Published online 8 October 2008 *Nucleic Acids Research, 2009, Vol. 37, Database issue* ***D315–D322***
doi:10.1093/nar/gkn675

CC+: a relational database of coiled-coil structures

Oliver D. Testa[1,2], Efrosini Moutevelis[1] and Derek N. Woolfson[1,2,*]

[1]School of Chemistry, University of Bristol, Bristol BS8 1TS and [2]Department of Biochemistry, School of Medical Sciences, University Walk, Bristol BS8 1TD, UK

Received August 15, 2008; Revised and Accepted September 22, 2008

ABSTRACT

We introduce the CC+ Database, a detailed, searchable repository of coiled-coil assignments, which is freely available at http://coiledcoils.chm.bris.ac.uk/ccplus. Coiled coils were identified using the program SOCKET, which locates coiled coils based on knobs-into-holes packing of side chains between α-helices. A method for determining the overall sequence identity of coiled-coil sequences was introduced to reduce statistical bias inherent in coiled-coil data sets. There are two points of entry into the CC+ Database: the 'Periodic Table of Coiled-coil Structures', which presents a graphical path through coiled-coil space based on manually validated data, and the 'Dynamic Interface', which allows queries of the database at different levels of complexity and detail. The latter entry level, which is the focus of this article, enables the efficient and rapid compilation of subsets of coiled-coil structures. These can be created and interrogated with increasingly sophisticated pull-down, keyword and sequence-based searches to return detailed structural and sequence information. Also provided are means for outputting the retrieved coiled-coil data in various formats, including PyMOL and RasMol scripts, and Position-Specific Scoring Matrices (or amino-acid profiles), which may be used, for example, in protein-structure prediction.

INTRODUCTION

The amino-acid sequence of a protein determines its three-dimensional structure and function. The rules by which proteins fold, however, remain elusive. Small and regular protein-folding motifs serve as model systems to test and develop understanding of protein folding. One such motif is the ubiquitous α-helical coiled coil, which is the structural subject of this article. On the bioinformatics side of the problem, tools are required that collate, relate and allow the analysis of sequence and structural information. A variety of web-based tools, databases and classification systems exist for proteins in general, for example, CATH, SCOP and Pfam (1–3). However, no single resource deals adequately with or is available for coiled-coil structures and sequences, though we note a sequence database of coiled coils predicted from the genome of *Arabidopsis thaliana*, ARABI-COIL (4). With the view of creating a relational database of coiled-coil structural and sequence information, we developed the CC+ Database described herein. Our starting point was SOCKET (5), an algorithm that locates coiled-coil interactions in protein structure coordinate sets deposited in the PDB and PQS databases (6,7).

Coiled coils comprise two or more α-helices that supercoil around each other to form rope-like structures. At the sequence level, these motifs usually have a repeating pattern of seven residues called the heptad repeat, and often designated ***abcdefg***. Residues at ***a*** and ***d*** positions are predominantly hydrophobic, which results in a hydrophobic stripe along each participating helix. These stripes come together to form the buried hydrophobic core of the coiled coil. Residues flanking the core also affect coiled-coil formation; charged amino acids at ***e*** and ***g*** positions can form salt bridges and electrostatic interactions to stabilize and specify coiled-coil structures further. A classical, dimeric coiled-coil motif is illustrated in Figure 1.

Despite the apparent simplicity of the heptad repeats and hence the relatedness of their sequences, there is considerable structural and functional diversity in their assemblies (10–13): classical bundles comprising 2, 3, 4, 5, 7 and 12 α-helices have all been observed, as well as more-complex assemblies; their functions include facilitating DNA binding, structural spacers and components of motor proteins, DNA-repair enzymes and cytoskeletal components, together with more dynamic units in the F_1F_0 ATPase and various membrane fusion apparatus.

Though many coiled-coil structures have now been designed *de novo* (14)—which suggests a good understanding of sequence-to-structure relationships—it is also apparent that at least certain natural coiled coils are very sensitive to small changes in sequence that can

*To whom correspondence should be addressed. Tel: +44 (0)117 95 46347; Fax: +44 (0)117 929 8611; Email: d.n.woolfson@bristol.ac.uk

