Journal of Structural Biology 170 (2010) 202–215

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

A novel coiled-coil repeat variant in a class of bacterial cytoskeletal proteins

John Walshaw [a,*], Michael D. Gillespie [b], Gabriella H. Kelemen [b,*]

[a] Department of Computational and Systems Biology, John Innes Centre, Norwich Research Park, Colney, Norwich, NR4 7UH, United Kingdom[1]
[b] University of East Anglia, School of Biological Sciences, Norwich, NR4 7TJ, United Kingdom

ARTICLE INFO

Article history:
Received 5 November 2009
Received in revised form 6 February 2010
Accepted 15 February 2010
Available online 21 February 2010

Keywords:
Bacterial cytoskeleton
Streptomyces
Coiled coil
Sequence repeat
Scy
FilP

ABSTRACT

In recent years, a number of bacterial coiled-coil proteins have been characterised which have roles in cell growth and morphology. Several have been shown to have a cytoskeletal function and some have been proposed to have an IF-like character in particular. We recently demonstrated in *Streptomyces coelicolor* a cytoskeletal role of Scy, a large protein implicated in filamentous growth, whose sequence is dominated by an unusual coiled-coil repeat. We present a detailed analysis of this 51-residue repeat and conclude that it is likely to form a parallel dimeric non-canonical coiled coil based on hendecads but with regions of local underwinding reflecting highly periodic modifications in the sequence. We also demonstrate that traditional sequence similarity searching is insufficient to identify all but the close orthologues of such repeat-dominated proteins, but that by an analysis of repeat periodicity and composition, remote homologues can be found. One clear candidate, despite a great size discrepancy and unremarkable sequence identity, is the known filament-former FilP in the same species. Both proteins appear distinct from the archetypal bacterial IF-like protein; they therefore may constitute a new class of bacterial filamentous protein. The similar sequence characteristics of both suggest their likely oligomer state and a possible mechanism for higher-order assembly into filaments. Another remote homologue in *Actinomyces* was highlighted by this method. Further, a known coiled-coil protein, DivIVA, appears to share some of these sequence characteristics.

1. Introduction

Several classes of filamentous proteins form dynamic, intracellular scaffolding within the cells of prokaryotes and eukaryotes. The microfilament and microtubule components of the cytoskeleton are formed by the polymerisation of globular proteins (Insall and Machesky, 2009; Wade, 2009) which occur in both animals and plants. Homologues of both actins and tubulins have been found in prokaryotes and they are essential for cell division and bacterial shape via the control of cell wall synthesis (Löwe and Amos, 2009; Cabeen and Jacobs-Wagner, 2005, 2007). A third class of cytoskeletal (and nucleoskeletal) protein, the intermediate filament (IF), has a variety of globular head and tail domains, but its filamentous nature results from its long α-helical domains winding around each other to form coiled-coil "rods" which then bundle together (Herrmann and Aebi, 2004; Sokolova et al., 2006). Although α-helical coiled coils are ubiquitous in all kingdoms of life, true IF proteins have been found throughout metazoans but not conclusively in bacteria or plants. A few notable IF-like proteins, which exhibit many, but not all (Herrmann and Aebi, 2004), of the key IF features, have been demonstrated in bacteria, however: CfpA (You et al., 1996), CreS (Ausmees et al., 2003; Charbon et al., 2009), Scc (Mazouni et al., 2006), FilP (Bagchi et al., 2008). In plants there is some evidence of IF-like proteins from antigen cross-reactivity (e.g. Ross et al., 1991; Yu and Moreno Díaz de la Espina, 1999) but classic IFs are yet to be confirmed.

The observation of a lengthy coiled-coil domain is necessary but insufficient to identify an IF protein. Several other classes of filaments made from long coiled-coil proteins have been characterised in many taxa. Amongst these are myosin II bipolar thick filaments in metazoans (Squire et al., 1998) and slime moulds (Bosgraaf and van Haastert, 2006). The families of 2–5 nm fine filaments (Roberts, 1987) include several largely α-helical proteins, with coiled-coil rod domains of several hundred residues. Amongst these are the protofilament-forming tektins of metazoans and *Chlamydomonas* (Norrander et al., 1992); giardin (Holberton et al., 1988) of *Giardia* microribbons; and SF-assemblin of the striated microtubule-associated fibres in apicomplexa and green flagellates (Weber et al.,

Abbreviations: FT, Fourier Transforms; IF, intermediate filament.

* Corresponding authors. Fax: +44 (0)1603450021 (J. Walshaw), +44 (0)1603592250 (G.H. Kelemen).
E-mail addresses: john.walshaw@bbsrc.ac.uk (J. Walshaw), g.kelemen@uea.ac.uk (G.H. Kelemen).

[1] The John Innes Centre is grant-aided from the United Kingdom Biotechnology and Biological Sciences Research Council (BBSRC).

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.02.008

1993). Some long filamentous coiled-coil proteins do not assemble into bundles but form single stalks or spikes expressed on the surface of prokaryote cells (e.g. Fischetti, 1989; Peters et al., 1996) or virus capsids (Olia et al., 2007).

The variety of these fibrous coiled-coil proteins is underlined by the fact that several of them possess non-canonical coiled-coil repeats. Besides the classic 7-residue (heptad) pattern which forms canonical α-helical coiled coils (Crick, 1953), a number of other repeats, with different supercoil geometries, have been characterised (Gruber and Lupas, 2003). All follow the principle that in an ideal repeat, a hydrophobic "seam" (the core) is formed by one amino acid side chain per turn of a supercoiled α-helix; two or more such helices are wound around each other in the coiled coil. Thus the heptad repeat forms two turns of a single supercoiled α-helix, meaning that the periodicity of hydrophobic residues in the sequence is 7 residues/2 turns, i.e. 3.5 residues. The 11/3 hendecad (or undecad) repeat, first characterised in an archaeal protein, tetrabrachion (Peters et al., 1996) thus has a periodicity of 3.667. In bacteria, the coiled-coil stalk of a *Yersinia* surface-expressed adhesin consists of a pentadecad repeat 15/4 (Hoiczyk et al., 2000). The fine filament proteins giardin and SF-assemblin have longer repeats, of 29 residues (29/8; Holberton et al., 1988; Hicks et al., 1997). The repeats found in nature have imperfect hydrophobic-polar patterns. Frequently there are hydrophobic amino acids at non-core positions (and vice versa), leading to weaker additional periods observed, such as 7/3 in heptads and 11/4 in hendecads. Further, local modifications in the repeat, and thus supercoiling, are relatively common in canonical coiled-coils (Brown et al., 1996). Also some, such as IF rods, are segmented and joined by non-helical linker domains (Herrmann and Aebi, 2004; Smith et al., 2002). To complicate matters, there is evidence that some coiled-coils switch from one repeat to another in a relatively short space, changing from a left-handed to right-handed supercoil (Shin et al., 2006; Parry, 2006).

We recently experimentally characterised Scy, a large protein in the filamentous bacterium *Streptomyces coelicolor* where Scy controls filamentous growth (Kelemen et al., submitted for publication). Analysis of the protein sequence strongly suggested that it is dominated by a very long non-canonical coiled coil, based on a novel repeat of 51 residues forming 14 supercoiled turns. Additionally there was evidence for a shorter, canonical coiled coil at the N-terminal end.

The aims of the current work are to use computational analyses to firstly establish the likely structure and size of this unusual repeat sequence, and to investigate how it might relate to Scy's cytoskeletal role. We supplemented this with a consideration of other domains of this protein and experimental binding assays. We also made a comparative analysis with one similar, notable protein in particular (FilP) in the light of previous experimental evidence on the latter (Bagchi et al., 2008), in order to gain insight into a possible assembly mechanism. Further, in the light of an expanding number of potentially cytoskeletal proteins apparent in many new, partially-annotated genomes, we investigate how Scy relates to these; and address the question of how best to identify such long fibrous coiled-coil proteins of unusual repeats. A long-established technique for analyzing long coiled-coil repeat sequences is the Fast Fourier Transform (Fast FT; McLachlan and Stewart, 1976), which calculates the periodicities of hydrophobic amino acids. Also very powerful are profile-derived methods for predicting heptad repeats (Lupas et al., 1991; Delorenzi and Speed, 2002; McDonnell et al., 2006) which although remarkably good at identifying even non-canonical coiled-coils, are unable to assign the register of non-heptad repeats correctly. FT has previously been combined with other methods (Gruber et al., 2005). On the other hand, traditional sequence similarity search techniques (Altschul et al., 1997) can be useful but have problems with low-complexity, very repetitive sequences such as coiled coils, especially very long ones. Aiming to find other examples of Scy-type structure we therefore used a combinatorial approach. First we used sequence similarity; then secondly employed a combination of coiled-coil predictions, two distinct hydrophobic repeat periodicities in tandem and also amino acid composition.

2. Materials and methods

2.1. Sequences and databases

We used the protein sequences of genes *SCO5397* (Scy) and *SCO5396* (FilP) from the StrepDB database (Bentley et al., 2002; http://strepdb.streptomyces.org.uk); the corresponding entries in UniProt (The UniProt Consortium, 2009) are Q9L2C3 and Q9K4B5.

2.2. Sequence similarity searches and pairwise alignments

We used BLASTP of NCBI Blast version 2.2.18 (Altschul et al., 1997), with low-complexity filtering turned off, and composition-based score adjustment turned first turned on, and then off for comparison. Default values of other parameters were used. We searched UniProt release 15.8 and the protein sequences of the Broad Institute's Actinomycetales group database (Fischbach et al., 2008). We used TBLASTN of the same package to search release 100 of the EMBL nucleotide database and the genome sequence of *Streptomyces scabies* (these data were provided by the Pathogen Genome Sequencing group at the Wellcome Trust Sanger Institute and can be obtained from http://www.sanger.ac.uk/Projects/S_scabies/). For local pairwise alignments to assess sequence identity we used WATER of the EMBOSS 6.0.1 package (Rice et al., 2000). Multiple alignment of *Streptomyces* Scy sequences was performed with MUSCLE (Edgar, 2004) and manual editing.

2.3. Sequence analysis, feature predictions and statistics

For coiled-coil prediction we used COILS (Lupas et al., 1991) and MARCOIL (Delorenzi and Speed, 2002). Assignment of the best repeats and registers, compilation of amino acid occupancies and charge pairs were performed with Perl scripts, coupled with manual intervention for the identification of register-shifts. We used the JNet secondary structure prediction method (Cuff et al., 1998), via the JPred web server, and the GLOBPLOT (Linding et al., 2003) tool for predicting disordered regions. Amino acid composition of domains was calculated with the COMPSEQ program of EMBOSS. Within the 11 positions of the hendecad-like units a^1–k^4, we examined pairwise charge frequencies of each position r and $r+x$ ($1 \leq x \leq 11$; this means some of the correlated positions lie in a^0–g^0, but the results for x and 11-x are near-identical).

2.4. Fast Fourier Transforms

We implemented Fourier Transforms (FT) of hydrophobic/polar residue distributions in the Scy and other amino acid sequences using Perl scripts. The approach, in which the input is a string of 1's substituted for hydrophobic residues (A, V, M, I, L, F, Y, W) and 0's for polar residues is essentially that employed by McLachlan and Stewart (McLachlan and Stewart, 1976). The Fourier signals x were calculated for each frequency k, which is measured in residues, where $k = 0, \Delta, 2\Delta, 3\Delta, \ldots N-1-2\Delta, N-1-\Delta, N-1$. N is the length of the sequence in residues. In the simplest form, $\Delta = 1$, meaning that the number of frequencies used is equal to the number of residues in the sequence. However, to obtain a higher-resolution plot, we used $\Delta = N/10{,}000$, providing 10,000 frequencies.

