Bioinformatics, 32(5), 2016, 792–794
doi: 10.1093/bioinformatics/btv637
Advance Access Publication Date: 30 November 2015
Applications Note

Structural bioinformatics

Fit3D: a web application for highly accurate screening of spatial residue patterns in protein structure data

Florian Kaiser[†,*], Alexander Eisold[†], Sebastian Bittrich and Dirk Labudde

Department of Applied Computer and Biosciences, University of Applied Sciences Mittweida, 09648 Mittweida, Germany

*To whom correspondence should be addressed.
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Anna Tramontano

Received on August 3, 2015; revised on September 25, 2015; accepted on October 24, 2015

Abstract

Summary: The clarification of linkage between protein structure and function is still a demanding process and can be supported by comparison of spatial residue patterns, so-called structural motifs. However, versatile up-to-date resources to search for local structure similarities are rare. We present Fit3D, an easily accessible web application for highly accurate screening of structural motifs in 3D protein data.
Availability and implementation: The web application is accessible at https://biosciences.hs-mittweida.de/fit3d and program sources of the command line version were released under the terms of GNU GPLv3. Platform-independent binaries and documentations for offline usage are available at https://bitbucket.org/fkaiser/fit3d.
Contact: florian.kaiser@hs-mittweida.de
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Background

The identification and thorough comprehension of spatial residue patterns—so-called structural motifs—can undoubtedly help to bridge the gap between protein structure and function. These patterns are often highly conserved and not necessarily dependent on primary structure due to occasionally large sequence insertions between motif residues. The close proximity of structural motif residues in tertiary structure forms essential 'building blocks' for synergy between protein structure and function (Hedstrom, 2002; Koutsotoli *et al.*, 2012).

Structural motifs are involved in the majority of biochemical processes, for example in catalytic activity of enzymes (Hedstrom, 2002). Additionally, they were observed to aid structure stabilization (Koutsotoli *et al.*, 2012). Even highly divergent protein superfamilies such as the enolase superfamily can be represented adequately by structural motif templates (Meng *et al.*, 2004).

Although protein structure is resilient to mutation, small structural changes can have a drastic impact on protein function (Samish *et al.*, 2015). Such small alterations are not yet fully understood and therefore the detailed investigation of spatial residue patterns is beneficial to understand complex implications.

Nowadays the contemporary and rapid increase of structure-related data results in the release of protein structures prior to correct functional annotation. Consequently, computational solutions for large-scale screening and analysis of diverse structural motifs are experiencing a high demand and are of significant interest for annotation of protein function. The Structure Function Linkage Database (Akiva *et al.*, 2013) or the Catalytic Site Atlas (Furnham *et al.*, 2014) are suitable starting points to deduce function representing spatial residue patterns. Furthermore, computational protein design approaches are often simplifying atomic structure, which can result

in misplaced sidechain orientations (Samish *et al.*, 2015). Here, the *in silico* comparison of sidechain placements with observations in experimentally determined structures can be helpful to appraise or improve design results.

The detection of local similarities in protein structures can be seen as a pattern matching problem in 3D space and simultaneous consideration of geometrical constraints—such as maximum allowed least root-mean-square deviation (LRMSD) to the query motif—as well as residue composition of match candidates. To date various algorithmic approaches were presented (for a brief overview see Supplementary Table S1), but most of them are either not accessible, difficult in usage or parametrization, deprecated, or limited in functionality. More specifically, existing software or web services lack at least one of the following primary features:

- convenient definition of structural motifs,
- consideration of position-specific exchanges (PSEs) of residues (Kaiser *et al.*, 2015),
- representation of structural motifs by C_α-, backbone- or side-chain-only atoms,
- detection of inter-molecular matches located in multiple protein chains,
- interactive visualization of the results or
- (if available) annotation of hits with EC numbers and Pfam accession codes to assess function or family affiliation.

Recently we developed an algorithm for motif screening and utilized it to demonstrate that multi-atom representation of motifs, based on a definable selection, is essential for proper screening results (Kaiser *et al.*, 2015). The presented method significantly outperforms existing approaches when identifying enolase superfamily members in terms of sensitivity and specificity. Furthermore, algorithmic capabilities are extended by allowing matching of non-identical residues to mimic residue substitutions, which might have occurred during protein evolution or were directly induced by mutagenesis or protein design approaches.

Consequently, the presented method tackles these limitations and aims at the combination and extension of advantages of existing methods to enable powerful analyses of structural motifs.

2 Functionality

Based on the previously published matching algorithm (Kaiser *et al.*, 2015) we designed an intuitive web application called Fit3D. Furthermore, we provide a platform-independent command line version that allows more sophisticated calculations and extended features for easy integration into existing processing pipelines. For an auxiliary overview of the basic concept please refer to Figure 1.

Fit3D offers an 'all-in-one' solution, covering the whole process of a structural motif screening pipeline from query motif definition (Fig. 1A) and screening (Fig. 1B), up to retrieval and visualization of the results (Fig. 1C). Compared to previous presented work, our parameter-independent web application offers several improvements: the definition of PSEs, different atom representations, the detection of intra- and inter-molecular motif occurrences, and interactive visualization of the results within the web browser (see Fig. 1). Besides matching, Fit3D provides a statistical assessment of the results by a *P*-value estimation (Fofanov *et al.*, 2008; Stark *et al.*, 2003).