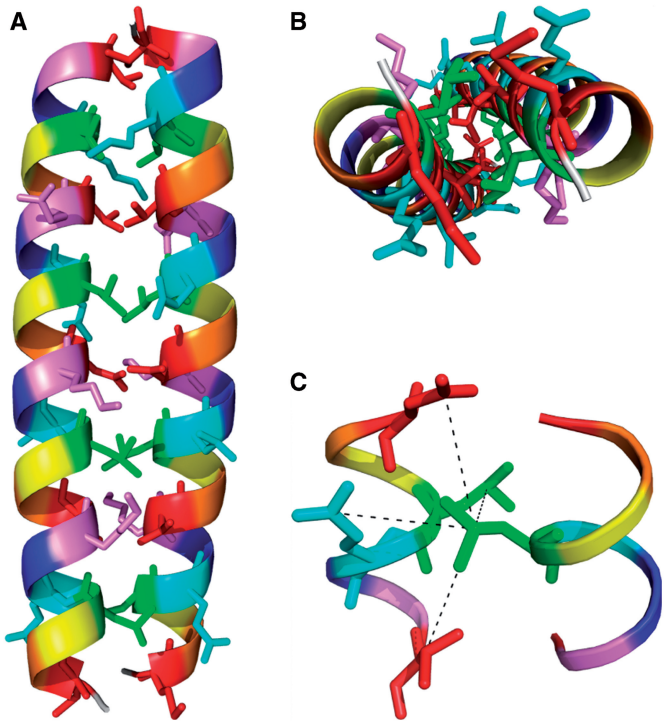

Figure 1. Views of a classical coiled-coil dimer. (**A** and **B**) Orthogonal views of the structure. The figures are shown as presented in the CC+ Database, with the backbone represented by ribbons, side chains that make up the KIH interactions as sticks and the heptad positions ***a–g*** coloured red to violet. In this way, the hydrophobic core can be easily followed in red and green. (**C**) A single KIH interaction between a 'knob' from one helix (shown in green on the rightmost helix) and a 'hole' formed by four residues on the other (left-hand helix). Images were created using PyMOL (8) and the PDB entry 2ZTA (9).

significantly affect the final coiled-coil structure adopted (15,16). Thus, understanding these balances and rules for coiled-coil assembly more precisely is key to developing the fields of coiled-coil prediction, engineering and design.

A number of computational methods have been developed to score query sequences and predict coiled-coil regions (17–20), oligomer state (19,21) and, more recently, helix orientation (22). Central to these are amino-acid propensities for specific register positions, compiled as Position-Specific Scoring Matrices, or PSSMs (17,21,23–25), and pairwise relationships (18,19,22). These approaches have met with high, but, nonetheless, varied degrees of success (26). Many, though not all (22), of these focus on information gleaned from coiled-coil sequences alone, and do not use hard structural information. An aim in creating the CC+ Database was to bridge this gap, and allow sequences and structures to be related directly.

At the structural level, coiled coils have a characteristic mode of side-chain packing known as knobs-into-holes (KIH) (27). KIH interactions are defined when an amino-acid side chain from one α-helix, the 'knob', interdigitates within a diamond-shaped constellation of four side chains, the 'hole', on a partnering helix (Figure 1C). For classical, two-, three- and four-helix coiled coils, the knobs and holes are made up of combinations of the ***a, d, e*** and ***g*** positions of the heptad repeat. There are direct relationships between the KIH interactions and sequence: for instance, in parallel dimers, a ***d***-knob from one helix combines with a hole formed by residues at ***a, d, e*** and $\boldsymbol{a}_{+1}$ of the other (subscript **+1** denotes the next heptad; Figure 1C). Thus, continuous seams of KIH interactions define the hydrophobic cores of coiled-coil assemblies. Therefore, identifying consecutive KIH constellations in protein structures presents a means to identify coiled coils on a structural rather than on a sequence basis and, in turn, to assign heptad register. The program SOCKET does this (5).

In more detail, SOCKET identifies KIH interactions trigonometrically by determining side chains whose centres of mass are within a distance of other side chains on a potential partner α-helix. The default distance cutoff is 7 Å, which does not have any chemical significance, but was derived empirically to minimize false positives. The resulting information is interpreted to describe the number and orientation of helices included in an identified helix–helix interaction and, as outlined above, to assign a register to the amino acid sequence. It is important to note that SOCKET assigns 'KIH-based helix–helix interactions' rather than identifying coiled coils *per se*, though the confidence of referring to assignments as coiled coils increases with the length over which KIH interactions are made, and we find that ≥15 residues is a useful cutoff (5).

With the aims of creating a database of coiled-coil structural assignments and relating these to underlying sequence features, we used SOCKET to scan the entire RSCB PDB (6) and PQS (7) databases. The resulting assignments were compiled into the CC+ Database presented here. In addition to information compiled from the SOCKET outputs, such as coiled-coil architecture (number of helices) and topology (helix orientation) and heptad assignments, the database harbours sequence information, metadata gleaned from the input PDB and PQS files and inter-side-chain distances. This information is stored in a relational database with a dynamic form-based web interface, which allows a range of straightforward to complex searches to retrieve user-specified coiled-coil subsets. In addition, a means of reducing statistical bias by removing redundant coiled-coil sequences has been implemented. In this article, the functionality of the database is illustrated with a small number of query examples.

DATABASE STRUCTURE AND CONTENT

The process of creating and updating the database is shown in Figure 2. The February 2008 release of the RCSB PDB (6) was scanned for coiled-coil containing structures using SOCKET (5). PQS (7) files were used to supplement the structural data set where the PDB files did not contain the biological unit. Coiled-coil assignments identified by this process were then organized into a relational MySQL database. The full set of these assignments, or 'default structures', are summarized in Figure 3, where they are also broken down according to the number and orientation of α-helices within the coiled coils.

The CC+ Database is available at http://coiledcoils.chm.bris.ac.uk/ccplus as part of a Linux, Apache,