Whole-sequence FT were performed with scripts. The period is equal to the sequence length divided by the frequency.

2.5. Calculation of the structural parameters of the putative coiled coil

We used the formula of Fraser and MacRae (Fraser and MacRae, 1973) to calculate the pitch (P) of the Scy long coiled-coil domain, i.e. the distance over which the supercoil completes one full turn. This also enables calculation of the pitch angle (α) between the coiled-coil axis and each α-helix axis, the rise per residue along the supercoil axis and thus the length (l) of the coiled coil. These parameters are also determined by a, the number of residues per turn in an α-helix; h, the axial rise per amino acid in an α-helix; p the periodicity of hydrophobic residues; r_0 the radius of the supercoil; and N the number of amino acid residues in each helix of the coiled coil. The length $l = Nh\cos\alpha$. We used $p = 51/14$ and $N = 1227$ residues in each case. The pitch is very sensitive to the number of residues per turn in the α-helix (Pauling and Corey, 1953), and increases, as does l, with a and h. We used different combinations of a, h and r_0 which occur in several structures of GCN4 leucine zipper-based peptides. Respectively these values are as follows: in the dimeric GCN4-p1 structure, 3.62, 1.51 Å, 4.9 Å (O'Shea et al., 1991); trimeric GCN4-pII, 3.60, 1.53 Å, 6.7 Å (Harbury et al., 1994); tetrameric GCN4-pLI, 3.59, 1.52 Å, 7.6 Å (Harbury et al., 1993). Additionally, parameters from the Asn16abu trimer are 3.63, 1.51 Å, 6.5 Å; those from the structure of the dimeric form are almost identical to GCN4-pII (Gonzalez et al., 1996a). We also used 3.65, 1.5 Å, 7.44 Å, inspired by the RHCC structure from tetrabrachion (Stetefeld et al., 2000). The pitch angle α decreases as a or h increase.

2.6. Hydrophobic moments

We calculated hydrophobic moments only for those R residues proposed to belong to the hendecad units, i.e. residues 8–51 of the penindaenad repeat. The x and y components of the hydrophobic moments are given by:

$$x = \frac{1}{R}\sum_r \phi_r \cos[2\pi\ n(i-1)/L], \quad y = \frac{1}{R}\sum_r \phi_r \sin[2\pi\ n(i-1)/L]$$

where r is each residue used, i is the position in the repeat, L is the repeat length, n is the number of turns of the α-helix spanned by one repeat and φ_r is the hydrophobicity measure for residue r. The magnitude of the hydrophobic moment is then $\sqrt{(x^2 + y^2)}$, and the angular direction in radians is $\tan^{-1}(y/x)$. In each case, we calculated moments from two different hydrophobicity scales (Kyte and Doolittle, 1982; Sweet and Eisenberg, 1983). Assuming the uniformly supercoiled arrangement, i ranges from 8 to 51, $L = 51$ and $n = 14$. In the variably supercoiled model, in which the 11-mer units supercoil as normal hendecads, i ranges from 1 to 11, $L = 11$ and $n = 3$.

2.7. Coiled coil/FT/composition-based search for protein sequences with similar domains

To avoid the problems associated with traditional similarity searches of proteins of low-complexity sequence repeats, we employed the following protocol. Since hydrophobic/polar (HP) repeat periods of between ~3.4 and 3.7 do not necessarily correspond to coiled coils in all cases, firstly we screened UniProt release 15.8 with MARCOIL, retaining only those sequences with any regions whose coiled-coil prediction probability was $\geqslant 0.5$. We then screened these hits with an in-house C-program (unpublished) which calculates the strongest periodicity in every consecutive window of w residues. This is similar in principle to the WINFT component of the repeat-identifying web-based tool REPPER (Gruber et al., 2005), but permits targeting of a particular periodicity p; within each group of s consecutive windows, the mean of the strongest periodicity must lie within $\pm r$ of p, with a standard deviation of $\leqslant\sigma$. To pass to the next stage, a sequence had to test positive in two runs, which correspond to the presence of both a potentially short heptad-based domain and a longer penindaenad repeat: (i) $p = 3.5$, $r = 0.01$, $\sigma = 0.5$, $w = 41$, $s = 40$; (ii) $p = 3.643$, $r = 0.001$, $\sigma = 0.25$, $w = 129$, $s = 65$. Thirdly, the hit regions of the remaining sequences were combined with the results of a screen to find all regions of $\geqslant 100$ consecutive non-Pro residues, where all such regions were $\leqslant 5\%$ Gly. This effectively trimmed, split or eliminated the target-period regions thus found. Sequences were then ranked in descending order of the proportion of their total length which fall in the resulting 3.643-period segments. Finally an extra constraint was imposed such that sequences required at least one further, separate, Pro-free segment (and thus at least one Pro) of $\geqslant 40$ residues and $\leqslant 5\%$ Gly, corresponding to a shorter coiled coil. In order to ascertain that expected known sequences were found, redundancy with regard to homologues was retained. Finally, the coiled-coil predictions (COILS) for the top hits were inspected manually to ensure that most of the 3.5 and 3.643 period regions were thus covered.

2.8. Bacterial strains

All bacterial strains, plasmids and oligonucleotides used in this work are listed in Table 1, Table 2 and Table 3, respectively. The BACTH Bacterial Two Hybrid system (Euromedex) was used according to the manufacturer's instructions.

2.9. Plasmid constructions

The pUT18c-ScyC construct was generated by moving the fragment encoding the hinge and the long C-terminal domain of Scy as a *Kpn*I-*Eco*RI fragment of pK48 (Kelemen et al., submitted for publication) into pUT18c. In order to have T18 and ScyC in frame, we inserted the annealed linkers THScy_KpnF and THScy_KpnR into the *Kpn*I site generating the final pUT18c-ScyC plasmid. From this construct *scyC* was moved as an *Xba*I and *Eco*RI fragment into pKT25 generating pKT25-ScyC. To create pUT18c-ScyN we moved the *scyN* fragment generated by PCR with primers THScy_F1a and Scy_Kpn followed by restriction digest with *Xba*I and *Kpn*I into pUT18c. This insert was also introduced to pKT25 to generate pKT25-ScyN. pUT18c-Scy carrying the entire *scy* gene was created by moving the fragment encoding the hinge and the long C-terminal domain of Scy as a *Kpn*I-*Eco*RI fragment of pK48 (Kelemen et al., submitted for publication) into pUT18c-ScyN. From pUT18c-Scy we moved the *Xba*I-*Eco*RI fragment carrying the entire *scy* gene into pKT25 to produce pKT25-Scy. Constructs pUT18c-ScyC and pUT18c-ScyN were confirmed by sequencing.

3. Results and discussion

3.1. Overall domain structure of Scy

We had previously obtained a COILS prediction (Kelemen et al., submitted for publication) of the Scy sequence; over 1000 residues had a probability (P) of >0.9 and for more than half of these $P = 1.0$. A second analysis using COILS with the weighted matrix confirmed that these should not be treated as false positives. A plot of the prediction is available as supplementary material (Fig. S1). The only notable sustained drop in P-value is at residues 910–1000, where it is mostly around 0.5.

Table 1
Bacterial strains used.

E. coli	Description	Source
DH5α	Cloning	Invitrogen
BTH101	*cya*- strain for screening using the bacterial two hybrid system	Euromedex

Table 2
Plasmid constructs.

Plasmid	Genotype or description	Reference or source
pUT18c	Vector carrying the T18 fragment for the bacterial two hybrid system	Euromedex
pKT25	Vector carrying the T25 fragment for the bacterial two hybrid system	Euromedex
pUT18c-Zip	T18 translational fusion to the leucine zipper of GCN4	Karimova et al. (1998)
pKT25-Zip	T25 translational fusion to the leucine zipper of GCN4	Karimova et al. (1998)
pUT18c-Scy	T18 translational fusion to Scy	This work
pUT18c-ScyC	T18 translational fusion to the C-terminal domain (66aa-1326aa) of Scy	This work
pUT18c-ScyN	T18 translational fusion to the N-terminal domain (1aa-73aa) of Scy	This work
pKT25-Scy	T25 translational fusion to Scy	This work
pKT25-ScyC	T25 translational fusion to the C-terminal domain (66aa-1326aa) of Scy	This work
pKT25-ScyN	T25 translational fusion to the N-terminal domain (1aa-73aa) of Scy	This work

Table 3
Oligonucleotide primers used.

Oligonucleotide	Sequence
THScy_KpnF	CGTTCGACGGCGGCGACATCGCGTAC
THScy_KpnR	GCGATGTCGCCGCCGTCGAACGGTAC
THScy_F1a	GATCATCTAGAGGTGCGGGGCTACGAGAGCCAG
Scy_Kpn	CGCGTTGCGCAGCAACTGCTCG

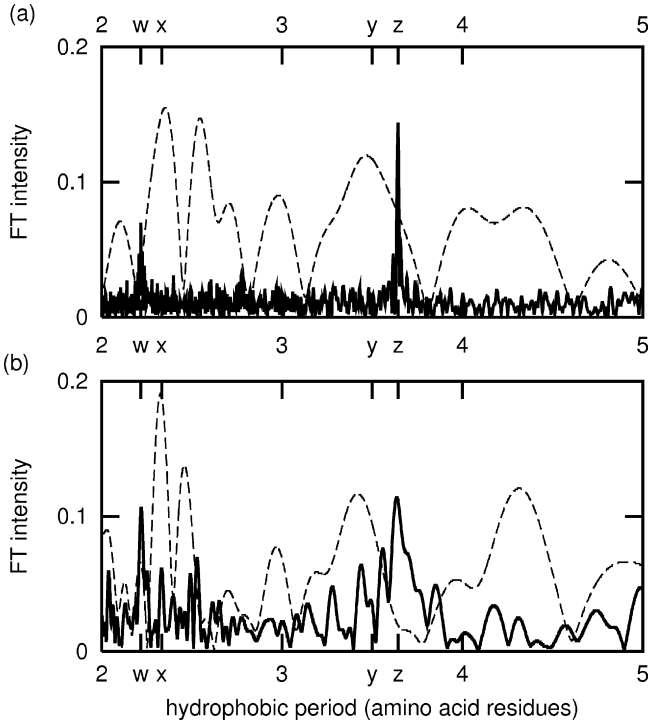

Fig. 1. FT plots of hydrophobic amino acids in (a) Scy and (b) FilP sequences. The FT of the CC7 (light broken line) and CC51 (bold unbroken line) domains are shown separately. The letters indicate 4 particular periods marked on the *x*-axis: $w = 51/23$; $x = 7/3$; $y = 7/2$; $z = 51/14$.

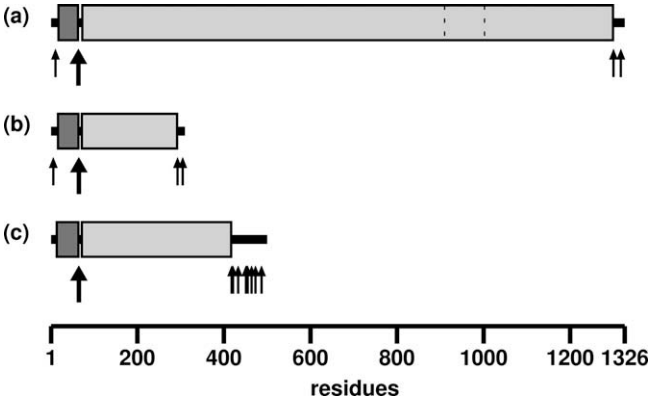

Fig. 2. Domain architecture of (a) Scy, (b) FilP and (c) *Actinomyces urogenitalis* low-complexity hydrophilic protein. CC7 and CC51 domains are shown as dark and light blocks respectively. All proline residues in the sequences are indicated by arrows; the large arrows point to the proline which begins the short flexible domain (hinge) connecting CC7 and CC51. (a) also shows the region in Scy with lower coiled-coil probability, between dashed lines.

