OXFORD

Structural bioinformatics

3DBIONOTES v2.0: a web server for the automatic annotation of macromolecular structures

Joan Segura*, Ruben Sanchez-Garcia, Marta Martinez, Jesus Cuenca-Alba, Daniel Tabas-Madrid, C.O.S. Sorzano and J.M. Carazo*

GN7 of the Spanish National Institute for Bioinformatics (INB), Biocomputing Unit, National Center of Biotechnology (CSIC), Instruct Image Processing Center, Madrid 28049, Spain

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

Abstract

Motivation: Complementing structural information with biochemical and biomedical annotations is a powerful approach to explore the biological function of macromolecular complexes. However, currently the compilation of annotations and structural data is a feature only available for those structures that have been released as entries to the Protein Data Bank.

Results: To help researchers in assessing the consistency between structures and biological annotations for structural models not deposited in databases, we present 3DBIONOTES v2.0, a web application designed for the automatic annotation of biochemical and biomedical information onto macromolecular structural models determined by any experimental or computational technique.

Availability and implementation: The web server is available at http://3dbionotes-ws.cnb.csic.es.

Contact: jsegura@cnb.csic.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Structural biology is a fundamental tool to further understand the mechanisms that control protein functions. Experimental and computational techniques for structure determination are continuously evolving and new macromolecular structures are submitted every day. In much the same way, the amount of biochemical and biomedical data available for genes and proteins grow rapidly and new databases appear every year. In this context, the mapping and analysis of biomedical and biochemical knowledge onto the residues of a newly proposed structural model constitutes a great help for a proper understanding of its function and cell role. This task was greatly facilitated by our former published web application 3DBIONOTES (Tabas-Madrid *et al.*, 2016); however, the tool was only available for those structures that were already released in structural databases. Then, if a structural model was still under analysis by the researcher and, consequently, had not yet been submitted to any structural database, there was no automatic manner to generate a proper compilation of the annotations associated with this model.

Answering the specific need analysed before, in this work we present a new version of 3DBIONOTES designed to submit structural models directly to automatically annotate them. Additionally, the range of annotations currently handled by the application has increased significantly, with a new panel on genomics variants organized by pathologies. Further to other tools like (O'Donoghue *et al.*, 2015; Stank *et al.*, 2016), 3DBIONOTES v2.0 has been designed for the automatic mapping of biomedical and biochemical information onto new structures and it integrates a large collection of annotation sources (see Supplementary Material Section S1).

3655

2 Methods

2.2 The web server

The web server has been implemented using the Ruby on Rails application framework (http://rubyonrails.org). The server performs three major tasks: first, it identifies the UniProt accessions of the different subunits contained in the submitted structure; second, it aligns the sequences of the subunits with their corresponding UniProt entries; and finally, it maps the protein annotations. Supplementary Figure S1 shows a schema of the communication and data processing between client and server. When a macromolecular structure is submitted, the server performs a BLAST search (Boratyn *et al.*, 2012) against all UniProt sequences to identify its different protein subunits; then, the best hits are sent back to the client and the user is asked to select the corresponding UniProt accessions for each chain of the submitted structure. Once each chain is identified, this information is returned to the server and the Smith–Waterman algorithm (Smith and Waterman, 1981) is used to align the sequence chains with their respective UniProt sequences. Finally, the server collects the biochemical and biomedical data from the different sources of information and the submitted structure is annotated and returned to the client.

2.3 The client

The web client provides an interactive environment linking protein sequences, structures and annotations. The client comprises three major panels (Supplementary Fig. S2): the structural panel, the annotation panel and the sequence panel. The structural panel uses the NGL viewer (Rose and Hildebrand, 2015) to display protein structures and cryo Electron Microscopy maps. The annotation panel was built using a bespoke version of the UniProt annotation viewer (Watkins *et al.*, 2017). The sequence panel shows the alignment between the sequence chains of the structure and their respective UniProt sequences. The alignment was fitted by using a modified version of the BioJS 'Sequence' package (Gomez and Jimenez, 2014). All panels are interconnected, allowing graphic interactivity among them. Thus, selections in the annotation or sequence panels are simultaneously highlighted in the three panels.

3 Human AKT1/PIN1 interaction

In this example, we illustrate how 3DBIONOTES v2.0 can be used together with protein structural docking to analyse and eventually select models consistent with other sources of biological knowledge. To model the structure of the AKT1/PIN1 interaction complex, we generated 50 potential models using GRAMM-X docking web server (Tovchigrechko and Vakser, 2006). Then, we used 3DBIONOTES v2.0 to visualize how well biological annotations matched to the different solutions.