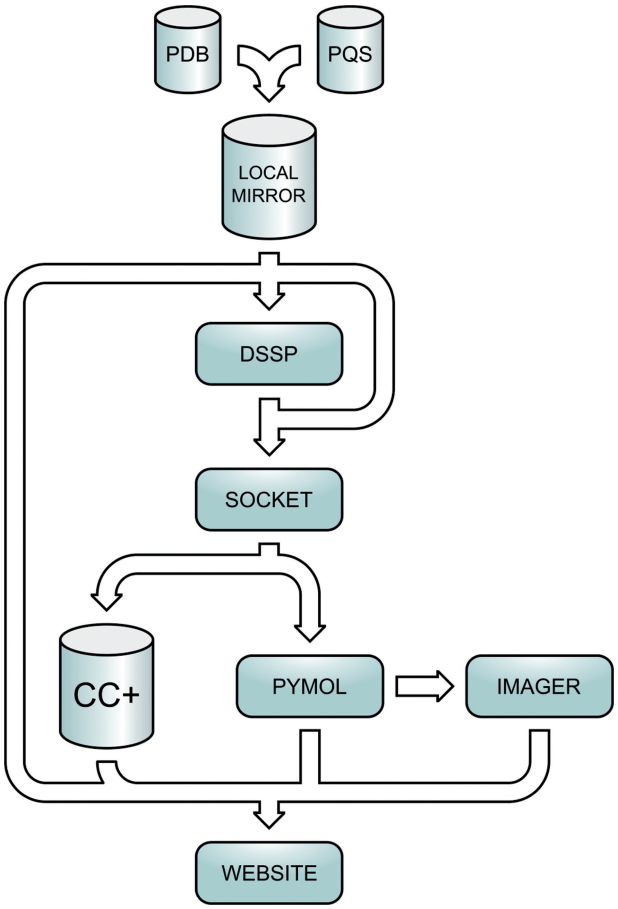

Figure 2. Flow chart illustrating processes compiling the CC+ Database and website. The CC+ Database is compiled from a mirror of the PDB (6) and PQS (7) structural files. These files are scanned using DSSP (28) and SOCKET (5) to identify KIH interactions and assign coiled coils. The organized output is stored in a relational MySQL database comprising tables of protein, coiled-coil, sequence and residue data; these data are also parsed into PYMOL (8) scripts for rendering, and then imaged automatically using an in-house algorithm. All data created by this process are available for download at http://coiledcoils.chm.bris.ac.uk/ccplus.

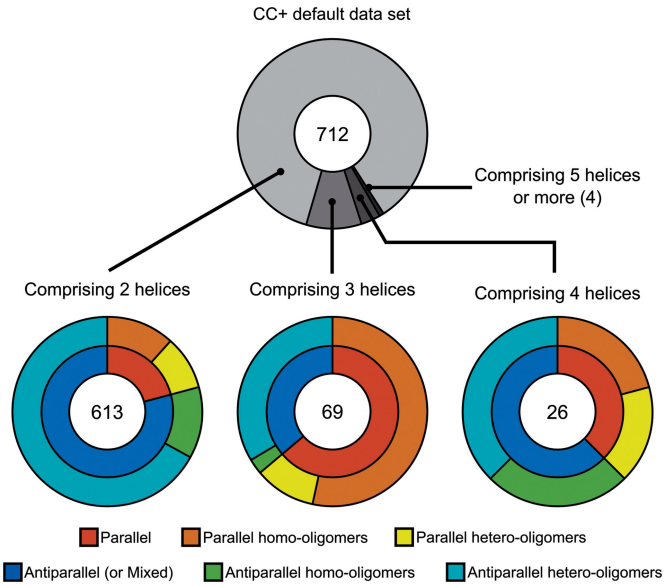

Figure 3. A summary of the coiled-coil composition of CC+. These doughnut charts illustrate the frequencies of different coiled-coil architectures returned using the default settings of the website's dynamic interface, i.e. canonical coiled coils with ≤50% sequence identity and α-helices longer than 11 amino acid residues. The number at the centre of each doughnut gives the number of assignments for each architecture (oligomer state). Note that coiled-coil architectures comprising five or more α-helices represent <1% of returned structures.

MySQL and PHP (LAMP) stack. CC+ has two interfaces: (i) a form-based 'Dynamic Interface' for customizable searching and (ii) a graphical interface for a 'Periodic Table of Coiled-coil Structures'.

The CC+ relational database comprises tables of coiled-coil data organized according distinct aspects of coiled-coil structure. The database is hierarchical, built on relationships between four Tables I 'proteins', II 'coiled coils', III 'amino-acid sequences' and IV 'individual amino acids': one protein may contain one or more coiled-coil assignments, and these assignments comprise two or more α-helical segments, the sequences of which contain many amino acids. Each table can be searched independently, or in combination with others, quickly to produce coiled-coil data scaling from relatively straightforward to complex queries seamlessly. Below, descriptions of example queries have sections (or tabs, shown in UPPERCASE); fields, i.e. the names parameters that can be specified (shown in *italics*); and arguments, i.e. entered keywords or selected values (shown in 'quotation marks').

KEYWORD SEARCHES—The first table in CC+ is dedicated to storing metadata from the PDB structural files containing coiled-coil assignments. Using this table, coiled-coil data sets can be gathered that incorporate or exclude terms associated with certain protein structures. This is implemented in the Dynamic Interface by the *keywords* field in the SPECIFY KEYWORDS tab. When using the *keyword* search to gather coiled-coil data sets, we recommend combining multiple terms, for example, searches using 'Leucine zipper', 'BZIP' and 'GCN4' together return a complete set of leucine-zipper motifs.

STRUCTURAL SEARCHES—Another table within CC+ is dedicated to storing data from the SOCKET output to describe the coiled-coil assignments in detail. Thus, searches using the SPECIFY STRUCTURES tab can restrict coiled-coil data sets to assignments: comprising a defined number of α-helices, with α-helices arranged in specific relative orientations, with α-helices of the same or different amino acid sequences (e.g. homo- and heterotypic interactions), with α-helices within the same or distinct polypeptide chains, with heptad or non-heptad repeats, and of a minimum number of residues in length. The default values specify: 'any' number of 'canonical' helices, greater than '11 residues' in length, in 'any' orientation, with 'any' type of partnering, from 'any' number of chains, and at '≤50%' sequence identity.