We noticed a difference in composition of two domains either side of a short glycine-rich region beginning with a proline (residues 64–72). The entire Scy sequence contains only 4 prolines; two of these at positions 11 and 64 bracket the first domain, which we hereafter refer to as CC7, whilst Pro64 and Pro1300 delimit the second coiled-coil domain, referred to as CC51. CC7 contains an uninterrupted heptad repeat (***a***bc***d***efg register) as predicted by COILS, with hydrophobic positions at the first (*a*) and fourth (*d*) positions; whereas there are many approximately equally-spaced jumps in the predicted register of the much longer CC51, despite a consistently high *P*-value. This is indicative of a non-canonical coiled-coil repeat, and Fast Fourier Transforms (FTs) of the two domains showed peaks at distinct periods (Fig. 1a). The CC7 FT is very noisy as expected due to this domain's short length, but broad peaks occur at 3.467, just short of the heptad periodicity of 7/2, and also at 2.353 (~7/3) and 2.102 (21/10). In contrast, CC51 has an extremely sharp and high peak at 3.643, which is 51/14. Fig. 2a shows the overall domain organisation.

3.2. Canonical six-heptad coiled-coil domain CC7

From Leu18 onwards, there are six and a half heptads in which the seven *a* positions are occupied by Leu, Met, Arg, Ala, Val, Leu and Ile, and the six *d* residues are Phe, Leu, Ala, Leu, Leu and Ala. These amino acids are compatible with a parallel homo-dimeric or trimeric assembly (Harbury et al., 1993; Lupas, 1996). A parallel tetramer seems much less likely, because it would have perpendicular core-packing interactions at the *a* layers, which disfavour the β-branched Val and Ile sidechains. Numerous classic dimers have a tendency for β-branched residues at *a* and Leu at *d*. The mostly polar and charged sidechains at *e* and *g* positions also are consistent with a parallel assembly of this motif. In such an arrangement, interactions are usually between each *g* sidechain of one helix and the *e* sidechain in the following heptad on a partner helix. Possible g_n to e_{n+1} pairings are His17-Glu22, Glu24-Lys29 and the polar-charge interaction Gln45-Arg50 (where *n* represents the heptad). An antiparallel arrangement would produce a similar number of oppositely-charged pairs, but also more like-charged

pairs. Overall, this domain has a small net positive charge. The amino acid composition is relevant to a comparison with other domains. For each of alanine, glutamate and arginine, the content of CC7 is 13%; for leucine it is 10%.

3.3. A putative hinge-region connects the CC7 and CC51 coiled-coil domains

The GLOBPLOT program predicts a possible 'hinge' region (Fig. 2a) of 8 disordered residues from positions 64–71; Gly72 completes a Gly-rich sequence clearly unlikely to be α-helical (PAFDGGDIG). Apart from short regions in the N- and C-terminal domains, GLOBPLOT predicts the whole of the Scy sequence to be structured. This is expected as the algorithm is based on secondary-structure preferences. The 26-residue portion following the proline at position 1300 is rich in glycine, and has a second proline at 1317. Unsurprisingly, predictions by JNET and GLOBPLOT indicate an almost complete lack of regular secondary or tertiary structure in this part.

3.4. Non-canonical 1227-residue coiled-coil domain CC51

The long central domain from residues 73 to 1299 is notable for a number of reasons. It is devoid of proline, has only 5 aromatic residues (two of which are relatively close to the ends – Tyr73, Phe1297) and 22 glycines (<2%). It is very rich in alanine (23%) and charged residues: 19% glutamate and 14% arginine (but only 4% aspartate and 3% lysine). This composition is somewhat similar to the much shorter, heptad-based, surface-expressed streptococcal M-proteins (Sumby et al., 2005; Fischetti, 1989), although the latter contain more leucine (12% compared to Scy's 7%).

The large, sharp peak at a periodicity of 51/14 in the FT represents 14 hydrophobic positions in a motif of 51 residues. Its magnitude is many times greater than any other peaks which could correspond to a signature coiled-coil periodicity (i.e. in the range of approximately 3.4–3.7 residues). A possible cause of this periodicity is a 51-residue repeat ("penindaenad") exhibiting on average 14 equally spaced hydrophobic residues, that is, 51 residues making 14 turns of a supercoiled α-helix. A 51-residue coiled-coil repeat is expected to have spacings between the hydrophobic residues of 3, 4, (3, 4, 4) × 4 positions, which we also refer to as a "7, 11, 11, 11, 11 repeat", made of one "heptad" unit and four "hendecad" units. The only other peak of note is considerably lower, at a period of 2.217, which is 51/23. This is essentially equivalent to the 7/3 period observed in FTs of canonical coiled coils; it represents 3 hydrophobic residues per 7-residue unit and 5 in each of the four 11-residue units. An examination of the FT signal in a 'sliding window' of 129 residues, indicating local hydrophobic repeats versus the position in the sequence, reveals a remarkably consistent intensity peak along most of the Scy protein, centred on a periodicity of 3.643 (Fig. 3a). Given that the number of residues per turn in a coiled coil α-helix is often ~3.6, this would result in a right-handed coiled coil. Interestingly, an occurrence of an 11/4 period, supporting a locally more hendecad like repeat (see Section 1), coincides with the ~90 residue $P = 0.5$ region identified by COILS at around residues 910–1000 (Fig. S1).

3.5. A parallel, in-register orientation is most likely for Scy coiled coils

Bacterial two hybrid analysis (Karimova et al., 1998) was used to demonstrate interactions between Scy monomers and truncated versions, ScyN and ScyC. ScyN contains the N-terminal coiled coil (CC7) and the hinge region of Scy whilst ScyC contains the hinge and the long penindaenad repeat (CC51). T18 and T25, the two complementary fragments of the catalytic domain of *Bordatella pertussis* adenylate cyclase were fused to the N-terminus of Scy, ScyN and ScyC. Protein–protein interaction was established when the two complementary fragments T18 and T25 restored the adenylate cyclase function when screened in a cAMP dependent β-galactosidase assay (Karimova et al., 1998). All homo-pair combinations (T18Scy-T25Scy; T18ScyC-T25ScyC; T18ScyN-T25ScyN) and the partial homo-pair T18Scy-T25ScyN gave positive interactions whereas the hetero-pair (T18ScyC-T25ScyN) did not (Fig. 4). As in all cases the T18 and T25 fragments were fused to the N-terminus of the constructs generated, the positive interaction suggests parallel organisation of the coiled coils; although higher-order assemblies (e.g. a dimer of dimeric coiled coils) could, in principle, generate interactions between two N-terminus fragments even in an antiparallel orientation. However, it is unlikely that both coiled-coil domains, on their own, are capable of forming higher-order assemblies (see section on Scy and FilP sequences). The fact that the short N-terminal domain still gave positive interactions when tested against either itself or the full length Scy; and the C-terminal domain gave positive interactions when tested against itself, suggests that parallel orientation is most likely.

3.6. Register of the 51-residue repeat

When assigning the continuous 7, 11, 11, 11, 11 repeat to CC51, the first and fourth positions of the groups of 7, and the first, fourth and eighth of the groups of 11, are ideally hydrophobic (representing the "core" of the putative coiled coil) whilst the remainder are ideally hydrophilic. In the 7- and four 11-residue units we denote the repeat positions as a^0–g^0,a^1–k^1,a^2–k^2,a^3–k^3,a^4–k^4. One register results in 241 hydrophobic positions out of a possible 337 (72%)

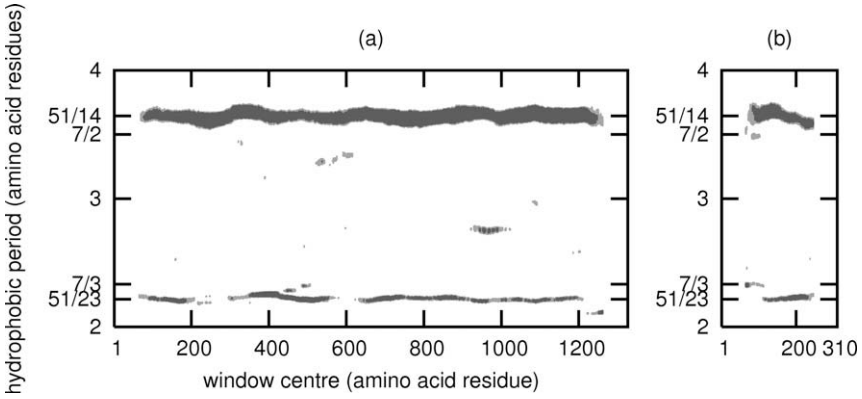

Fig. 3. Contours of FT intensities of hydrophobic residue period (*y*-axis) versus centre of a 129-residue sliding window (*x*-axis) for (a) Scy and (b) FilP. Periods where the intensities are greater than 2.5× (light shade) and 3.0× (dark) the mean intensity are shown.

Fig. 4. Bacterial Two Hybrid analysis of interactions between the different Scy domains. The appropriate pUT18c and pKT25 derivatives were co-transformed into BTH101 and the co-transformants were screened in LB medium containing 0.5 mM IPTG and 40 μg/ml Xgal after 48hours incubation at 30 °C. The pUT18c and pKT25 vectors were used as a negative control, whilst the pUT18c-Zip and pKT25-Zip plasmid-pair was used as a positive control (Karimova et al., 1998). T18 fragments are shown in green and T25 fragments shown in yellow. The colour coding for Scy domains are: red, N-terminal domain; black, hinge; blue, C-terminal domain.

and 671 of 890 possible hydrophilic positions (75%) matching a perfect repeat. The 51-residue repeat far out-performed other known and speculative coiled-coil repeats which we assigned (Table 4).

Although only 53 of 337 ideally hydrophobic positions are occupied by charged residues, which does not compare especially unfavourably with some known coiled coils, inspection of the continuous register revealed that these tended to cluster in several groups of consecutive core positions (data not shown). Although individual charged and polar residues in core positions are known to confer oligomer state preferences in left-handed coiled coils (Harbury et al., 1993; Gonzalez et al., 1996b), the zonal character of the core-charge indicated that the register we assigned is likely to be systematically wrong in these regions, and might be corrected by a small number of interruptions.

We therefore modified the original register, to include a few insertions or deletions with respect to the 51-residue repeat. Although FT calculates not only the magnitude but also the phase of any period, there was insufficient signal from phase-jumps to indicate such shifts. We therefore used a computerised approach to finding the best local register encompassing the above preferences, complemented by manual inspection, resulting in the six register-shifts producing the segmentation shown in Table 5; the whole register is shown in Fig. 5.

Although two of the uninterrupted segments are rather short – number three in particular has only one '7' unit – four others each have at least three entire 51-residue repeats and all have a peak in the Fourier Transform-derived periodicity at or near the signature. The 3.643 peak is distinct from the 3.667 hydrophobic period of a continuous hendecad; however some of the segments do have local regions within them where the peak is shifted towards the hendecad signature periodicity or slightly in the opposite direction (Fig. 3a).