The mapping of meta information contributes to fast selection of relevant matches. Consequently, structures where matches occurred are annotated with EC number or Pfam accession codes (if available) and cross-referenced to the Protein Data Bank (PDB) (Berman *et al.*, 2000).

For proper visualization of the results the Protein Viewer (Biasini, 2014) was implemented. Different visualization types of matches can be picked: alignment of all matches against the query motif, alignment of a single match against the query motif, or visualization of a single match in the structure of origin.

All results are provided as individual files or an archive file, bundling all matches in a column-separated format (CSV) and their corresponding structures in PDB format. The distribution of LRMSD values of all matches, which can be an important signature pattern depending on characteristics of the query motif, is computed and available for download.

Beside algorithmic efficiency the usability was a primary aspect during development. The web application is optimized for highly accurate screening of small (usually $\leq$ five residues) and compact (close spatial proximity) structural motifs and a filter is applied to estimate the complexity of the query motif. For a three residue motif the whole process takes only minutes, but the integrated email notification allows the user to recall the results within 72 hours after submission.

Summarizing the above, Fit3D can be valuable for versatile applications, offers an ample set of features and combines advantages of existing methods. The usability and intuitive usage stands out and facilitates quick acquisition of results.

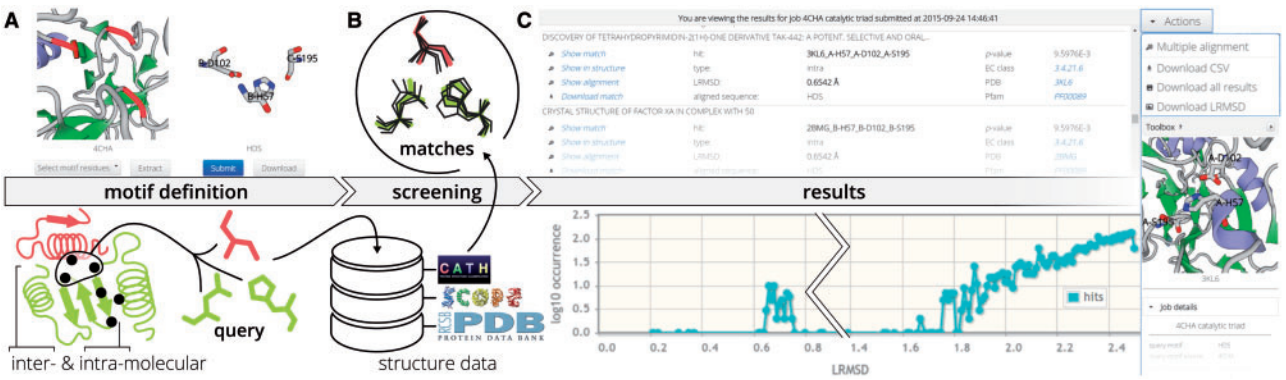

Fig. 1. Fit3D screening pipeline shown for the catalytic triad of serine proteases (Hedstrom, 2002). (**A**) Based on structure data (uploaded PDB file or given PDB-ID), the motif extraction wizard allows inter- or intra-molecular motif definition and extraction. Individual chains of the source structure are highlighted in red and green. (**B**) Screening of the query motif in non-redundant versions of protein structure databases returns a set of matches. Allowed PSEs (Kaiser *et al.*, 2015) can be defined for each individual residue of the query motif. **(C)** The comprehensive application output features detailed results for each match. The distribution of all LRMSD values is visualized and available for download. Different interactive visualization types are selectable and all results can be downloaded as a single archive file, which contains results in comma-separated value format and structures of matches in PDB format

Acknowledgement

We would like to thank our colleagues, especially Christoph Leberecht and Steffen Grunert, for extensively testing the web interface and providing valuable feedback.

Funding

Research and publication funding: Europäischer Sozialfonds (ESF), Free State of Saxony.

Conflict of Interest: none declared.

References

Akiva,E. *et al*. (2013) The structure–function linkage database. *Nucleic Acids Res.*, **42**, D521–D530.

Berman,H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.

Biasini,M. (2014) PV—WebGL-based protein viewer. *Zenodo*, doi:10.5281/zenodo.12620.

Fofanov,V.Y. *et al*. (2008) A statistical model to correct systematic bias introduced by algorithmic thresholds in protein structural comparison algorithms. *Bioinf. Biomed. Workshops*, 1–8. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4686202.

Furnham,N. *et al*. (2014) The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.*, **42**, D485–D489.

Hedstrom,L. (2002) Serine protease mechanism and specificity. *Chem. Rev.*, **102**, 4501–4524.

Kaiser,F. *et al*. (2015) A novel algorithm for enhanced structural motif matching in proteins. *J. Comput. Biol.*, **22**, 698–713.

Koutsotoli,A. and Tzakos,A.G. (2012) Host-pathogen crosstalking: the mastery of taking the helm of the host. *Structure*, **20**, 1613–1615.

Meng,E.C. *et al*. (2004) Superfamily active site templates. *Proteins*, **55**, 962–976.

Samish,I. *et al*. (2015) Achievements and challenges in structural bioinformatics and computational biophysics. *Bioinformatics*, **31**, 146–150.

Stark,A. and Russel,R.B. (2003) A model for statistical significance of local similarities in structure. *J. Mol. Biol.*, **326**, 1307–1316.