SEQUENCE REDUNDANCY—Many coiled-coil assignments identified and stored within the database have high sequence identity due to the size and repetitive

nature of coiled-coil heptads. Some proteins differ only outside the coiled-coil assignment; in the case of leucine zippers, for example, many coiled-coil sequences are mutations differing by only a few amino acids. Typically hydrophobic amino-acid residues at ***a*** and ***d*** positions, as well as similar polar residues at ***e*** and ***g***, serve to further increase the overall sequence identity between assigned coiled coils. To reduce coiled-coil assignments with high sequence identity, a method of culling redundant assignments has been implemented (Figure 4). Separate α-helical sequences were not simply subject to BLAST comparisons (29) as coiled-coil sequence redundancy is deemed a function of a whole coiled-coil assignment (Figure 4) not just of its individual amino-acid sequences.

SEQUENCE SEARCHES—A further table in CC+ is dedicated to amino-acid sequences and their SOCKET-assigned coiled-coil (typically heptad) registers. Searches using this table are implemented under the SPECIFY SEQUENCES tab and are specified using either plain text or PROSITE pattern syntax (30). There is also an option to specify heptad-register positions (e.g. combinations of ***a–g***).

INTERACTION SEARCHES—The final table of CC+ is dedicated to each individual amino acid from the SOCKET assignment. Using the SPECIFY INTERACTIONS tab, it is possible to identify coiled-coil assignments with individual amino acids, optionally at a specific register position, in a manner similar to sequence searches. More powerful searches can be conducted by specifying a second amino-acid residue, and optionally its register, that occurs within a certain distance cutoff of the first. This facilitates the compilation of coiled-coil data sets that contain potential residue–residue interactions for probing sequence-to-structure relationships in coiled coils in detail.

COMPLEX/LAYERED SEARCHES—It is important to note that searches are combined: all specified values are used cumulatively in searches, facilitating deeper and increasingly sophisticated interrogation of the CC+ Database.

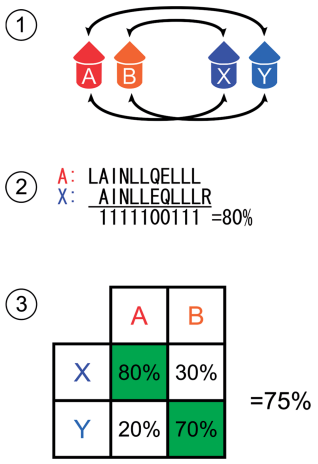

Figure 4. Scheme illustrating how coiled-coil sequence identity (redundancy) is calculated. (**1**) A coiled-coil assignment is only deemed redundant to another if their α-helices are configured identically, e.g. parallel dimers as in this case of coiled-coil AB, and coiled-coil XY. Every α-helical sequence of one assignment is compared to every α-helix in the other coiled coil being considered. (**2**) The sequences are compared using a sliding-window hashing algorithm. This records the maximum sequence identity of the overlap, which we call its redundancy. (**3**) With all pairwise comparisons complete, the highest scoring pairs (shown in green) are used to calculate an average sequence identity for the entire coiled-coil assignment.

Periodic Table of Coiled-coil Structures

A subset of the coiled coils in CC+ was selected with a maximum of 50% sequence identity. The protein structures were manually validated in order to obtain insight into the conformational space explored by coiled coils and to generate pristine sets of assignments applicable to structure prediction and protein engineering and design studies. This procedure included automated and manual methods and resulted in a classification system that we call the 'Periodic Table of Coiled-coil Structures', which will be described elsewhere (Moutevelis,E. and Woolfson,D.N., submitted for publication).

Downloadable resources

Data from analysing PDB and PQS files with SOCKET are available for download, including PYMOL (8) and RASMOL (31) scripts for visualization of identified coiled-coil assignments in more detail. Using these scripts, ribbon diagrams for α-helical segments are colour coded according to their assigned register, and interacting residues are grouped for easy identification, for example, constellations of residues involved in KIH packing can be selected. These scripts and their corresponding structural files have been made available to download and render locally using their respective programs. The PYMOL and RASMOL scripts are rendered automatically during compilation of the database to generate orthogonal views of each coiled-coil assignment. These high-resolution images are presented along with coiled-coil data and are also available for download. SOCKET outputs have also been made available.

Finally, the CC+ Database facilitates the compilation of coiled-coil data and its formatting into text-based outputs and tabulated PSSMs. Once a coiled-coil data set has been compiled, the user can select to output the data set in several text-based formats, described using simple tags similar to XML. If the available formats do not meet the requirements of the user, tags can be entered manually into the appropriate field, and the format can be updated to use the new specifications. In this way, users can quickly compile coiled-coil data sets and export them in a format that is easily incorporated into their own applications. If the user wishes to tabulate the propensities of amino acid residues for each position in an α-helical heptad, an option exists to create such PSSM tables, both of raw amino-acid frequencies and normalized against the amino-acid frequency in SWISS-PROT (32).

Updates

At the time of writing, the entries in the CC+ Database were based on the February 2008 release of the RCSB

PDB (6). We aim to update the entries, for dynamic interface at least, automatically on a monthly basis.