We note that for any sequence or sub-sequence, several registers will give an almost equally good (or bad) match. In particular, in two registers offset by $11x + y$ where x is an integer <5 and $y \in \{3, 4, 7\}$, many of the hydrophobic positions will coincide. Therefore, the precise number of register-shifts introduced and the offsets used are somewhat arbitrary, but we believe the register shown depicts the likely core, or exposed, nature of most of the residues in this putative coiled coil. We also note that two of these shifts restore the registers of earlier segments (Table 5), which explains why even a single uninterrupted penindaenad produced a very good match.

The amino acid frequencies at each of the 51 positions are available in the supplementary material (Table S1) and the most common amino acids are shown in Fig. 6a. It is clear that the preferences for corresponding positions of the four '11' units (e.g. a^1, a^2, a^3, a^4) are very similar, but there are cases where the most common amino acid differs. Frequencies of hydrophobic, polar and charged residues at all 51 positions are available in the supplementary material (Table S2); Table S3 shows these data where a^{1-4}, b^{1-4}, etc. have been averaged over the four hendecad units. About 90% of the proposed a and h positions are hydrophobic, as are 60% of d^1–d^4,70% of the d^0 and 40% of the e^1–e^4 positions. 70–90% of the putative non-core positions are occupied by hydrophilic residues. As shown by Table 6, hendecad amino acid types are broadly

Table 5
Segmentation of the Scy CC51 domain resulting from the 6 shifts in the register of the 51-residue repeat. An offset of 0 represents position a^0, 1 is b^0, etc.

First and last residues of segment		Offset (register) of first residue of segment		Offset with respect to residue 73
73	256	5	(f^0)	5
257	358	40	(a^4)	9
359	428	32	(d^3)	1
429	715	37	(i^3)	38
716	846	40	(a^4)	9
847	1047	0	(a^0)	42
1048	1299	11	(e^1)	5

Table 4
Highest possible frequencies of residues in the Scy CC51 sequence that match hydrophobic or polar positions in perfect repeats. H is the proportion of all the putative 'core' positions which are hydrophobic (F, A, M, I, L, Y or V). P is the proportion of all the *non*-core positions which are polar. C′ is the proportion of all the core positions which are *not* charged (D, E, K, R or H). All possible registers were evaluated for each repeat, to find the best value of H. Repeats are described as comprising units of 7, 10 or 11 residues where the hydrophobic spacings are respectively 3, 4 (*abcdefg*); 3, 4, 3 (*abcdefghij*); and 3, 4, 4 (*abcdefghijk*). These units are numbered 0, 1, 2 etc. and this number appears in superscript beside the best registers found.

Repeat	Period		Best offset (register)	H	P	C′
7, 10, 7	3.429	(24/7)	7 (a^1)	0.42	0.64	0.60
7	3.500	(7/2)	6 (g)	0.40	0.63	0.60
7, 11, 7	3.571	(25/7)	9 (c^1)	0.40	0.64	0.59
7, 11	3.600	(18/5)	3 (d^0)	0.43	0.65	0.61
11, 7, 11	3.625	(29/8)	3 (d^0)	0.41	0.64	0.63
7, 11, 11, 11	3.636	(40/11)	26 (i^2)	0.54	0.69	0.68
7, 11, 11, 11, 11	3.643	(51/14)	27 (j^2)	0.72	0.75	0.84
7, 11, 11, 11, 11, 11	3.647	(62/17)	46 (g^4)	0.59	0.71	0.75
11	3.667	(11/3)	4 (e)	0.46	0.66	0.67

Fig. 5. A 51-residue repeat register assigned to the sequence of the Scy CC51 domain. The top two lines indicate the repeat positions a^0 to k^4. Predominantly hydrophobic positions are highlighted in grey. The e^1–e^4 positions are about 40% hydrophobic.

similar to a 4-stranded right-handed coiled coil (Peters et al., 1996), whereas specific amino acid preferences are very different. In contrast the amino acid consensus is in most positions strikingly similar to the "second stalk" (*b* subunit) of bacterial F_0F_1-ATPase (Del Rizzo et al., 2002). Although the sequences of the *b* subunit coiled coil appear to match an unbroken hendecad whilst heptads can only be assigned with several interruptions, its handedness is controversial; there is experimental evidence consistent with both the right (Del Rizzo et al., 2002, 2006; Bi et al., 2008) and left (Wise and Vogel, 2008; Hornung et al., 2008). Compared to tetrameric hendecad coiled coils, lower-order oligomer arrangements of this repeat would be expected to have only one, not both, of the *d* and *e* positions tending to a hydrophobic character (Strelkov, personal communication), which appears the case with Scy and the *b* subunit. Interestingly however, in contrast to Scy the *b* subunit is strongly apolar at the *e*, not *d* position, which implies there is a difference in the packing arrangement of these layers. There is one cysteine residue, Cys1190, at an a^0 position which might afford covalent cross-linking, but this is not conserved in all species.

3.7. Charge properties of CC51

Although charge density of CC51 is very high, the gross feature of the domain is that the central region is the most negatively charged whilst the extremities have a mild positive charge (Fig. 7). We also performed a Fourier Transform of the periodicity of charged residues and found only small peaks (data not shown), which are at the same locations as the hydrophobic analysis, i.e. at 3.643, and at the reciprocal periodicity 1.378 (51/(51–14)). These peaks result from the negative charge, as shown by the FT of the periodicity of such residues alone. The highest peak of positive charge is low in comparison.

We examined correlations of oppositely-charged residues at *r*, *r* + *x* that would be compatible with a hendecad-based coiled-coil structure. In this arrangement, a residue at position *r* is close on the helix surface to a residue at *r* + 4, and to a lesser extent *r* + 3 (Fig. 6b). These two sequence separations represent potential intra-helical interactions. The former is represented by interactions between, for example, *c*–*g* in the same hendecad; or between *j* in one hendecad and *c* at the next (denoted *j*–*c*′), as well as other pairs (*f*–*j*, *i*–*b*′). The *r* + 3 pairing is represented by amongst others *k*–*c*′ and *b*–*e*. In contrast a pair of residues at *r*, *r* + 2 or *r*, *r* + 5 in the sequence are on opposite sides of an α-helix, but *i* and *k* positions in different helices have the potential to interact in a parallel hendecad of a coiled coil; likewise, *k*–*e*′ and *g*–*e*′ (Fig. 6b). This is somewhat similar to the *e*–*g* situation of a heptad, which is usually an *r*, *r* + 5 interaction (between *g* and *e*′; Lupas, 1996).

Thus according with the proposed coiled-coil arrangement, more than 60% of *k*, *g* and *c* residues are charged, and a little more than half of *b*, *f*, *i* and *j* positions. Of the charged cases, over 75% of *b* and *k* and > 65% of *g* and *f* are negatively charged, whilst *c*, *j* and *i* have a mild positive preference. The strongest correlations are compatible with electrostatic interactions of both intra-helical and inter-helical nature (Table 7). In Scy CC51 40% of *e* positions are charged but 90% of these are positive; these correlate strongly with negative charge at *b*, *g* and *k*, thereby promoting both intra- and inter-helical stability.

3.8. Two alternative supercoil models of a 51-residue repeat

The 51-residue (7, 11, 11, 11, 11) Scy CC51 repeat sequence is remarkably regular. The first characterised hendecad sequence in contrast has very few 7-residue units, which are irregularly spaced in a hendecad environment (Peters et al., 1996; Stetefeld et al.,

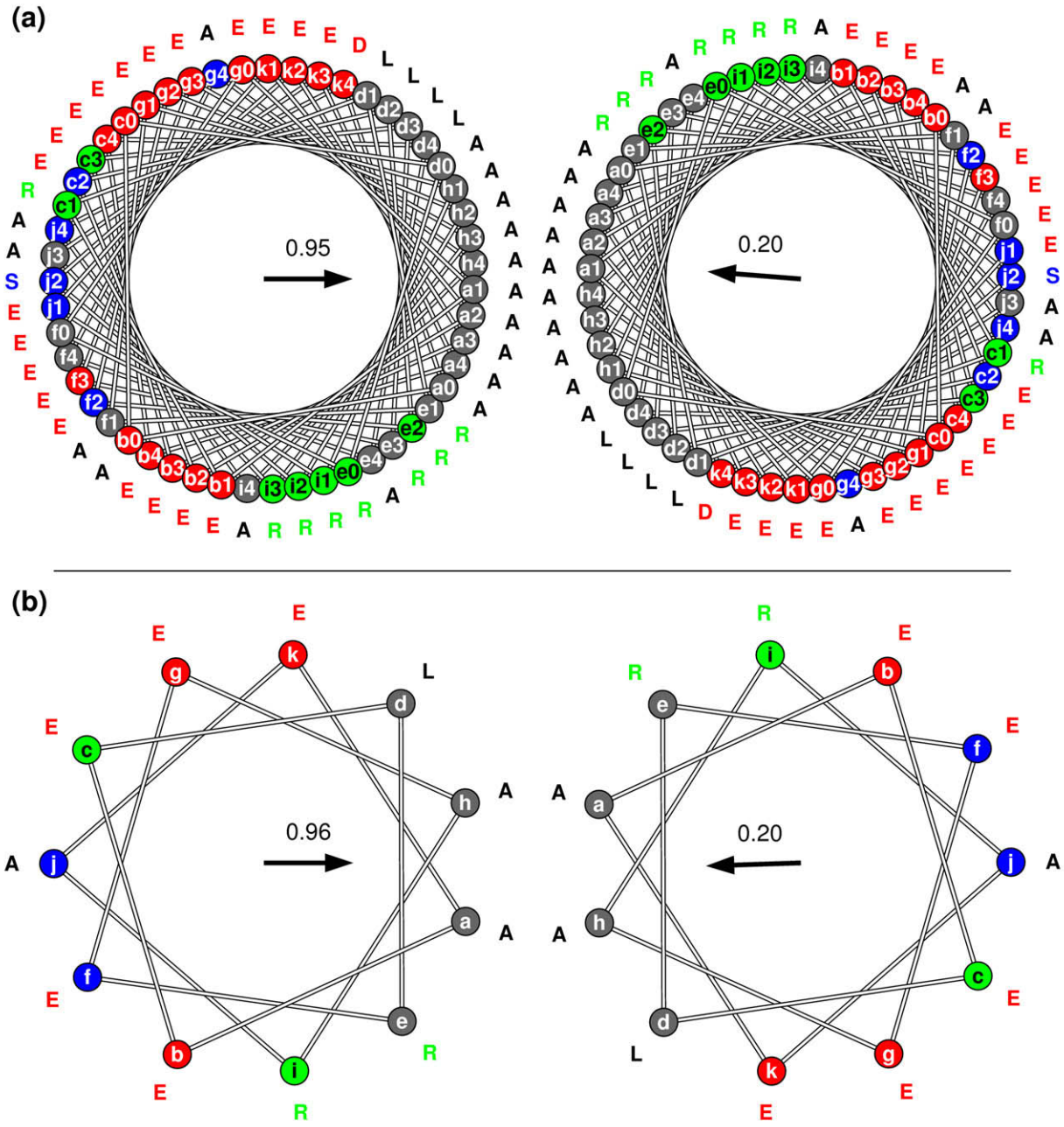

Fig. 6. Amino acid preferences at each position of the 51-residue repeat in Scy CC51 looking from the N-terminal end of a supercoiled α-helix. The most common amino acid at each position is shown, whilst each circle is coloured according to the majority type of amino acid (grey, hydrophobic; blue uncharged polar; red negatively charged; green positively charged). These indicate intra- and inter-helical interactions that would occur in a dimer. From the residues occupying positions a^1–k^4 only, arrows indicate the mean hydrophobic moments per residue using two different hydropathy index scales (on left Kyte and Doolittle, 1982, and right Sweet and Eisenberg, 1983). (a) Uniform supercoil arrangement with 51 unique positions. (b) Standard hendecad supercoil, which the 11-residue units would adopt in the variable supercoil model. Preferences only for positions a^1–k^4, where a is averaged over all positions of $a^{1,2,3,4}$, b over $b^{1,2,3,4}$, etc.

