Bioinformatics, 31(15), 2015, 2580–2582
doi: 10.1093/bioinformatics/btv162
Advance Access Publication Date: 19 March 2015
Applications Note

OXFORD

Structural bioinformatics

Solubis: optimize your protein

Greet De Baets[1,2,†], Joost Van Durme[1,2,†], Rob van der Kant[1], Joost Schymkowitz[1,*] and Frederic Rousseau[1,*]

[1]VIB Switch Laboratory, Department of Cellular and Molecular Medicine, Katholieke Universiteit Leuven, 3000 Leuven, Belgium and [2]Vrije Universiteit Brussel, 1050 Brussels, Belgium

*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 December 16, 2014; revised on February 26, 2015; accepted on March 16, 2015

Abstract

Motivation: Protein aggregation is associated with a number of protein misfolding diseases and is a major concern for therapeutic proteins. Aggregation is caused by the presence of aggregation-prone regions (APRs) in the amino acid sequence of the protein. The lower the aggregation propensity of APRs and the better they are protected by native interactions within the folded structure of the protein, the more aggregation is prevented. Therefore, both the local thermodynamic stability of APRs in the native structure and their intrinsic aggregation propensity are a key parameter that needs to be optimized to prevent protein aggregation.
Results: The Solubis method presented here automates the process of carefully selecting point mutations that minimize the intrinsic aggregation propensity while improving local protein stability.
Availability and implementation: All information about the Solubis plugin is available at http://solubisyasara.switchlab.org/.
Contact: joost.schymkowitz@switch.vib-kuleuven.be or Frederic.Rousseau@switch.vib-kuleuven.be
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Protein aggregation remains a major bottleneck in the production of many therapeutic proteins. To reduce this aggregation problem, several approaches have been developed such as (i) generation of a fusion protein containing a solubilizing tag (Park *et al.*, 2008); (ii) careful formulation of the buffer (Wang, 1999) or (iii) increasing the colloidal stability by increasing the net charge (Kramer *et al.*, 2012). Nowadays, it is, however, common knowledge that aggregation is mediated by the presence of aggregation-prone regions (APRs) resulting in several computational methods that can predict the aggregation tendency and therefore, the effect of a mutation on protein aggregation (Supplementary Table S1). As none of the available methods currently takes into account protein stability, there is a need of integrating these with protein stability predictors (Supplementary Table S1). The Solubis method is such an integration approach that selects mutations that reduce the aggregation tendency by combining both TANGO (Fernandez-Escamilla *et al.*, 2004) and the empirical forcefield FoldX (Schymkowitz *et al.*, 2005). This combination results in a tool to guide the design of aggregation-resistant protein sequences, by identifying mutations that reduce TANGO while respecting the thermodynamic stability. We experimentally validated the approach by improving the solubility and abundance of both protective antigen and alpha-galactosidase (unpublished results). In addition, we analyzed the performance of our method on published variants that increase solubility (Supplementary Table S2). Eighteen of the 24 variants (75%) lower the aggregation tendency (ΔTANGO < 0), where the other six mutations do not increase he aggregation tendency. In contrast, many of them do not pass our conservative criteria of respecting structural stability ($\Delta\Delta G < 1$). However, for a large fraction of mutations with a $\Delta\Delta G$ higher than 1, the calculation were based on homology models, which is known to negatively affect the performance of FoldX. Overall, 6 out of the 24 analyzed variants that increase solubility can be identified following the Solubis method presented here but

this number increases to six out of seven if homology models are not included. Therefore, the use of homology models is not recommended to obtain an accurate prediction of the effect on protein stability.

For user friendliness, we here present an implementation of Solubis in the YASARA molecular viewer environment (Krieger *et al.*, 2002).

2 Implementation and required software

The Solubis plugin combines two established methods, FoldX and TANGO into YASARA. YASARA (View) can be downloaded freely at http://www.yasara.org. FoldX and TANGO are free for non-commercial purposes and can be downloaded from, respectively, http://foldx.crg.es and http://tango.crg.es. This plugin is written in python, which needs to be installed on MS Windows (http://www.python.org), and is platform independent. Installing the plugin adds a new dropdown 'Solubis' menu in the main 'Analyze' menu (Fig. 1).

3 The solubis menu items

3.1 Complete analysis of object

Complete analysis of object runs the TANGO algorithm and returns the short APRs present in your protein. To have an idea whether your protein contains such a stretch, you can analyze an object of choice. The APRs are colored, if desired according to their aggregating strength, in the same object or in a newly created object. The output of the algorithm is printed to the YASARA console and to a file (see 'Save last calculation').

3.2 Complete analysis of molecule

This item has the same function as 'Complete analysis of Object' but allows analyzing only certain molecules.

3.3 Mutate residue

Mutate residue mutates one residue to one or more new residues and returns the effect on aggregation tendency. The necessary steps are (i) residue selection to select the residue to mutate; (ii) selection menu to choose the desired replacement residue(s); (iii) set options (temperature, ionic strength and pH) to run the aggregation predictor, default shown; (iv) settings to define an aggregating stretch, default shown and (v) menu to indicate whether your structure contains a gap. This will return a new object for each selected mutation and outputs whether a mutation affects (both increase and decrease) the aggregation tendency. If the effect on protein stability is also required, we advise to use the FoldX plugin (Van Durme *et al.*, 2011) or use the menus 'Solubis run on complete molecule' or 'Solubis run in marked region'.