USING THE DATABASE

Inspection of the CC+ Database (Figure 3) reveals that it comprises predominantly antiparallel, two-helix, heteromeric assignments. Structures based on four helices typically also have mixed parallel/antiparallel topologies but are more evenly divided between homo- and heteromeric interactions. Coiled-coil assignments based on three helices, in contrast, are predominantly parallel, homomeric assemblies. Currently, there are too few structures with five or more helices for detailed comparisons.

These structures present a resource for interrogating coiled-coil structures rapidly and routinely. The CC+ Database allows subsets of coiled-coil assignments, together with their sequences and other information, to be garnered quickly from the RCSB PDB (6). The results of such searches provide the basis for developing sequence-to-structure relationships for this important class of protein-folding and protein–protein interaction motif. Such studies should improve our understanding of coiled coils, and our ability to predict, engineer and design them. To illustrate this potential, we outline a number of searches of the CC+ Database. These example queries, which were conducted using the interface illustrated in Figure 5, refer to sections (or tabs) using UPPERCASE, fields in *italics*, and entered keywords or selected values in 'quotation marks'. (The search results presented here refer to the CC+ Database current at the time of writing, which was drawn from the February 2008 release of then PDB and therefore, may change with updates of the Database.)

Specifying keywords—finding coiled coils within known proteins

At the simplest level, a subject protein structure can be tested to see if it contains a coiled-coil assignment(s) in CC+ by entering its PDB code into the *keywords* field of the SPECIFY KEYWORDS tab, and selecting the 'Search!' button.

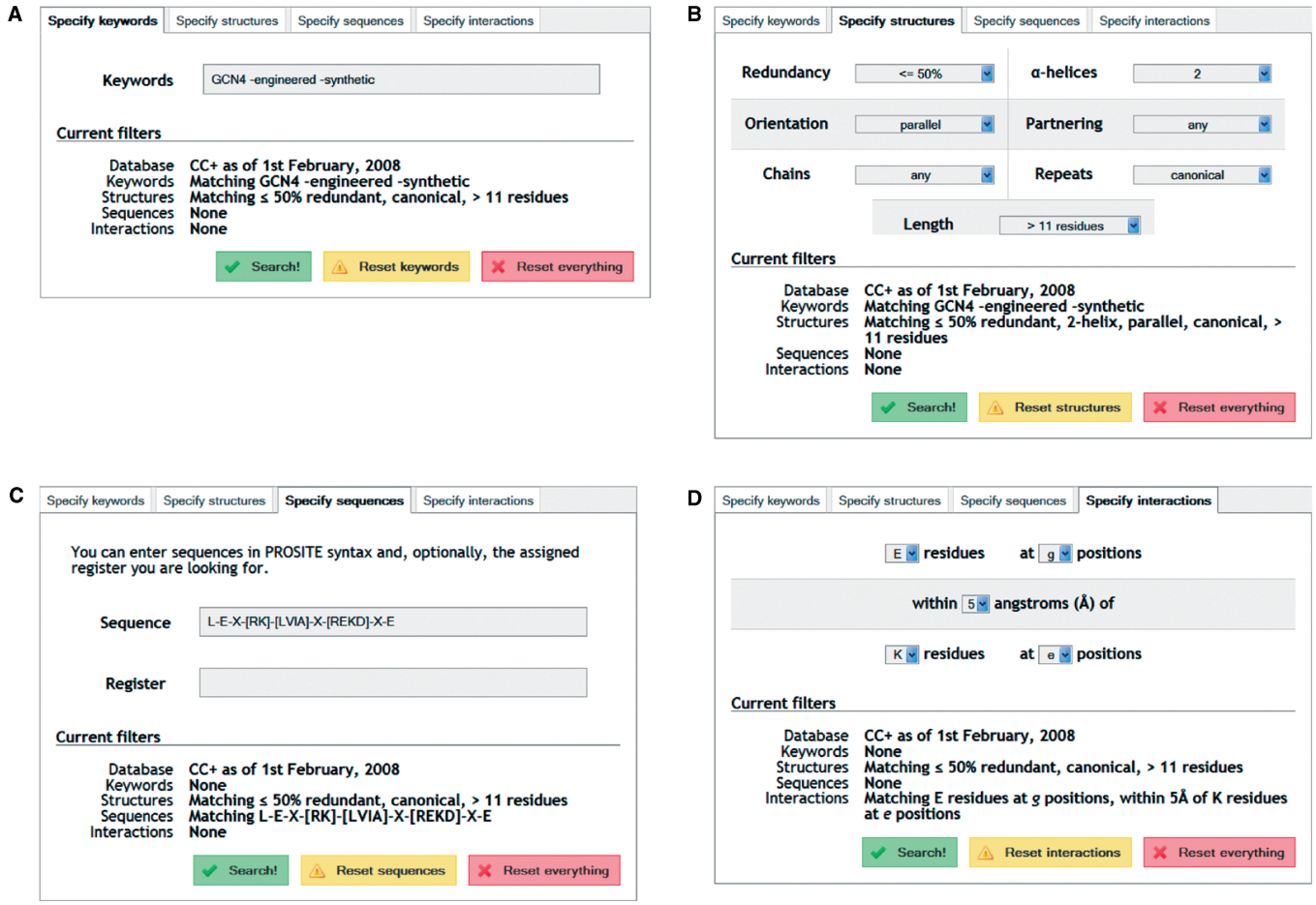

Figure 5. Interrogating CC+ via the Dynamic Interface. This is available through the 'search page' and provides a tabbed interface to the four main tables of CC+. (**A**) An example keyword search for finding coiled coils within known proteins. The 'Current filters' information is updated upon the selection of a new tab, or selecting the 'Search!' button. (**B**) An example structural search for parallel dimers. Note the keywords specified in A are still present, filtering data further. (**C**) An example sequence search using a PROSITE pattern; plain text is also accepted. (**D**) An example interaction search for potential salt bridges. In each case, selecting the 'Search!' button returns the data and downloadable resources or formatting returned assignments for export.