2000). Scy is therefore novel in this respect. There is a precedent for a periodically modified coiled-coil sequence, although that is a canonical heptad repeat punctuated by stutters (Lupas et al., 1995).

There is then the consideration of the structure adopted by such a regularly interrupted hendecad sequence. Intuitively, the 7-residue units would be assumed to form coiled coils with a canonical left-handed twist, with the 11-residue units adopting a regular right-handed twist as in uninterrupted hendecad sequences. We refer to this as the "variable supercoil" arrangement. However, given the remarkable regularity of the punctuation of the hendecads with single heptads, it is appropriate to consider an alternative conformation whose structure reflects the regularity of the sequence repeat; in other words, a "uniform supercoil" model.

The structural basis of the variable supercoil is that the periodic 7-residue insertions are accommodated by (i) periodic local distortions, in this case underwinding (decrease in right-handedness), of the hendecad-based coiling and/or (ii) periodic changes in the number of residues per turn of the α-helix; counterpart modifications have been characterised in canonical coiled coils (Strelkov and Burkhard, 2002). For the 7-residue insertions to adopt a canonical heptad structure in CC51, the first type of distortion would predominate. The CC51 amino acid profile's similarities with hendecad repeats in known coiled coils are indeed consistent with

Table 6
Characteristics of amino acids at positions 8–51 (a^1–k^4) of the CC51 repeat, and at each position of the hendecads in tetrabrachion (Peters et al., 1996) and the dimerization domain of the *b* subunit of ATP synthase; the latter were calculated from the alignment of sequences from 9 species (Del Rizzo et al., 2002).

Hendecadposition	Proportion hydrophobic			Most common amino acid and their relative frequencies		
	Scy	Tetrabrachion	*b*-Subunit	Scy	Tetrabrachion	*b*-Subunit
a	0.86	0.97	0.80	A 0.63	I 0.58	A 0.78
b	0.19	0.10	0.22	E 0.32	S 0.29	E 0.22
c	0.09	0.16	0.26	E 0.28	S 0.26	E 0.24
d	0.59	0.45	0.41	L 0.26	V 0.19	E 0.28
e	0.46	0.44	0.81	R 0.31	N 0.19	L 0.30
f	0.24	0.12	0.17	E 0.26	D 0.28	E 0.33
g	0.15	0.12	0.15	E 0.35	T 0.25	E 0.24
h	0.86	0.97	0.73	A 0.64	L 0.48	A 0.42
i	0.24	0.32	0.24	R 0.25	L 0.25	K 0.31
j	0.22	0.33	0.18	A 0.20	N 0.19	E 0.24
k	0.20	0.43	0.13	E 0.47	K 0.18	E 0.22

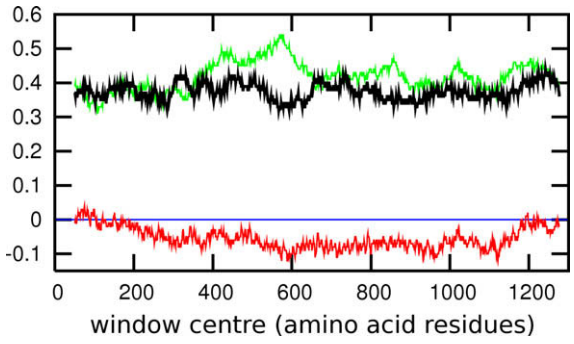

Fig. 7. Hydrophobicity and charge features of the full Scy sequence, averaged in a sliding window of 100 residues. The proportions of hydrophobic (black) and positively or negatively charged (green) residues is shown. The red line shows the mean net charge in the window.

a standard right-handed conformation (Table 6) with the 7-residue units accommodated by a local structural distortion.

In principle, the alternative possibility for any unusual coiled-coil repeat is that it adopts a uniformly supercoiled conformation. Novel repeats were found in *Giardia* proteins (Holberton et al., 1988; Marshall and Holberton, 1993, 1995), including 2 with periodicities (29/8 and 25/7) greater than 3.6. These may be interpreted as continuous right-handed coiled coils (Hicks et al., 1997) rather than coiled-coils with stutters. The notable contrast in the *e* position compared to ATPase subunit *b*, and the very different core residues compared to tetrabrachion (Table 6), also imply that CC51's structure might fundamentally differ to both. However, the precise arrangement of core sidechains in such a supercoil remains speculative without model-building. Therefore to evaluate the two alternative models, we calculated the hydrophobic moment of all residues at the 44 a^1–k^4 positions arrangement in (i) a standard hendecad supercoil with 11 distinct positions and (ii) a uniform supercoil of 51 positions. We omitted all residues in positions a^0–g^0 from these calculations, since in (i) the locally modified-hendecad model, the nature of the structural distortions accommodating them remains more speculative. We found that the modified hendecad arrangement (i) gives a slightly larger moment (Fig. 6).

When considering the structural parameters either of a uniform or periodically modified supercoil, a complication is that the number of residues per turn in an α-helix is variable between different coiled-coils, and is frequently >3.60 (Strelkov and Burkhard, 2002). Although it generally decreases as the number of helices increases (Seo and Cohen, 1993), it is around 3.65 in the solved structure of a tetrabrachion fragment (Stetefeld et al., 2000). Perhaps depending on its oligomer state, it is therefore possible for CC51 to form a shallow left- instead of right-handed coiled coil. Also possible is an essentially straight hydrophobic stripe, should the α-helices have around 3.64 residues per turn (if uniformly supercoiled) or approaching 3.66 (if variably supercoiled but largely hendecad based). The large number of inter-helical charge interactions might be expected to overcome such an arrangement's reduced stability. So too in the latter case would the periodic changes in the twist, providing localised tightening of what would otherwise be a straight assembly.

For the sake of simplicity we assumed the uniformly supercoiled arrangement of CC51 when calculating its structural parameters. We found that a right-handed dimer would have a minimum pitch of 370Å and a maximum of 1573Å, depending on the assumed α-helical parameters and oligomer state. This resulted in a coiled coil of between 183–188 nm. The only case where the number of residues per turn (3.65) of the α-helix was greater than the domain's hydrophobic periodicity resulted in a pitch of -2848Å (the sign indicates left-handedness), and also a length of 188 nm. The values of pitch and length decreased only fractionally as oligomer state increased. The largest pitch angles possible were calculated to be 4.7° in a dimer, 6.5° in a trimer and 7.2° in a tetramer. The very large pitch values (and correspondingly very small pitch angles of ~1°) describe the almost straight helical assemblies.

3.9. Expected Scy oligomer state

The simplest assembly of Scy is a long "rod" formed by the two coiled-coil domains. However, the likely flexible hinge domain intervening between CC7 and CC51 means that more complex structures may be possible. A consideration of potential means of higher-order assembly is relevant to Scy's cytoskeletal role. For example, CC51 might form a long "stalk" with the CC7 domains splaying apart at the hinge to form other associations, but this is highly speculative. Our experimental data is consistent with compact, rod-like parallel coiled-coil oligomers, and we therefore consider this arrangement only, in which both CC7 and CC51 have the same oligomer state.

The amino acids at *a* and *d* of CC7 described earlier most strongly suggest that Scy forms a (parallel) dimeric coiled coil, with a trimer possible and a tetramer unlikely. The composition of the CC51 core positions is compatible with this conclusion; further, a consideration of the extreme preference for alanine at *a* and *h* but not *d* (or *e*) positions also may suggest a dimer, as follows. Hendecad *a* and *h* positions have a somewhat similar orientation to the *a* and *d* positions of heptads, but with the packing geometry shifted towards that in *x*-positions of stutters and stammers in canonical coiled coils (Peters et al., 1996; Brown et al., 1996; Stetefeld et al., 2000). That is, the packing considerations at *a*, *h*

Table 7
Potential charge interactions in the hendecad units of the Scy CC51 domain. The first three columns show the total number of charged/polar, and hydrophobic, residues at r where r is a^{1-4}–k^{1-4} only. Of those instances where r is occupied by a charged residue, the total numbers of oppositely/like charged, polar and hydrophobic residues at $r + x$ is shown by the final four columns. The pairs in the top 6 rows represent intra-helical interactions, and those in the bottom 3 rows, inter-helical.

r	Charged or polar residues at r	Hydrophobic residues at r	x	$r+x$	Opposite-charged residues at $r+x$	Like-charged residues at $r+x$	Uncharged polar residues at $r+x$	Hydrophobic residues at $r+x$
b	74	21	+3	*e*	22	1	8	22
c	83	12	+4	*g*	31	6	10	11
k	77	19	+3	*c*	28	13	14	12
j	74	22	+4	*c*	21	7	5	9
f	71	26	+4	*j*	16	2	14	11
i	72	24	+4	*b*	19	7	11	17
k	77	19	+5	*e*	24	5	8	30
e	51	46	+2	*g*	22	2	10	5
i	72	24	+2	*k*	31	6	3	14

positions are subject to some of the same spatial restrictions as *x* layers, but not to as extreme a degree. The significance of this is that the space available for these interactions reduces significantly as oligomer state decreases. The side chains packing head-to-head in *x*-positions in canonical dimers are probably restricted to glycine, and those in trimers to alanine, threonine or valine (Lupas et al., 1995). In comparison, the extra space afforded in hendecad *a*, *h* layers means that alanine should be tolerated at these positions in dimers. Further, a tetrameric arrangement of Scy might well be disfavoured by the resulting cavities in the middle of these alanine layers. "Unpacked" alanines have been shown to be unfavourable in designed coiled-coil peptides (Gonzalez et al., 1996c; Monera et al., 1996). Cavities alone might not prevent a tetramer, however: three water-filled cavities are present in the tetrameric tetrabrachion structure (Stetefeld et al., 2000), bounded by a mixture of sidechain and mainchain atoms.

There is also the question of why, if Scy were a trimer or tetramer, is there not more diversity at *a* and *h*; medium-sized and larger side chains are expected to be relatively destabilizing in dimers, but are fully compatible with higher oligomer states. In CC51, more than 60% of *a* and *h* are occupied by alanine; this is similar to the occupancy in the dimeric b subunit (Table 6) and contrasts with the isoleucine/leucine preference of tetrabrachion. Finally, whereas tetrabrachion has a weaker preference for hydrophobic residues at both its central *d*, *e* positions, Scy has a relatively strong hydrophobic character at one of these, reminiscent of the dimeric ATPase *b* subunit (see Section 3.6). In summary, the CC51 sequence features are highly compatible with the dimeric arrangement implied by those of CC7 alone.

