Bioinformatics, 33(12), 2017, 1889–1891
doi: 10.1093/bioinformatics/btx085
Advance Access Publication Date: 10 February 2017
Applications Note

OXFORD

Structural bioinformatics

FELLS: fast estimator of latent local structure

Damiano Piovesan[1], Ian Walsh[1,2], Giovanni Minervini[1] and Silvio C.E. Tosatto[1,3,*]

[1]Department of Biomedical Sciences, University of Padua, Padova 35121, Italy, [2]Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), Singapore, Singapore and [3]CNR Institute of Neuroscience, Padova 35121, Italy

*To whom correspondence should be addressed.
Associate Editor: Alfonso Valencia

Received on November 7, 2016; revised on January 4, 2017; editorial decision on February 4, 2017; accepted on February 4, 2017

Abstract

Motivation: The behavior of a protein is encoded in its sequence, which can be used to predict distinct features such as secondary structure, intrinsic disorder or amphipathicity. Integrating these and other features can help explain the context-dependent behavior of proteins. However, most tools focus on a single aspect, hampering a holistic understanding of protein structure. Here, we present Fast Estimator of Latent Local Structure (FELLS) to visualize structural features from the protein sequence. FELLS provides disorder, aggregation and low complexity predictions as well as estimated local propensities including amphipathicity. A novel fast estimator of secondary structure (FESS) is also trained to provide a fast response. The calculations required for FELLS are extremely fast and suited for large-scale analysis while providing a detailed analysis of difficult cases.
Availability and Implementation: The FELLS web server is available from URL: http://protein.bio.unipd.it/fells/. The server also exposes RESTful functionality allowing programmatic prediction requests. An executable version of FESS for Linux can be downloaded from URL: protein.bio.unipd.it/download/.
Contact: silvio.tosatto@unipd.it
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Protein structure is encoded in its amino acid sequence. This simple paradigm has been at the heart of protein science for decades. At the lowest structural level, secondary structure is determined by the propensity of amino acids to form regular α-helices and β-strands (Chou and Fasman, 1974). More recently, a large portion of the polypeptide chain in higher organisms has been identified to be intrinsically disordered (ID) (Potenza *et al.*, 2015). Similarly, specific residues have been suggested to promote specific behavior, such as amyloid aggregation (Walsh *et al.*, 2014), hydrophobic clusters (Faure and Callebaut, 2013) or amphipathicity (Eisenberg *et al.*, 1982). More recently, there has been a growing realization that ID proteins in particular may undergo interesting phase transitions depending on their amino acid composition (Wu and Fuxreiter, 2016). An early classification of ID proteins distinguishes five main categories based on their content of charged residues (Das and Pappu, 2013), which has been recently applied to a large-scale analysis (Necci *et al.*, 2016). Here, we present a novel tool, Fast Estimator of Latent Local Structure (FELLS), to systematically visualize several computed aspects of the protein sequence in a user-friendly way.

2 Implementation

FELLS aggregates structural predictions and sequence propensities in a single view. A novel fast secondary structure predictor (FESS) has been developed based on the same single-sequence neural network architecture as ESpritz (Walsh *et al.*, 2012) and provides a similar speed vs. accuracy tradeoff. A full description of the method training and benchmarking is provided in the Supplementary Material. Signal peptides and transmembrane segments are predicted with Phobius (Käll *et al.*, 2004) and intrinsic disorder with ESpritz-NMR (Walsh *et al.*, 2012). The SEG algorithm for sequence

complexity (Wootton and Federhen, 1996) and hydrophobic cluster analysis (Faure and Callebaut, 2013) have been implemented in house. Aggregation propensity is predicted by Pasta 2.0 (Walsh *et al.*, 2014). Amino acid propensities along the sequence are averaged over a window of 7 residues. Single residue hydrophobicity is weighted by the normalized Kyte-Doolittle scale, while positively (K, H, R) and negatively (E, D) charged residues contribute equally. Amphipathicity provides the propensity of a residue to be part of a fragment in which hydrophobic (or charged) residues are separated into opposite surfaces, especially for α-helices (Eisenberg *et al.*, 1982). We have generalized the concept to include both hydrophobic and charged amphipathicity for α-helices and β-strands (see Supplementary Material for details). The FELLS web server is implemented using the REST (Representational State Transfer) architecture, allowing its services to be accessed both from a web interface and programmatically from URL: protein.bio.unipd.it/fells/. The FELLS output is plotted dynamically on the website and allows users to navigate the sequence prediction exploiting the functionality of the Plotly.js library (URL: plot.ly/javascript/).

