

Structural bioinformatics

TADOSS: computational estimation of tandem domain swap stability

Aleix Lafita ^{1,*}, Pengfei Tian², Robert B. Best ² and Alex Bateman ^{1,*}

¹European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK and ²Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

*To whom correspondence should be addressed.

Associate Editor: Alfonso Valencia

Received on April 18, 2018; revised on October 13, 2018; editorial decision on November 25, 2018; accepted on November 28, 2018

Abstract

Summary: Proteins with highly similar tandem domains have shown an increased propensity for misfolding and aggregation. Several molecular explanations have been put forward, such as swapping of adjacent domains, but there is a lack of computational tools to systematically analyze them. We present the TAndem DOrain Swap Stability predictor (TADOSS), a method to computationally estimate the stability of tandem domain-swapped conformations from the structures of single domains, based on previous coarse-grained simulation studies. The tool is able to discriminate domains susceptible to domain swapping and to identify structural regions with high propensity to form hinge loops. TADOSS is a scalable method and suitable for large scale analyses.

Availability and implementation: Source code and documentation are freely available under an MIT license on GitHub at <https://github.com/lafita/tadoss>.

Contact: aleixlafita@ebi.ac.uk or agb@ebi.ac.uk

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Protein misfolding and aggregation is a major problem for cells and organisms, and the cause of severe human diseases like Alzheimer's. Recent studies have shown an increased propensity of aggregation in proteins containing identical domains in tandem (Borgia *et al.*, 2015; Wright *et al.*, 2005). One of the misfolded conformations identified in these experiments are domain swaps, i.e. part of the structure of a domain folds into its adjacent domain. Domain swapping has been associated with protein aggregation (Rousseau *et al.*, 2012), so its computational prediction is of widespread biomedical and biotechnological interest.

A recent study by Tian and Best (2016) demonstrated using coarse-grained simulations of tandem pairs of identical domains that the difference in stability between the native and the domain-swapped conformations correlates with the swapping propensity. They also described an alchemical approach (i.e. simplified model) to estimate the free energy difference of the two conformations that can be generally used to predict domain swapping. Here we present an improved and fully automated version of the method, named TAndem DOrain Swap

Stability predictor (TADOSS), which can be used to systematically find domains susceptible to domain swapping and identify the regions of the structure with the highest propensity to form hinge loops.

2 Description

As described originally by Tian and Best (2016), the total free energy difference between the native and swapped conformations can be split into the energy of joining the N- and C- termini of the domain (ΔG_J) and the energy of cutting the domain (ΔG_C), i.e. forming a hinge loop between the swapped domains. TADOSS systematically evaluates all possible cut positions in the domain and calculates the free energy contribution of forming a hinge loop of at least three residues. The profile of ΔG_C is valuable to identify the regions of the domain that are more susceptible to form hinge loops in the tandem domain swaps, as shown in Figure 1.

The length of the linker between adjacent domains also plays an important role in the stability of tandem domain swaps. Longer linkers allow the connection of the N- and C-termini of the domain, thereby increasing the ΔG_J . To account for this effect, we have introduced an

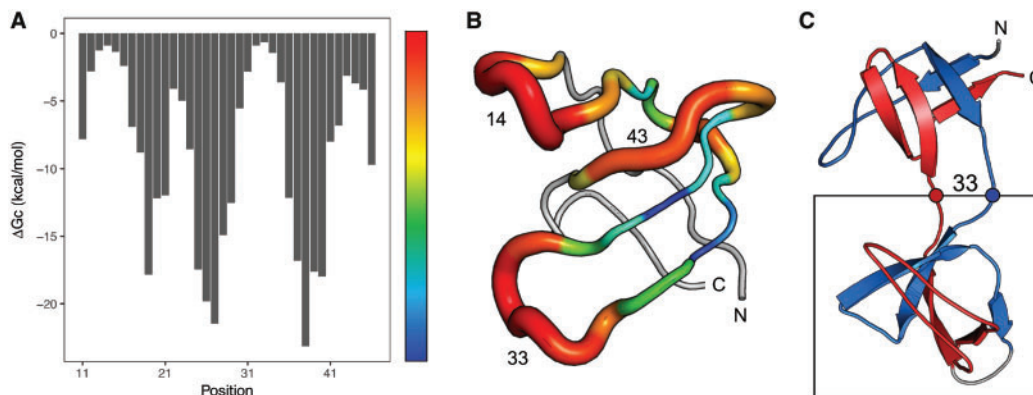


Fig. 1. Alchemical free energy of forming a hinge loop (ΔG_c) centered at every residue position in an SH3 domain (ECOD: e1shgA1). Energies can be visualized as a profile along the domain sequence (A) or mapped in 3D to the domain structure (B). Higher (more positive) free energy differences (red) indicate regions of the structure more likely to form hinge loops. For the SH3 domain, the alchemical free energy of joining the termini (ΔG_j) is -6.1 kcal/mol. The most stable domain swap conformation is predicted to be forming a hinge loop centered at Position 33, with a total $\Delta\Delta G$ of -6.8 kcal/mol. The corresponding tandem domain-swapped structure extracted from the simulations by Tian and Best (2016) is also shown (C)