More sophisticated searches, for example, to gather leucine-zipper like motifs, and engineered assemblies based on those structures, can be achieved by entering *keywords* such as 'GCN4'. (This keyword was chosen because the archetypal leucine zipper, the PDB entry 2ZTA (9), is from the yeast transcriptional activator GCN4.) This returned 10 coiled-coil assignments, but these did not include 2ZTA itself as it is redundant to another structure. However, with the *redundancy* field of the SPECIFY STRUCTURES tab set to 'redundant', the search returned 67 coiled-coil assignments, including 2ZTA. Similarly, other fields in the SPECIFY STRUCTURES tab can be changed to gather more, or to focus on coiled coils with specific features. For example, with the *redundancy* field set back to its default ($\leq$50%), and the *repeats* field set to 'all' (to include both canonical and non-canonical sequence repeats), 14 assignments are returned with a search for the keyword 'GCN4'. Should the user wish to restrict the search, for instance in the GCN4 case to only natural coiled coils, the *keyword* search can be reconfigured to reject protein structures referred to as 'engineered' and 'synthetic'. This is done by specifying the keyword as 'GCN4 -engineered-synthetic'; in this case, three coiled-coil assignments were returned. It is important to note that restriction terms as presented here must be preceded by a desirable term; searches comprising exclusively undesired *keywords* are not possible.

Specifying structures I—compiling amino-acid profiles for given data sets

This example illustrates how PSSMs, or amino-acid profiles, can be compiled to compare and contrast two or more distinct coiled-coil types. The process of compiling a data set and generating a profile is conducted once for each targeted configuration, but simple changes to the filter criteria require minimal reconfiguration between searches. This example generates PSSMs for both dimeric and trimeric parallel coiled coils. We have chosen these as they are often targets for coiled-coil prediction algorithms, which currently draw on sequence rather than structurally derived data (19,21).

The SPECIFY STRUCTURES tab facilitates the selection of dimeric coiled-coil assignments as follows: setting the *α-helices* field to '2' configures searching for dimeric coiled coils; setting the *orientation* input to 'parallel' configures the remaining filter. In all, 129 coiled-coil assignments were returned.

Selecting the 'Format these coiled-coil data' button, at the top of the Dynamic Interface, links to a tool for exporting retrieved data in various ways. Selecting the 'Profiles' button from the list of preset formats compiles data into PSSMs. The raw and SWISS-PROT-normalized data for the matching coiled-coil assignments are then presented below the formatting tool, for copying into text or spreadsheet programs accordingly.

With the first PSSM saved, compilation of the next data set is straightforward. Returning to the Dynamic Interface restores the last search, and the 'Specify structures' tab, the dropdown input for 'α-helices' changed to '3'. In all, 44 coiled-coil assignments were returned in this case, and the new data can be formatted as above.

Specifying structures II—non-canonical coiled coils

SOCKET identifies KIH interactions between α-helices. By examining patterns within and between these KIH constellations, it assigns the ***a–g*** heptad register to the amino acid sequences. As a result, SOCKET can identify both canonical—that is, contiguous runs of ***abcdefg***, heptad repeats—and non-canonical coiled coils that incorporate non-heptad repeats. CC+ captures this information and, via the *repeats* field of the SPECIFY STRUCTURES tab, facilitates searches to return canonical assignments, non-canonical assignments or both.

The prevalence, importance and potential roles of non-canonical coiled-coil motifs are contemporary issues in coiled-coil research (33–36). With this in mind, CC+ was searched for examples of canonical and non-canonical coiled-coil assignments with $\leq$50% sequence redundancy. In all, 712 canonical and 50 non-canonical examples were returned. The latter included thirteen 4-residue inserts of ***abcd*** and seven ***defg*** between the ***g*** and ***a*** of otherwise canonical repeats; three 3-residue inserts (all ***abc***-type between ***g*** and ***a***); ten single-residue inserts (four ***a*** and six ***d***); and five reassignments of what would be ***g*** sites in a standard heptad repeat to ***d***-knobs. The remainder were irregular structures or artefacts with multiple consecutive residues assigned as ***a***- or ***d***-knobs; these were mainly short, antiparallel two-helix assignments within larger helical globules.

Specifying sequences—finding trigger motifs

A number of sequence motifs have been reported from experimental studies of various coiled-coil systems to be essential for coiled-coil folding, assembly, stabilization and specificity. These have been termed 'trigger motifs' (37,38). We performed searches to identify examples of such motifs within the coiled-coil assignments of CC+. This was done using the SPECIFY SEQUENCES tab and standard PROSITE syntax (30). Sequences from a non-identical data set—which can be selected by specifying 'non-identical' in the *redundancy* field of the SPECIFY STRUCTURES tab—were queried for two 9-residue patterns: L-E-X-[RK]-[LVIA]-X-[REKD]-X-E and L-E-X-E-[LVIA]-X-[REKD]-X-[RK], proposed to promote dimer formation (37).