3.10. Alternative interpretations

An uninterrupted periodicity of 11/3 alone can give rise to a left-handed coiled-coil, such as in apolipoprotein E3 (Wilson et al., 1991). However, in that structure there are short turns between helices in an antiparallel arrangement which contribute to an adjustment of the core position. In the long Scy domain, there are no prolines and few glycines – for example there are none in the first 250 residues, and other stretches of over 100 and 150 Gly-free residues occur. The right-handed, looping arrangement of the helices of apolipoprotein A-I can also be ruled out (Borhani et al., 1997) as this "22/6" repeat involves a proline in almost every unit of 22. However, Scy does have several conserved glycines, a notable case being 3 equally spaced 17 residues apart starting at 1111; one or more breaks in the coiled-coil structure are certainly plausible. Also discounted is the glycine-rich, helical, 11-residue repeat of α-synuclein, whose hydrophobic interface interacts directly with micelle membranes (Bisaglia et al., 2005), although computational evidence implies it too may form coiled coils at high concentrations (Mihajlovic and Lazaridis, 2008).

3.11. The Scy and FilP sequences have near-identical domain organisation

In the *S. coelicolor* genome the gene adjacent to *scy* (*SCO5397*) is *filP* (*SCO5396*) (Bentley et al., 2002; http://strepdb.streptomyces.org.uk). FilP is strongly predicted to contain coiled coils (Bagchi et al., 2008), and it forms filamentous assemblies *in vitro* and *in vivo*. FilP is required to maintain normal hyphal rigidity of *S. coelicolor*. Its length is 310 amino acids, less than a quarter of Scy's. In FilP COILS prediction shows very high probability between Asp17 and Asp127, and between Glu134 and Ala288. However, a closer analysis revealed very striking sequence similarities with the Scy protein. Although the first COILS-predicted domain includes a proline (Pro65) near its end, this residue appears equivalent to Pro64 of Scy. The region N-terminal to it can clearly be assigned an uninterrupted, canonical heptad repeat (residues 17–64), by manual inspection as well as by COILS and consistent with FT analysis (Fig. 1b); thus, a CC7 domain of almost 7 heptads. As in Scy, the proline which follows the canonical coiled-coil domain marks the beginning of a long (in this case, 227 residues) stretch in which prolines are absent, ended by Pro293. The hinge region is apparent too, residues 65–71 (PSYAGLG) (Fig. 2b). FT analysis of the long proline-free region produces a peak in hydrophobic periodicity of 3.636 (40/11) (Fig. 1b) indicating a non-canonical, right-handed coiled-coil. As in Scy, this is also apparent as a series of discontinuities in the COILS prediction despite high probability being maintained. Although this peak is slightly different to Scy's 3.643, we found that the best hydrophobic-polar patterns which could be ascribed to this domain are 7, 11, 11, 11 (period 40/11) and, marginally better, 7, 11, 11, 11, 11 (51/14) with 73% of expected hydrophobic positions, and 85% of polar, satisfied. Towards the end of the putative FilP CC51 domain, the period appears to shorten somewhat (Fig. 3b); so too does Scy's in the equivalent region. Also telling are the changes in composition between the CC7 and CC51 domains, which are very similar to Scy's. FilP CC7 is 12% alanine and 12% glutamate; these change to 21% and 15% in FilP CC51. Also near-identical to Scy is the drop in leucine from 10% to 8%, and the fractional increase in arginine.

If the N-terminal canonical coiled coil of FilP forms a parallel arrangement, then possible g_n to e_{n+1} pairings are Arg-Val, Asp-Leu, Arg-Glu, Arg-His, Glu-Gln and Gln-Glu. The amino acids at *a* and *d* positions in each heptad are Arg_a,Val_d; Ile_a,Leu_d; Arg_a,Ala_d; Ile_a,Leu_d; Ile_a,Leu_d; Thr_a,Ala_d; Val_a,Ala_d: there are β-branched amino acids at five out of seven *a* positions (expected to strongly disfavour a tetramer), and three leucines at *d*. However, the notion that this combination appears even more classically dimer-like than Scy's, seems confounded by the demonstration that FilP's 93%-identical homologue in *S. reticuli*, known as AbpS, was shown to be tetrameric (Walter and Schrempf, 2003). However, the same authors showed that the progressive deletion of the N-terminus, including this coiled coil, resulted in the successive AbpS mutants becoming dimeric and unable to tetramerise. This supports the hypothesis that the minimal assembly unit of FilP is a dimer, but that a dimer of dimers is a stable association, mediated by some part of the N-terminal region of each dimer; these would interact with each other or with a more C-terminal part of the partner dimer. This is reminiscent of the assembly of dimers-of-dimers in some IF proteins (Sokolova et al., 2006), but key interactions of the latter involve large, non-coiled-coil head domains which Scy and FilP appear to lack. The segment of AbpS implicated in mediating assembly beyond dimers straddles both CC7 and CC51 (resi-

dues 30–109; Walter and Schrempf, 2003). This suggests that at most one, and possibly neither, of the two domains alone possesses the ability to form the higher-order assemblies. The conclusion from our Bacterial Two Hybrid analysis of Scy fragment association, that the most fundamental coiled-coil of Scy is parallel, is therefore unaffected.

Given the great similarity between Scy and FilP/AbpS, it can be hypothesised that Scy's cytoskeletal function involves a role as a giant unit of filament-formation. If the case, the interaction between the more N-terminal CC7 and a different part of the partner dimer would seem likely, given the charge profile (Fig. 7). Neither a parallel nor antiparallel pairing of dimers would be ruled out; these would respectively produce unipolar or bipolar filaments.

3.12. Homologues found by BLAST searches

With composition-based scoring turned off, BLASTP unsurprisingly returned a vast number of hits from UniProt when the Scy sequence was used as a query. These included a diverse array of canonical coiled-coil proteins. Interestingly, even extending the hit list to 1000 sequences failed to find FilP. We note that a local sequence alignment of FilP and Scy produced by WATER with default parameters was only 26% identical over 336 residues (c.f. Scy and rabbit tropomyosin 24% over 333). With composition-based scoring on, we found only a small number of clear BLASTP hits, all in *Streptomyces* species, with a large step-change in E-value beyond that. Sequence C4DLU2 is a possible Scy homologue in the Actinomycete *Stackebrandtia nassauensis*. Amongst the UniProt, Broad Institute and Sanger databases, clear Scy homologues were found in 13 *Streptomyces* strains/species, of 80–99% identity; several of these appear as two large fragments rather than a single sequence. A multiple alignment is available in the supplementary material (Fig. S2). The significance of these sequences for the structural consideration is that the indels relative to Scy's CC51 domain are nearly all of length $11x$, where x is a small integer. Relative to Scy there are numerous deletions of 11, 22 or 33 residues. A 44-residue deletion (936–979) occurs in 3 species; in 4 others, a 33-residue deletion occurs in the same region (939–971). Interestingly this coincides with the only brief drop in the COILS-predicted probability of the domain. Several insertions of 11 and 22 residues are apparent relative to *S. coelicolor*. What is true for Scy also applies to FilP and some of its Actinomycetales homologues in UniProt, found by BLASTP. Sequences C2A9R0 (*Thermomonospora curvata*) and A6W7H8 (*Kineococcus radiotolerans*) are both putative proteins with 31% identity to FilP. Compared to FilP's CC51 domain, the former has 5 separate deletions, of 33, 22 or 11 residues length. The latter has two deletions, of 40 (interestingly, an $11x + 7$ deletion) and 11 residues. These results support the idea that the CC51 structure of both proteins is mostly composed of normal hendecads, with local 7-residue register-shifts. If a uniformly supercoiled structure, indels of $51x$ would be expected.

3.13. Results of the coiled coil/Fourier Transform/composition-based screen

Of the ~10 million UniProt sequences searched, 4604 passed our screening procedure for a Scy-type domain structure. Many of these we discounted as they have only a minimal proportion of their sequence in a proline-free region of 51/14 periodicity; this proportion is how we ranked the hits. The top 3 hits are Scy or a fragment thereof in different *Streptomyces* species. FilP is ranked 19 and its homologues are ranked 21–26 and 50. The screen found proteins in all kingdoms of life but bacteria appear to dominate the upper part of the list, part of which is shown in Table 8. A larger sample is included in the supplementary material (Table S4). We closely examined COILS predictions of the top sequences, which

Table 8
Top 25 hits of coiled-coil/FT/composition screen. r3.643 is the number of residues falling inside a region which fulfils all of the criteria of period of hydrophobic residues (3.6429 ± 0.001) and Pro/Gly content. The proportion p3.643 is r3.643/seq. length.

UniProt accession	Seq. length	r3.643	p3.643	Species	Notes/UniProt annotation
Q9L2C3	1326	1235	0.931	*S. coelicolor*	Scy
B1W080	1297	1184	0.913	*S. griseus*	Scy
B5HMN0	527	434	0.824	*S. sviceus*	Scy N-terminal segment
B8EH97	236	194	0.822	*Borrelia burgdorferi* 156a	Haemolysin accessory protein
A3K1K8	542	430	0.793	*Sagittula stellata* E-37	Putative uncharacterised protein
Q06431	474	334	0.705	*Betula verrucosa* (White birch) (*Betula pendula*)	BP8 protein
Q5CT86	282	197	0.699	*Cryptosporidium parvum* Iowa II	497P1.13-like
Q661A9	2162	1509	0.698	*Borrelia garinii*;	Putative uncharacterised protein
B5Z6L3	285	195	0.684	*Helicobacter pylori* (strain G27)	Putative uncharacterised protein
C4BNQ6	304	207	0.681	*Saccharomonospora viridis* DSM 43017	Putative uncharacterised protein
C4DLU2	684	464	0.678	*Stackebrandtia nassauensis* DSM 44728	Putative uncharacterised protein; putative Scy homologue
Q4T579	293	193	0.659	*Tetraodon nigroviridis* (Green puffer)	Chromosome undetermined SCAF9366, WGS sequence
Q4RZQ4	366	236	0.645	*Tetraodon nigroviridis* (Green puffer)	Chromosome 18 SCAF14786, WGS sequence
O44236	650	415	0.638	*Ciona intestinalis* (Transparent sea squirt)	mRNA,endostyle-specific
Q1CU40	308	196	0.636	*Helicobacter pylori* (strain HPAG1)	Putative uncharacterised protein
A2EBP2	360	227	0.631	*Trichomonas vaginalis* G3	Putative uncharacterised protein
C6XGJ9	1828	1152	0.63	*Liberibacter asiaticus* (strain psy62)	Chemotaxis sensory transducer
C1WH33	316	199	0.63	*Kribbella flavida* DSM 17836	Putative uncharacterised protein
Q9K4B5	310	195	0.629	*S. coelicolor*	FilP
C0QY77	7854	4933	0.628	*Brachyspira hyodysenteriae* (str. ATCC49526/WA1)	Putative uncharacterised protein
B5GL19	311	195	0.627	*S. clavuligerus* ATCC 27064	Cellulose-binding protein (FilP/AbpS)
B5HMN3	311	195	0.627	*S. sviceus*	FilP/AbpS annotated as 'Cellulose-binding protein'
B1W081	312	195	0.625	*S. griseus*	FilP/AbpS annotated as 'Putative cellulose-binding protein'
Q82JA6	312	195	0.625	*S. avermitilis*	FilP/AbpS annotated as 'Putative cellulose-binding protein'
B5H9C8	312	195	0.625	*S. pristineaespiralis*	FilP/AbpS annotated as 'Cellulose-binding protein'
B5GI28	311	194	0.624	*S.* sp. SPB74	FilP/AbpS annotated as 'Cellulose-binding protein';
C5C1X1	619	381	0.616	*Beutenbergia cavernae* (strain ATCCBAA-8)	Putative uncharacterised
C1RAU5	317	195	0.615	*Catenulispora acidiphila* DSM 44928	Putative uncharacterised protein
Q7RKP8	281	172	0.612	*Plasmodium yoelii yoelii*	Putative uncharacterised protein PY02852
C0VBN9	461	278	0.603	*Xylanimonas cellulosilytica* DSM 15894	Putative uncharacterised protein

indicated that a fairly large proportion are false positives. In contrast to Scy and FilP, these sequences appear to have several regions of canonical heptad repeats, which are separated by obvious non-coiled-coil segments, often of considerable length. An extreme example is the 8th hit, a putative *Borrelia* protein of >2000 amino acids which appears to be a long series of segmented canonical coiled coils. The inclusion of non-coiled-coil domains adjacent to or within coiled coils affects the hydrophobic periodicity, producing misleading peaks as far as interpreting novel repeats is concerned. The approach also produced false negatives; for example, *S. avermitilis* (Omura et al., 2001) Scy is missing because its N-terminal deletion means that it lacks a 40-residue proline-free region, which our screen required in addition to a separate, longer one.

