE**L**KA**I**AKE**L**KA**I**AWE**L**KA**I**AQGAG-NH_2	3373.9/3373.5
CC-Hex-Φ_{22}	AC-GE**L**KA**I**AQE**L**KA**I**AKE**L**KA**I**AΦE**L**KA**I**AQGAG-NH_2	3460.8/3460.4
CC-Hex-H_{24}	AC-GE**L**KA**I**AQE**L**KA**I**AKE**L**KA**I**AWE**H**KA**I**AQGAG-NH_2	3397.9/3397.4
CC-Hex-D_{24}	AC-GE**L**KA**I**AQE**L**KA**I**AKE**L**KA**I**AWE**D**KA**I**AQGAGY-NH_2	3538.9/3538.5
CC-Hex-$D_{24}Y_{15}$	AC-GE**L**KA**I**AQE**L**KA**I**AYE**L**KA**I**AKE**D**KA**I**AQG-NH_2	3224.8/3224.3