3 Web server overview

The main FELLS page features a search box, which accepts any valid multi FASTA string or, alternatively, the input can be uploaded as multi FASTA file. Input size is limited to 10 000 sequences. The submitted job can be retrieved at a later time by providing the session identifier or the URL to the result page. Predictions are stored permanently in a database where entries are indexed by their sequence in order to speed up the service when requesting a cached protein. Moreover FELLS runs in parallel and single sequence results appear in the result page independently as soon as processed. Figure 1 shows FELLS output for CDKN1B (cyclin-dependent kinase

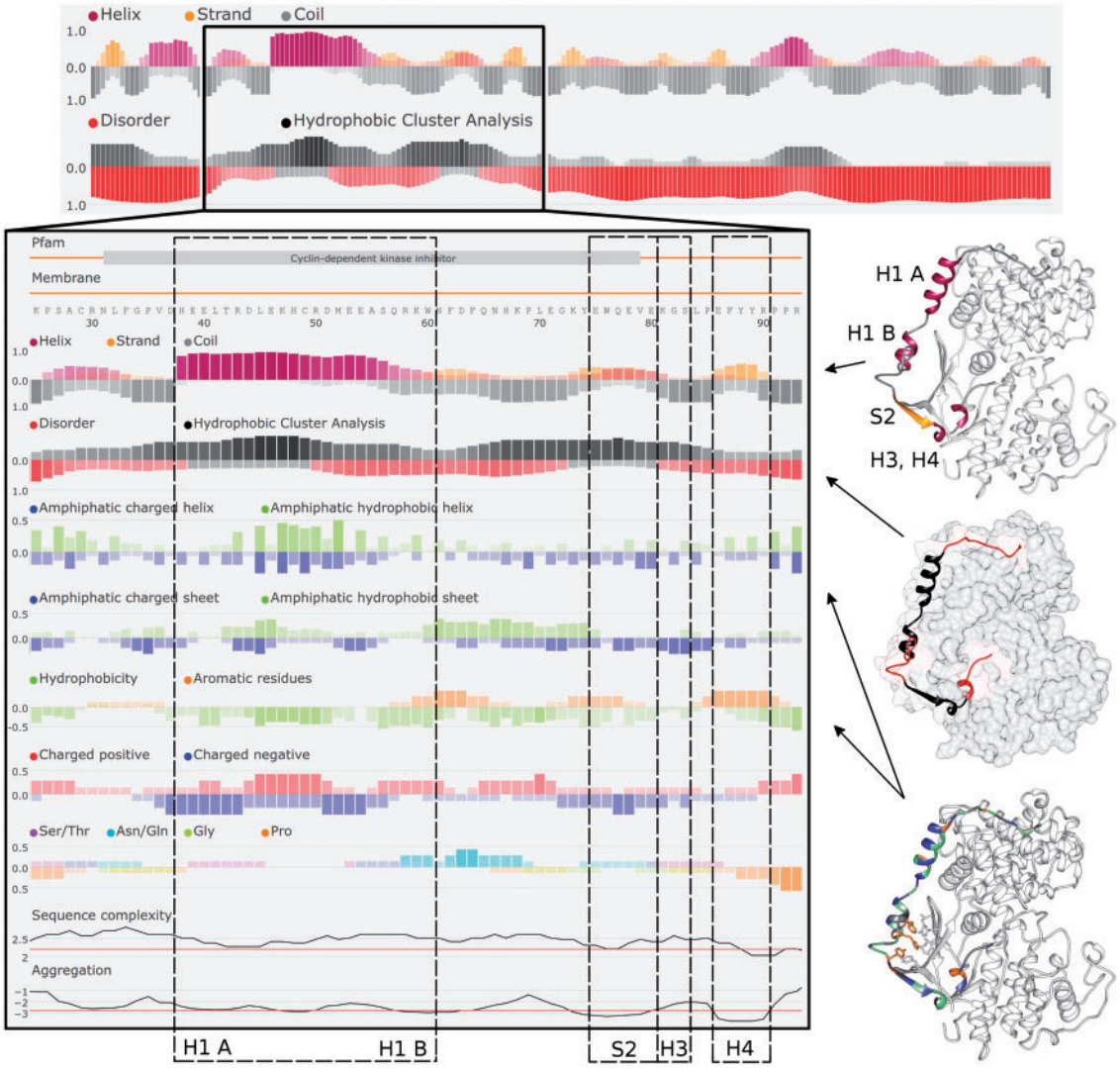

Fig. 1. FELLS usage example with CDKN1B (cyclin-dependent kinase inhibitor 1B). The output for the secondary structure and disorder output for the CDKN1B sequence is shown on the top part. The central part is a close-up of the full FELLS output for the crystallized part (residues 25–93) of the CDKN1B structure bound to the cyclinA-Cdk2 complex (PDB code: 1JSU). The CDKN1B structure is shown right with labels for the secondary structure elements and different FELLS features highlighted. The eleven FELLS tracks are, from top to bottom: Pfam domains and query sequence, signal peptides and transmembrane segments from Phobius, secondary structure predicted with FESS, intrinsic disorder predicted with ESpritz-NMR and hydrophobic clusters, amphipathic helices, amphipathic sheets, hydrophobicity and aromatic residues, charge clusters, special amino acid clusters. SEG sequence complexity and PASTA aggregation are shown as continuous lines with lower threshold in red. Color codes are shown separately in each graph and boxes highlight the secondary structure elements labeled in the CDKN1B structure (Color version of this figure is available at *Bioinformatics* online.)