The first pattern returned 11 sequences within five coiled-coil assignments, of which three were dimeric and two trimeric. Given that CC+ comprises more dimers than trimers (Figure 3) it is appropriate to normalize these data by the total number of non-identical dimeric and trimeric coiled-coil assignments, respectively. The resulting frequencies of trigger motifs are 0.4% (3/843) in dimers and 1.6% (2/127) in trimers. The second pattern returned 23 sequences within 11 coiled-coil assignments, of which 8 (0.9%) were in dimers and 3 (2.4%) in trimers.

More recently, the sequence R-[ILVM]-X-X-[ILV]-E has been proposed to be important in trimer specification (38). Searching CC+ for this returned 34 sequences, of

which 16 (1.9%) occur within dimers, 16 (11.8%) in trimers, and 2 (5%, 2/40) in tetramers.

Given that we found so few potential 'trigger motifs', our analyses do not provide strong support for the trigger hypothesis, at least not for such specific sequences within the structural assignments comprising CC+, though the third sequence, R-[ILVM]-X-X-[ILV]-E, does show some preference for trimers by ~2- to 6-fold.

Specifying interaction—finding interhelical salt bridges

This final example illustrates how potential salt bridges—which are also of contemporary interest in coiled-coil research (39)—can be identified between residues at ***g*** and ***e*** positions in the CC+ Database. Amino acid residue interactions can be returned from the database by applying filters from the SPECIFY INTERACTIONS tab. Here, parameters are strung together as part of a sentence. For example, we searched for any glutamic acid residues at ***g*** positions that occur within 5 Å of a lysine residue at an ***e*** position. (5 Å was chosen as a nominal cutoff for potential salt bridges, and the distances within CC+ are calculated between side chain centres of mass.) This is configured by selecting 'E' and '***g***' for the first residue parameters, '5' for the distance parameter, and 'K' and '***e***' for the second residue parameters. In all, 159 amino acid residues in 28 coiled-coil assignments were returned.

One of the coiled-coil assignments within this subset was for a heterodimeric leucine-zipper, PDB entry 1CI6 (40). Selecting its icon presents metadata from the PDB structural file, and a simple illustration of the amino acid sequence and its assigned register. Note that lysine 321 of chain A is highlighted as an interacting residue matching the above criteria. Downloading the appropriate script and corresponding PDB structure facilitates rendering of this coiled-coil assignment using RASMOL (31) or PYMOL (8). In this case, a tight interaction between the γ-carboxylate of E260 and the interaction with the ε-amino group of K321 is clear.

CONCLUSION AND FUTURE PLANS

The CC+ Database provides a straightforward, yet powerful, tool for probing the structures of and sequence-to-structure relationships in coiled coils. Indeed, we have already used it to provide bioinformatic data to support experimental studies of antiparallel dimeric coiled coils (41). The database has been designed to incorporate new structural data easily in order to remain up to date, via regular synchronization with other web-based databases. In future, we plan to supplement coiled-coil assignments with other data, for example, from sequence-based prediction tools such as COILS, PAIRCOIL2, MULTICOIL and MARCOIL (17,19,20,26,42). This will provide a source of coiled-coil data tailored to a wide range of individual user needs. We encourage feedback from our users to improve and expand the tool further.

ACKNOWLEDGEMENTS

We thank John Walshaw, Andrei Lupas, Richard Sessions, Marryat Stevens and members of the DNW group for many helpful discussions and feedback; Aisyah Syed Abdullah for helping with the studies of trigger motifs; and Chris Tothill for assistance with and maintenance of the web server.

FUNDING

BBSRC- and EPSRC-funded studentship (to O.D.T.); a BBSRC grant (BB/D003016/1 to D.N.W.). Funding for open access charge: UK BBSRC grant BB/D003016/1.

Conflict of interest statement. None declared.

REFERENCES

1. Orengo,C.A., Michie,A.D., Jones,S., Jones,D.T., Swindells,M.B. and Thornton,M. (1997) CATH—a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.
2. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
3. Finn,R.D., Tate,J., Mistry,J., Coggill,P.C., Sammut,S.J., Hotz,H.R., Ceric,G., Forslund,K., Eddy,S.R., Sonnhammer,E.L.L. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.
4. Rose,A., Manikantan,S., Schraegle,S.J., Maloy,M.A., Stahlberg,E.A. and Meier,I. (2004) Genome-wide identification of Arabidopsis coiled-coil proteins and establishment of the ARABI-COIL database. *Plant Physiol.*, **134**, 927–939.
5. Walshaw,J. and Woolfson,D.N. (2001) Socket: a program for identifying and analysing coiled-coil motifs within protein structures. *J. Mol. Biol.*, **307**, 1427–1450.
6. Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F. Jr., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.
7. Henrick,K. and Thornton,J.M. (1998) PQS: a protein quaternary structure file server. *Trends Biochem. Sci.*, **23**, 358–361.
8. DeLano, W.L. (2002) http://www.pymol.org.
9. O'Shea,E.K., Klemm,J.D., Kim,P.S. and Alber,T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, **254**, 539–544.
10. Lupas,A. (1996) Coiled coils: new structures and new functions. *Trends Biochem. Sci.*, **21**, 375–382.
11. Burkhard,P., Stetefeld,J. and Strelkov,S.V. (2001) Coiled coils: a highly versatile protein folding motif. *Trends Cell. Biol.*, **11**, 82–88.
12. Walshaw,J. and Woolfson,D.N. (2003) Extended knobs-into-holes packing in classical and complex coiled-coil assemblies. *J. Struct. Biol.*, **144**, 349–361.
13. Lupas,A.N. and Gruber,M. (2005) The structure of a-helical coiled coils. *Adv. Protein Chem.*, **70**, 37–78.
14. Woolfson,D.N. (2005) The design of coiled-coil structures and assemblies. *Adv. Protein Chem.*, **70**, 79–112.
15. Harbury,P.B., Zhang,T., Kim,P.S. and Alber,T. (1993) A switch between 2-stranded, 3-stranded and 4-stranded coiled coils in Gcn4 leucine-zipper mutants. *Science*, **262**, 1401–1407.
16. Liu,J., Zheng,Q., Deng,Y., Cheng,C.S., Kallenbach,N.R. and Lu,M. (2006) A seven-helix coiled coil. *Proc. Natl Acad. Sci. USA*, **103**, 15457–15462.
17. Lupas,A., Vandyke,M. and Stock,J. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.
18. Berger,B., Wilson,D.B., Wolf,E., Tonchev,T., Milla,M. and Kim,P.S. (1995) Predicting coiled coils by use of pairwise residue correlations. *Proc. Natl Acad. Sci. USA*, **92**, 8259–8263.