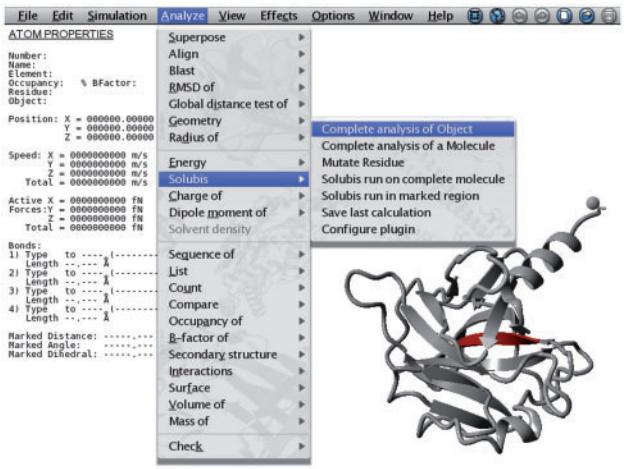

Fig. 1. Solubis interface in YASARA. This menu is part of the Analyze main menu and offers several implemented Solubis commands

3.4 Solubis run on complete molecule

Solubis run on complete molecule search for a selected number of mutants that minimize the intrinsic aggregation tendency and, if selected, do not destabilize the protein. As minimizing the aggregation tendency is done by introducing so called gatekeepers (P, R, K, D and E) (Rousseau *et al.*, 2006) into APRs, it is advised to filter out the destabilizing mutations with FoldX. The required steps are (i) select the molecule to mutate; (ii) set options to run the aggregation predictor, default shown; (iii) set thresholds to define an aggregating stretch, default shown (iv) menu to enable the FoldX analysis to calculate the effect on protein stability. As FoldX needs to start from a repaired structure, you need to select 'FoldX RepairPDB' if not the case; (v) options for FoldX specific settings: number of runs, temperature, pH, Ionic strength and Vander Waals penalty, default shown; (vi) menu to select the threshold for aggregation tendency and protein stability, default shown; (vii) box to select how many mutations you want to retrieve and (viii) menu to indicate whether your structure contains a gap.

If a mutation minimizes the aggregation tendency with a certain threshold and the $\Delta\Delta G$ is below the selected value (setting point 6), this mutation is considered as a Solubis mutant. Stringent conditions require a more negative $\Delta\Delta G$ (more stabilizing) and more positive aggregation threshold.

3.5 Solubis run in marked region

Solubis run in marked region performs the same calculation as above but on a selected region of the protein. When analyzing a large protein, these calculations can take a considerable amount of time, as multiple APRs are present. To limit the calculation time, this menu allows you to first select a certain region for further investigation. After this step, the same steps are followed as in 'Solubis run on complete molecule'. The marked region has to contain an APR, so it is recommended to first run 'Complete analysis of Object/Molecule' if unknown.

3.6 Save last calculation

The user can save all files or just the SUMMARY file. This file contains, dependent on the performed action, the APRs present in a protein, the effect of a mutation on the aggregation tendency or the Solubis mutations. If the user chooses to save all files, this includes the raw output and the used Options file.

4 Summary

Solubis is a protein redesign method to abrogate aggregation through disrupting aggregation-nucleating sequences with mutations that maximally reduce their intrinsic aggregation propensity while preserving thermodynamic stability of the functional protein. The graphical interface makes it very straightforward to identify mutants that lead to marked reduction in protein aggregation upon overexpression and will attract non-experienced computer users to perform these complex calculations. However, we want to emphasize that our method does not take into account available information about the active site or post-translational modifications but focus on the

structural integrity. Therefore; users should always control whether the selected mutations do not affect important functional sites.

Funding

The Switch Laboratory was supported by grants from VIB, University of Leuven, the Funds for Scientific Research Flanders (FWO), the Flanders Institute for Science and Technology (IWT) and the Federal Office for Scientific Affairs of Belgium (Belspo, IAP network P7/16). R.v.d.K. was supported by Boehringer Ingelheim Pharma GmbH & Co. KG.

Conflict of Interest: none declared.

References

Fernandez-Escamilla,A.M. *et al*. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.*, **22**, 1302–1306.

Kramer,R.M. *et al*. (2012) Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophys. J.*, **102**, 1907–1915.

Krieger,E. *et al*. (2002) Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins*, **47**, 393–402.

Park,J.S. *et al*. (2008) Solubility enhancement of aggregation-prone heterologous proteins by fusion expression using stress-responsive *Escherichia coli* protein, RpoS. *BMC Biotechnol.*, **8**, 15.

Rousseau,F., *et al*. (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. *J. Mol. Biol.*, **355**, 1037–1047.

Schymkowitz,J. *et al*. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**(Web Server issue), 382–W388.

Van Durme,J. *et al*. (2011) A graphical interface for the FoldX forcefield. *Bioinformatics*, **27**, 1711–1712.

Wang,W. (1999) Instability, stabilization, and formulation of liquid protein pharmaceuticals. *Int. J. Pharm.*, **185**, 129–188.