inhibitor 1B), an intrinsically disordered protein (Piovesan *et al.*, 2016) characterized by an extended conformation of consecutive secondary structure elements not interacting with each other. It is an important kinase inhibitor involved in the regulation of cell cycle progression (Polyak *et al.*, 1994) and mutation of CDKN1B is causative of cancer progression (Philipp-Staheli *et al.*, 2003; Russo *et al.*, 1996). FELLS predicts five distinct α-helices (13–21, 27–31, 38–59, 141–149, 163–173) and, with lower confidence, five short β-strands at the N-terminus (74–131). CDKN1B is also predicted to be largely disordered and with a hydrophobic cluster at position 36–85. The detailed characterization of amino acid composition and helix amphipathicity suggests this region to be a protein-protein interaction interface. Inspection of the crystal structure of CDKN1B bound to the cyclinA-Cdk2 complex (PDB code: 1JSU) confirms the hydrophobic cluster corresponds to an extended secondary structure-rich segment interacting with both cyclinA and Cdk2 (top and middle structure in Fig. 1). The high amphipathicity of the larger helix correctly identifies the segregation of charged resides (exposed to the solvent) and hydrophobic amino acids interacting with the complex (green and blue residues in the bottom structure of Fig. 1). Notably, FELLS also highlights a cluster of aromatic residues. In the structure these interact with an affine pocket (orange residues in the figure). Overall, the FELLS results derived from the CDKN1B sequence are in excellent agreement with the crystal structure, suggesting it can be a very useful tool to analyze ID proteins. In the future, we plan to integrate it into the MobiDB database (Potenza *et al.*, 2015) of protein disorder annotations.

Funding

Fondazione Italiana per la Ricerca sul Cancro [16621] to D.P.; Associazione Italiana per la Ricerca sul Cancro [IG17753] to S.T. ELIXIR-IIB (elixir-italy.org), the Italian Node of the European ELIXIR bioinformatics infrastructure, is acknowledged for supporting the development and maintenance of MobiDB lite.

Conflict of Interest: none declared.

References

Chou,P.Y. and Fasman,G.D. (1974) Conformational parameters for amino acids in helical, beta-sheet, and random coil regions calculated from proteins. *Biochemistry (Mosc.)*, **13**, 211–222.

Das,R.K. and Pappu,R.V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 13392–13397.

Eisenberg,D. *et al.* (1982) The helical hydrophobic moment: a measure of the amphiphilicity of a helix. *Nature*, **299**, 371–374.

Faure,G. and Callebaut,I. (2013) Comprehensive repertoire of foldable regions within whole genomes. *PLoS Comput. Biol.*, **9**, e1003280.

Käll,L. *et al.* (2004) A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.*, **338**, 1027–1036.

Necci,M. *et al.* (2016) Large-scale analysis of intrinsic disorder flavors and associated functions in the protein sequence universe. *Protein Sci. Publ. Protein Soc.*, **25**, 2164–2174.

Philipp-Staheli,J. *et al.* (2003) Distinct roles for p53, p27Kip1, and p21Cip1 during tumor development. *Oncogene*, **23**, 905–913.

Piovesan,D. *et al.* (2016) DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.*, **45**, D219–D227.

Polyak,K. *et al.* (1994) Cloning of p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals. *Cell*, **78**, 59–66.

Potenza,E. *et al.* (2015) MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Res.*, **43**, D315–D320.

Russo,A.A. *et al.* (1996) Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature*, **382**, 325–331.

Walsh,I. *et al.* (2012) ESpritz: accurate and fast prediction of protein disorder. *Bioinformatics*, **28**, 503–509.

Walsh,I. *et al.* (2014) PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.*, **42**, W301–W307.

Wootton,J.C. and Federhen,S. (1996) Analysis of compositionally biased regions in sequence databases. *Methods Enzymol.*, **266**, 554–571.

Wu,H. and Fuxreiter,M. (2016) The structure and dynamics of higher-order assemblies: amyloids, signalosomes, and granules. *Cell*, **165**, 1055–1066.