A very large proportion of the hits are annotated as putative and uncharacterised. An exception is the 6th best hit, a late embryogenesis abundant (LEA) protein BP8, from birch (Puupponen-Pimiä et al., 1993). This is a group 3 LEA protein (Pfam LEA_4 family Finn et al., 2008), which is distinct from several other LEA classes. This group contains repeats of 11- and 22-residues (Liu and Zheng, 2005) and strong coiled-coil predictions occur in these regions; some of these proteins were missed by our search due to their glycine content. However, in the identified 334-residue segment of BP8 with a mean strongest FT peak of 51/14, inspection showed the signal arises from a relatively short central region of 150–200 residues; fewer than 4 repeats.

In the upper echelons of the hit list there are, perhaps unsurprisingly, numerous Actinomycetales proteins. One example of an apparently genuine hit, ranked 73rd, is a 500 amino acid protein from *Actinomyces urogenitalis*, annotated in UniProt (C0W8W5) as a predicted low-complexity hydrophilic protein. It was ranked 854th in the non composition-based scoring BLASTP search we performed using Scy against the same database, but a local alignment shows 31% identity to Scy. Inspection and COILS suggests it has the hallmark CC7 and CC51 domains (summarised in Fig. 2c). Although the N-terminal part of CC51 appears that it might well be an additional CC7 domain, the remainder of the 261-residue region we denote as CC51 has high COILS probability throughout but fairly evenly-spaced changes in the heptad register, many of which (such as three 5-residue insertions in the heptad) do not resemble the discontinuities seen in canonical coiled coils; it also has several stammers, stutters and a skip. It has a 7-residue hinge, starting at Pro65 (PSYAGLG).

Also noteworthy are a number of hits, starting at rank 60 (UniProt:A4FLT9), which are annotated as putative homologues of DivIVA, another bacterial cytoskeletal protein with predicted coiled coils (Edwards et al., 2000). We note that previously one DivIVA-related protein has been described as containing a hendecad motif (Bi et al., 2008). However, the non-canonical domains we identified are rather short, <150 residues. Inference of a periodically modified hendecad in fewer than 3 instances of the repeat is therefore more questionable, but the N-terminal coiled coil and hinge are apparent, and the resemblance may be significant.

4. Conclusions

We have shown that the primary structure of the *Streptomyces* Scy protein exhibits a marked, novel repeat which is highly compatible with an unusual coiled-coil structure. Apart from a canonical coiled-coil domain near its N-terminal end, the possibility that the remainder of the sequence is canonical with intervening serial non-coiled-coil segmentation, is discounted.

The repeat appears to be essentially hendecad-based but very periodically modified by the replacement of a hendecad by a heptad. Evidence from structural calculations and comparative sequence analysis suggests that the likely structure reflects this pattern, with a hendecad-associated, probably right-handed twist that is distorted locally to accommodate the regular insertions, perhaps underwound to a more canonical-like twist. However, coiled coils are not static structures; different modes of dealing with register-shifts are observed even in different copies of the same protein (Strelkov and Burkhard, 2002) and alternative dynamic local twists along the long Scy CC51 domain cannot be ruled out. Notable differences in amino acid selection at some core positions also suggest that even the 11-mer units may not entirely share the structural characteristics of known hendecad coiled coils. Sequence features of both this domain and the canonical N-terminal domain suggest the most likely oligomerisation state is a dimer.

We propose that like Scy, FilP is best described as consisting of two distinct coiled-coil domains of different nature, rather than a segmented arrangement of several canonical coiled coils separated by non-coiled-coil linkers. FilP is essentially the same as Scy but with a much truncated CC51 domain. We believe that the primary structures of Scy and FilP are clear evidence that they are homologues, despite the unremarkable sequence identity and their difference in size (Scy is four times longer). Given their almost identical domain structure including the same unusual repeat, and that they are products of neighbouring genes on the chromosome, it seems extremely likely that Scy and FilP share descent from a common ancestor. The fact that mutant phenotypes of *scy* and *filP* are different suggests that Scy and FilP are paralogues and that they evolved related, but distinct functions. Further structural and biological characterisation of these proteins will establish the relevance of these structural similarities.

Bundled filament-formation by FilP has been demonstrated *in vivo* (Bagchi et al., 2008) which raises the interesting question as to whether Scy is an unusually large (190 nm) formation-unit of filaments. The primary structure of both suggests a dimeric coiled coil; this and the previously-demonstrated tetramerisation of a FilP homologue, AbpS (Walter and Schrempf, 2003), are compatible with a mechanism of assembly not unlike that of intermediate filaments. However, in contrast to other bacterial cytoskeletal components such as CreS (Ausmees et al., 2003), the domain organisation and coiled-coil basis of Scy and FilP appears to be fundamentally distinct from that of IF and IF-like proteins.

FT of hydrophobic/polar repeat periodicity is long established as a powerful tool for analyzing sequences of fibrous proteins (McLachlan and Stewart, 1976; Gruber et al., 2005). We used such a method complemented with coiled-coil prediction (Lupas et al., 1991; Delorenzi and Speed, 2002) and amino acid composition, to find multiple domains with differing periods in the same sequence. We have identified proteins with a broadly similar domain organisation as Scy, whose homology is however far less clear. These proteins do however appear to occur mostly in Actinomycetes with filamentous hyphae.

The same structural principles and behaviour as Scy/FilP are likely to apply to at least some of the proteins we found. Interestingly, one of these with similar domain organisation is DivIVA, another bacterial cytoskeletal protein that is essential for polarised growth in *Actinomycetes* (Flardh, 2003; Letek et al., 2008) whilst it controls cell division in others, such as *Bacillus subtilis* (Edwards et al., 2000; Edwards and Errington, 1997). It will be of great interest to establish to what extent Scy, FilP and DivIVA might share structural characteristics with potential relevance to their biological function. Our search approach requires improvement, but appears to have overcome some of the problems of more traditional alternatives in which target sequences of interest were missing or hidden. This may serve as a starting point for investigating a putative cytoskeletal role of uncharacterised sequences rapidly appearing from genome sequencing projects, especially in some Actinomycetes which, like *S. coelicolor*, are of biomedical interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jsb.2010.02.008.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Ausmees, N., Kuhn, J.R., Jacobs-Wagner, C., 2003. The bacterial cytoskeleton: an intermediate filament-like function in cell shape. Cell 115, 705–713.

Bagchi, S., Tomenius, H., Belova, L.M., Ausmees, N., 2008. Intermediate filament-like proteins in bacteria and a cytoskeletal function in *Streptomyces*. Mol. Microbiol. 70, 1037–1050.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L.D., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J.R., Barrell, B.G., Parkhill, J., Hopwood, D.A., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417, 141–147.

Bi, Y., Watts, J.C., Bamford, P.K., Briere, L.K., Dunn, S.D., 2008. Probing the functional tolerance of the *b* subunit of *Escherichia coli* ATP synthase for sequence manipulation through a chimera approach. Biochim. Biophys. Acta 1777, 583–591.

Bisaglia, M., Tessari, I., Pinato, L., Bellanda, M., Giraudo, S., Fasano, M., Bergantino, E., Bubacco, L., Mammi, S., 2005. A topological model of the interaction between α-synuclein and sodium dodecyl sulfate micelles. Biochemistry 44, 329–339.

Borhani, D.W., Rogers, D.P., Engler, J.A., Brouillette, C.G., 1997. Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation. Proc. Natl. Acad. Sci. USA 94, 12291–12296.

Bosgraaf, L., van Haastert, P.J., 2006. The regulation of myosin II in Dictyostelium. Eur. J. Cell Biol. 85, 969–979.

Brown, J.H., Cohen, C., Parry, A.D., 1996. Heptad breaks in α-helical coiled coils: stutters and stammers. Proteins Struct. Funct. Genet. 26, 134–145.

Cabeen, M.T., Jacobs-Wagner, C., 2005. Bacterial cell shape. Nat. Rev. Microbiol. 3, 601–610.

Cabeen, M.T., Jacobs-Wagner, C., 2007. Skin and bones: the bacterial cytoskeleton, cell wall, and cell morphogenesis. J. Cell Biol. 179, 381–387.

Charbon, G., Cabeen, M.T., Jacobs-Wagner, C., 2009. Bacterial intermediate filaments: in vivo assembly, organization, and dynamics of crescentin. Genes Dev. 23, 1131–1144.

Crick, F.H.C., 1953. The packing of α-helices: simple coiled-coils. Acta Crystallogr. 6, 689–697.

Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M., Barton, G.J., 1998. JPred: a consensus secondary structure prediction server. Bioinformatics 14, 892–893.

Del Rizzo, P.A., Bi, Y., Dunn, S.D., Shilton, B.H., 2002. The second stalk of *Escherichia coli* ATP synthase: structure of the isolated dimerization domain. Biochemistry 41, 6875–6884.

Del Rizzo, P.A., Bi, Y., Dunn, S.D., 2006. ATP synthase *b* subunit dimerization domain: a right-handed coiled coil with offset helices. J. Mol. Biol. 364, 735–746.

Delorenzi, M., Speed, T., 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18, 617–625.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.

Edwards, D.H., Errington, J., 1997. The *Bacillus subtilis* DivIVA protein targets to the division septum and controls the site specificity of cell division. Mol. Microbiol. 24, 905–915.

Edwards, D.H., Thomaides, H.B., Errington, J., 2000. Promiscuous targeting of *Bacillus subtilis* cell division protein DivIVA to division sites in *Escherichia coli* and fission yeast. EMBO J. 19, 2719–2727.

Finn, R.D., Tate, J., Mistry, J., Coggill, P.C., Sammut, J.S., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L., Bateman, A., 2008. The Pfam protein families database. Nucleic Acids Res. (Database Issue) 36, D281–D288.

Fischbach, M., Ward, D., Young, S., Jaffe, D., Gnerre, S., Berlin, A., Heiman, D., Hepburn, T., Sykes, S., Mehta, T., Alvarado, L., Kodira, C.D., Straight, P., Clardy, J., Hung, D., Kolter, R., Mekalanos, J., Walker, S., Walsh, C.T., Lander, E., Galagan, J., Nusbaum, C., Birren, B., 2008. <http://www.broadinstitute.org/annotation/genome/streptomyces_group/MultiHome.html>.

Fischetti, V.A., 1989. Streptococcal M protein: molecular design and biological behavior. Clin. Microbiol. Rev. 2, 285–314.

Flardh, K., 2003. Essential role of DivIVA in polar growth and morphogenesis in *Streptomyces coelicolor* A3(2). Mol. Microbiol. 49, 1523–1536.