Among the retrieved biological annotations for the AKT1 protein, we analysed the short linear motifs (SLiMs) (Supplementary Fig. S2B, 'Domains & sites' section). SLiMs are short conserved segments of residues involved in the targeting and recognition of other macromolecules, which mediate many protein–protein interactions. By clicking the second SLiM, a panel that shows the SLiM information from Eukaryotic Linear Motifs database (EML DB) (Dinkel *et al.*, 2016) is displayed. According to this information, we found that AKT1 SLiM, comprised between residues I447 and P452 (Fig. 1, purple spheres), may interact with modular protein domains of type WW (Aragon *et al.*, 2011). Given that PIN1 protein contains a WW domain between L7 and P37 residues (Fig. 1, pink spheres) we explored all 50 docking models searching for solutions that involved contacts
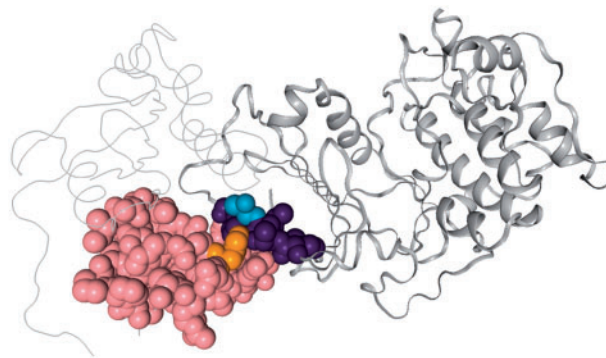

Fig. 1. Docking model of the AKT1/PIN1 human complex. In purple colour the AKT1 SLiM region from I447 to P452 residue. In blue colour the AKT1 phosphorylation site T450. In pink colour the PIN1 protein domain WW comprised between residues L7 and P37. In orange colour the W34 PIN1 residue; note that mutation W34A disrupts interaction of PIN1 with phosphorylated proteins

between the I447-P452 region of AKT1 and the L7-P37 residues of PIN1 WW domain. Noteworthy, model number 33, displayed in Figure 1, was the unique solution that satisfied this restraint. Other relevant annotations showed that phosphorylation site T450 of AKT1 (Fig. 1, blue spheres) is involved in the interaction with PIN1 protein (Liao *et al.*, 2009) and mutations on the W34 PIN1 residue (Fig. 1, orange spheres) disrupt interactions with other proteins (Min *et al.*, 2012). A more detailed analysis of the biochemical annotation along with another example is available in Supplementary Material. This use case illustrates how 3DBIONOTES v2.0 associates biological annotations to macromolecular structures, helping to select the interaction model that better fits those annotations.

References

Aragon, E. *et al.* (2011) A Smad action turnover switch operated by WW domain readers of a phosphoserine code. *Genes Dev.*, **25**, 1275–1288.

Boratyn, G.M. *et al.* (2012) Domain enhanced lookup time accelerated BLAST. *Biol. Direct*, **7**, 12.

Dinkel, H. *et al.* (2016) ELM 2016–data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Res.*, **44**, D294–D300.

Gomez, J., and Jimenez, R. (2014) Sequence, a BioJS component for visualising sequences. *F1000Res.*, **3**, 52.

Liao, Y. *et al.* (2009) Peptidyl-prolyl cis/trans isomerase Pin1 is critical for the regulation of PKB/Akt stability and activation phosphorylation. *Oncogene*, **28**, 2436–2445.

Min, S.H. *et al.* (2012) Negative regulation of the stability and tumor suppressor function of Fbw7 by the Pin1 prolyl isomerase. *Mol. Cell*, **46**, 771–783.

O'Donoghue, S.I. *et al.* (2015) Aquaria: simplifying discovery and insight from protein structures. *Nat. Methods*, **12**, 98–99.

Rose, A.S., and Hildebrand, P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res*., **43**, W576–W579.

Smith, T.F., and Waterman, M.S. (1981) Identification of common molecular subsequences. *J. Mol. Biol*., **147**, 195–197.

Stank, A. *et al*. (2016) ProSAT+: visualizing sequence annotations on 3D structure. *Protein Eng. Des. Sel*., **29**, 281–284.

Tabas-Madrid, D. *et al*. (2016) 3DBIONOTES: a unified, enriched and interactive view of macromolecular information. *J. Struct. Biol*., **194**, 231–234.

Tovchigrechko, A., and Vakser, I.A. (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res*., **34**, W310–W314.

Watkins, X. *et al*. (2017) ProtVista: visualization of protein sequence annotations. *Bioinformatics*, **33**, 2040–2041.