optional parameter in TADOSS to reduce the effective distance between the termini of the domain proportional to the length of the linker.

Finally, the total free energy difference $\Delta\Delta G$ of a tandem domain swap is obtained by summing up the free energy differences for cutting (ΔG_c) and joining (ΔG_j) the domain. The most susceptible domain swaps are those with the maximum $\Delta\Delta G$ (most positive). More information about the energetic model and its parameters can be found in the [supplementary materials](#).

3 Results

The alchemical free energy difference from TADOSS correlates well with the free energies obtained in the simulations by Tian and Best (2016), although the scale differs by a factor of two approximately (Supplementary Fig. S1). We also find an agreement in the effect of the linker length between the simulated and alchemical free energy differences (Supplementary Fig. S2).

Using the free energy profile from TADOSS, it is possible to reproduce experimental ΔG measurements and identify with good accuracy the folding units of a DHFR domain characterized by Iwakura et al. (2000) (Supplementary Fig. S3). Furthermore, hinge loops of experimentally determined domain swap dimers presented by Ding et al. (2006) correspond to maximums of the ΔG_c profile, as expected (Supplementary Fig. S4).

We also provide a dataset of alchemical $\Delta\Delta G$ estimations for T-group (topology) representatives in the ECOD database (Cheng et al., 2014). We find a significant proportion of domains susceptible to tandem domain swapping (Supplementary Fig. S5).

4 Implementation

TADOSS is written in Python and bundled in a Bash script with a simple interface to the user. The program takes the structure of a protein domain as a PDB file and generates output files with the alchemical free energy differences. The method requires BioPython (Cock et al., 2009) to parse and manipulate the domain structures and either GROMACS (Abraham et al., 2015) or Reduce (Word et al., 1999) to add the hydrogens.

The structure of the input domain is represented using a coarse-grained structure-based (Gō-like) model, as described by

Karanicolas and Brooks (2002). Native interactions in the structure are attractive and the relative contact energies are set according to the Miyazawa-Jernigan matrix (Miyazawa and Jernigan, 1996).

The running time for an example domain of 159 residues on a MacBook Pro 2.9 GHz Intel Core i5 with 16 GB RAM is about 4 s. The method scales quadratically with the number of residues in the input domain structure (Supplementary Fig. S6).

Funding

This work was supported by the intramural research program of the National Institute of Diabetes and Digestive and Kidney Diseases (grant number ZIA DK075104-06) of the National Institutes of Health to P.T. and R.B.B. This work was funded by the European Molecular Biology Laboratory.

Conflict of Interest: none declared.

References

- Abraham, M.J. et al. (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, 19–25.
- Borgia, A. et al. (2015) Transient misfolding dominates multidomain protein folding. *Nat. Commun.*, 6, 8861.
- Cheng, H. et al. (2014) ECOD: an evolutionary classification of protein domains. *PLoS Comput. Biol.*, 10, e1003926.
- Cock, P.J. et al. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25, 1422–1423.
- Ding, F. et al. (2006) Topological determinants of protein domain swapping. *Structure*, 14, 5–14.
- Iwakura, M. et al. (2000) Systematic circular permutation of an entire protein reveals essential folding elements. *Nat. Struct. Biol.*, 7, 580–585.
- Karanicolas, J. and Brooks, C.L. (2002) The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.*, 11, 2351–2361.
- Miyazawa, S. and Jernigan, R.L. (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.*, 256, 623–644.
- Rousseau, F. et al. (2012) Implications of 3D domain swapping for protein folding, misfolding and function. *Adv. Exp. Med. Biol.*, 747, 137–152.
- Tian, P. and Best, R.B. (2016) Structural determinants of misfolding in multidomain proteins. *PLoS Comput. Biol.*, 12, e1004933.
- Word, J.M. et al. (1999) Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.*, 285, 1735–1747.
- Wright, C.F. et al. (2005) The importance of sequence diversity in the aggregation and evolution of proteins. *Nature*, 438, 878–881.