Fraser, R.D.B., MacRae, T.P., 1973. Conformation in Fibrous Proteins and Related Synthetic Polypeptides. Academic Press, London. pp. 456–465.

Gonzalez, L., Brown, R.A., Richardson, D., Alber, T., 1996a. Crystal structures of a single coiled-coil peptide in two oligomeric states reveal the basis for structural polymorphism. Nat. Struct. Biol. 3, 1002–1010.

Gonzalez, L., Woolfson, D.N., Alber, T., 1996b. Buried polar residues and structural specificity in the GCN4 leucine zipper. Nat. Struct. Biol. 3, 1011–1018.

Gonzalez, L., Plecs, J.J., Alber, T., 1996c. An engineered allosteric switch in leucine-zipper oligomerization. Nat. Struct. Biol. 3, 510–515.

Gruber, M., Lupas, A.N., 2003. Historical review: another 50th anniversary – new periodicities in coiled coils. Trends Biochem. Sci. 28, 679–685.

Gruber, M., Söding, J., Lupas, A.N., 2005. REPPER – repeats and their periodicities in fibrous proteins. Nucleic Acids Res. (Web Server issue) 33, W239–W243.

Harbury, P.B., Zhang, T., Kim, P.S., Alber, T., 1993. A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. Science 262, 1401–1407.

Harbury, P.B., Kim, P.S., Alber, T., 1994. Crystal structure of an isoleucine-zipper trimer. Nature 371, 80–83.

Herrmann, H., Aebi, U., 2004. Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular scaffolds. Annu. Rev. Biochem. 73, 749–789.

Hicks, M.R., Holberton, D.V., Kowalczyk, C., Woolfson, D.N., 1997. Coiled-coil assembly by peptides with non-heptad sequence motifs. Fold. Des. 2, 149–158.

Hoiczyk, E., Roggenkamp, A., Reichenbecher, M., Lupas, A., Heesemann, J., 2000. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. EMBO J. 19, 5989–5999.

Holberton, D., Baker, D.A., Marshall, J., 1988. Segmented α-helical coiled-coil structure of the protein giardin from the *Giardia* cytoskeleton. J. Mol. Biol. 204, 789–795.

Hornung, T., Volkov, O.A., Zaida, T.M., Delannoy, S., Wise, J.G., Vogel, P.D., 2008. Structure of the cytosolic part of the subunit *b*-dimer of *Escherichia coli* F0F1-ATP synthase. Biophys. J. 94, 5053–5064.

Insall, R.H., Machesky, L.M., 2009. Actin dynamics at the leading edge: from simple machinery to complex networks. Dev. Cell 17, 310–322.

Karimova, G., Pidoux, J., Ullmann, A., Ladant, D., 1998. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. Proc. Natl. Acad. Sci. USA 95, 5752–5756.

Kelemen, G.H., Walshaw, J., Thibessard, A., Holmes, N., Dalton, K.A., Leggett, R.M., Gillespie, M.D., Hemmings, A., Gust, B., Hunter, J.I.B., submitted for publication.

Kyte, J., Doolittle, R., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Letek, M., Ordonez, E., Vaquera, J., Margolin, W., Flardh, K., Mateos, L.M., Gil, J.A., 2008. DivIVA is required for polar growth in the MreB-lacking rod-shaped actinomycete *Corynebacterium glutamicum*. J. Bacteriol. 190, 3283–3292.

Linding, R., Russell, R.B., Neduva, V., Gibson, T.J., 2003. GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Res. 31, 3701–3708.

Liu, Y., Zheng, Y., 2005. PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in *Escherichia coli*. Biochem. Biophys. Res. Commun. 331, 325–332.

Löwe, J., Amos, L.A., 2009. Evolution of cytomotive filaments: the cytoskeleton from prokaryotes to eukaryotes. Int. J. Biochem. Cell Biol. 41, 323–329.

Lupas, A., 1996. Coiled coils: new structures and new functions. Trends Biochem. Sci. 21, 375–382.

Lupas, A., Van Dyke, M., Stock, J., 1991. Predicting coiled coils from protein sequences. Science 252, 1162–1164.

Lupas, A., Müller, S., Goldie, K., Engel, M., Engel, A., Baumeister, W., 1995. Model structure of the Ompα rod, a parallel four-stranded coiled coil from the hyperthermophilic eubacterium *Thermotoga maritima*. J. Mol. Biol. 248, 180–189.

Marshall, J., Holberton, D.V., 1993. Sequence and structure of a new coiled coil protein from a microtubule bundle in *Giardia*. J. Mol. Biol. 231, 521–530.

Marshall, J., Holberton, D.V., 1995. *Giardia* gene predicts a 183 kDa nucleotide-binding head-stalk protein. J. Cell Sci. 108, 2683–2692.

Mazouni, K., Pehau-Arnaudet, G., England, P., Bourhy, P., Saint Girons, I., Picardeau, M., 2006. The scc spirochetal coiled-coil protein forms helix-like filaments and binds to nucleic acids generating nucleoprotein structures. J. Bacteriol. 188, 469–476.

McDonnell, A.V., Jiang, T., Keating, A.E., Berger, B., 2006. Paircoil2: improved prediction of coiled coils from sequence. Bioinformatics 22, 356–358.

McLachlan, A.D., Stewart, M., 1976. The 14-fold periodicity in α-tropomyosin and the interaction with actin. J. Mol. Biol. 103, 271–298.

Mihajlovic, M., Lazaridis, T., 2008. Membrane-bound structure and energetics of α-synuclein. Proteins 70, 761–778.

Monera, O.D., Zhou, N.E., Lavigne, P., Kay, C.M., Hodges, R.S., 1996. Formation of parallel and antiparallel coiled-coils controlled by the relative positions of alanine residues in the hydrophobic core. J. Biol. Chem. 271, 3995–4001.

Norrander, J.M., Amos, L.A., Linck, R.W., 1992. Primary structure of tektin A1: comparison with intermediate-filament proteins and a model for its association with tubulin. Proc. Natl. Acad. Sci. USA 89, 8567–8571.

Olia, A.S., Casjens, S., Cingolani, G., 2007. Structure of phage P22 cell envelope-penetrating needle. Nat. Struct. Mol. Biol. 14, 1221–1226.

Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., 2001. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. Proc. Natl. Acad. Sci. USA 98, 12215–12220.

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

Parry, D.A., 2006. Hendecad repeat in segment 2A and linker L2 of intermediate filament chains implies the possibility of a right-handed coiled-coil structure. J. Struct. Biol. 155, 370–374.

Pauling, L., Corey, R.B., 1953. Compound helical configurations of polypeptide chains: structure of proteins of the α-keratin type. Nature 171, 59–61.

Peters, J., Baumeister, W., Lupas, A., 1996. Hyperthermostable surface layer protein tetrabrachion from the archaebacterium *Staphylothermus marinus*: evidence for the presence of a right-handed coiled coil derived from the primary structure. J. Mol. Biol. 257, 1031–1041.

Puupponen-Pimiä, R., Saloheimo, M., Vasara, T., Ra, R., Gaugecz, J., Kurtén, U., Knowles, J.K., Keränen, S., Kauppinen, V., 1993. Characterization of a birch (*Betula pendula* Roth.) embryogenic gene, BP8. Plant Mol. Biol. 23, 423–428.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277.

Roberts, T.M., 1987. Fine (2–5-nm) filaments: new types of cytoskeletal structures. Cell Motil. Cytoskelet. 8, 130–142.

Ross, J.H.E., Hutchings, A., Butcher, G.W., Lane, E.B., Lloyd, C.W., 1991. The intermediate filament-related system of higher plant cells shares an epitope with cytokeratin 8. J. Cell Sci. 99, 91–98.

Seo, J., Cohen, C., 1993. Pitch diversity in α-helical coiled coils. Proteins Struct. Funct. Genet. 15, 223–234.

Shin, D.H., Kim, J.-S., Yokota, H., Kim, R., Kim, S.-H., 2006. Crystal structure of the DUF16 domain of MPN010 from *Mycoplasma pneumoniae*. Protein Sci. 15, 921–928.

Smith, T.A., Strelkov, S.V., Burkhard, P., Aebi, U., Parry, D.A., 2002. Sequence comparisons of intermediate filament chains: evidence of a unique functional/structural role for coiled-coil segment 1A and linker L1. J. Struct. Biol. 137, 128–145.

Sokolova, A.V., Kreplak, L., Wedig, T., Muëcke, N., Svergun, D.I., Herrmann, H., Aebi, U., Strelkov, S.V., 2006. Monitoring intermediate filament assembly by small-angle X-ray scattering reveals the molecular architecture of assembly intermediates. Proc. Natl. Acad. Sci. USA 103, 16206–16211.

Squire, J., Cantino, M., Chew, M., Denny, R., Harford, J., Hudson, L., Luther, P., 1998. Myosin rod-packing schemes in vertebrate muscle thick filaments. J. Struct. Biol. 122, 128–138.

Stetefeld, J., Jenny, M., Schulthess, T., Landwehr, R., Engel, J., Kammerer, R.A., 2000. Crystal structure of a naturally occurring parallel right-handed coiled coil tetramer. Nat. Struct. Biol. 7, 772–776.

S.V. Strelkov, personal communication.

Strelkov, S.V., Burkhard, P., 2002. Analysis of α-helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. J. Struct. Biol. 137, 54–64.

Sumby, P., Porcella, S.F., Madrigal, A.G., Barbian, K.D., Virtaneva, K., Ricklefs, S.M., Sturdevant, D.E., Graham, M.R., Vuopio-Varkila, J., Hoe, N.P., Musser, J.M., 2005. Evolutionary origin and emergence of a highly successful clone of serotype M1 group A Streptococcus involved multiple horizontal gene transfer events. J. Infect. Dis. 192, 771–782.

Sweet, R.M., Eisenberg, D., 1983. Correlation of sequence hydrophobicities measures similarity in three-dimensional protein structure. J. Mol. Biol. 171, 479–488.

The UniProt Consortium, 2009. The Universal Protein Resource (UniProt). Nucleic Acids Res. 37, D169–D174.

Wade, R.H., 2009. On and around microtubules: an overview. Mol. Biotechnol. 43, 177–191.

Walter, S., Schrempf, H., 2003. Oligomerization, membrane anchoring, and cellulose-binding characteristics of AbpS, a receptor-like *Streptomyces* protein. J. Biol. Chem. 278, 26639–26647.

Weber, K., Geisler, N., Plessmann, U., Bremerich, A., Lechtreck, K.-F., Melkonian, M., 1993. SF-assemblin, the structural protein of the 2-nm filaments from striated microtubule associated fibers of algal flagellar roots, forms a segmented coiled coil. J. Cell Biol. 121, 837–845.

Wilson, C., Wardell, M.R., Weisgraber, K.H., Mahley, R.W., Agard, D.A., 1991. Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. Science 252, 1817–1822.

Wise, J.G., Vogel, P.D., 2008. Subunit *b*-dimer of the *Escherichia coli* ATP synthase can form left-handed coiled-coils. Biophys. J. 94, 5040–5052.

You, Y., Elmore, S., Colton, L.L., Mackenzie, C., Stoops, J.K., Weinstock, G.M., Norris, S.J., 1996. Characterization of the cytoplasmic filament protein gene (cfpA) of *Treponema pallidum* subsp. *pallidum*. J. Bacteriol. 178, 3177–3187.

Yu, W., Moreno Díaz de la Espina, S., 1999. The plant nucleoskeleton: ultrastructural organization and identification of NuMA homologues in the nuclear matrix and mitotic spindle of plant cells. Exp. Cell Res. 246, 516–526.