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.