19. Wolf,E., Kim,P.S. and Berger,B. (1997) MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein Sci.*, **6**, 1179–1189.
20. Delorenzi,M. and Speed,T. (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics*, **18**, 617–625.
21. Woolfson,D.N. and Alber,T. (1995) Predicting oligomerization states of coiled coils. *Protein Sci.*, **4**, 1596–1607.
22. Apgar,J.R., Gutwin,K.N. and Keating,A.E. (2008) Predicting helix orientation for coiled-coil dimers. *Proteins*, **72**, 1048–1065.
23. Parry,D.A. (1982) Coiled-coils in alpha-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins. *Biosci. Rep.*, **2**, 1017–1024.
24. Conway,J.F. and Parry,D.A.D. (1990) Structural features in the heptad substructure and longer range repeats of 2-stranded alpha-fibrous proteins. *Int. J. Biol. Macromol.*, **12**, 328–334.
25. Conway,J.F. and Parry,D.A.D. (1991) 3-Stranded alpha-fibrous proteins—the heptad repeat and its implications for structure. *Int. J. Biol. Macromol.*, **13**, 14–16.
26. Gruber,M., Soding,J. and Lupas,A.N. (2006) Comparative analysis of coiled-coil prediction methods. *J. Struct. Biol.*, **155**, 140–145.
27. Crick,F. (1953) The packing of a-helices: simple coiled-coils. *Acta Crystallogr.*, **6**, 689–697.
28. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
29. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
30. Hulo,N., Bairoch,A., Bulliard,V., Cerutti,L., Cuche,B.A., de Castro,E., Lachaize,C., Langendijk-Genevaux,P.S. and Sigrist,C.J. (2008) The 20 years of PROSITE. *Nucleic Acids Res.*, **36**, D245–D249.
31. Sayle,R.A. and Milner-White,E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.
32. Boeckmann,B., Bairoch,A., Apweiler,R., Blatter,M.C., Estreicher,A., Gasteiger,E., Martin,M.J., Michoud,K., O'Donovan,C., Phan,I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **31**, 365–370.
33. Brown,J.H., Cohen,C. and Parry,D.A. (1996) Heptad breaks in a-helical coiled coils: stutters and stammers. *Proteins*, **26**, 134–145.
34. Hicks,M.R., Holberton,D.V., Kowalczyk,C. and Woolfson,D.N. (1997) Coiled-coil assembly by peptides with non-heptad sequence motifs. *Fold. Des.*, **2**, 149–158.
35. Hicks,M.R., Walshaw,J. and Woolfson,D.N. (2002) Investigating the tolerance of coiled-coil peptides to nonheptad sequence inserts. *J. Struct. Biol.*, **137**, 73–81.
36. Gruber,M. and Lupas,A.N. (2003) Historical review: another 50th anniversary—new periodicities in coiled coils. *Trends. Biochem. Sci.*, **28**, 679–685.
37. Kammerer,R.A., Schulthess,T., Landwehr,R., Lustig,A., Engel,J., Aebi,U. and Steinmetz,M.O. (1998) An autonomous folding unit mediates the assembly of two-stranded coiled coils. *Proc. Natl Acad. Sci. USA*, **95**, 13419–13424.
38. Kammerer,R.A., Kostrewa,D., Progias,P., Honnappa,S., Avila,D., Lustig,A., Winkler,F.K., Pieters,J. and Steinmetz,M.O. (2005) A conserved trimerization motif controls the topology of short coiled coils. *Proc. Natl Acad. Sci. USA*, **102**, 13891–13896.
39. Meier,M. and Burkhard,P. (2006) Statistical analysis of intrahelical ionic interactions in a-helices and coiled coils. *J. Struct. Biol.*, **155**, 116–129.
40. Podust,L.M., Krezel,A.M. and Kim,Y. (2001) Crystal structure of the CCAAT box/enhancer-binding protein beta activating transcription factor-4 basic leucine zipper heterodimer in the absence of DNA. *J. Biol. Chem.*, **276**, 505–513.
41. Hadley,E.B., Testa,O.D., Woolfson,D.N. and Gellman,S.H. (2008) Preferred side-chain constellations at antiparallel coiled-coil interfaces. *Proc. Natl Acad. Sci. USA*, **105**, 530–535.
42. McDonnell,A.V., Jiang,T., Keating,A.E. and Berger,B. (2006) Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics*, **22**, 356–358.