bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400432; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FrustratometeR: an R-package to compute Local frustration in protein structures, point mutants and MD simulations

Atilio O. Rausch[1*], Maria I. Freiberger[2*], Cesar O. Leonetti[2], Diego M. Luna[1], Leandro G. Radusky[3], Peter G. Wolynes[4], Diego U. Ferreiro[2], and R. Gonzalo Parra[5] ✉

[1]Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Argentina., [2] Laboratorio de Fisiología de Proteínas, Departamento de Química Biológica - IQUIBICEN/CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina., [3] Center for Genomic Regulation, Barcelona Institute for Science and Technology, Barcelona, Spain., [4] Center for Theoretical Biological Physics and Department of Chemistry, Rice University, Houston, TX, 77005, USA., [5] European Molecular Biology, Laboratory, Heidelberg, Germany., [*] Jointly 1st authors

Once folded natural protein molecules have few energetic conflicts within their polypeptide chains. Many protein structures do however contain regions where energetic conflicts remain after folding, i.e. they have highly frustrated regions. These regions, kept in place over evolutionary and physiological timescales, are related to several functional aspects of natural proteins such as protein-protein interactions, small ligand recognition, catalytic sites and allostery. Here we present FrustratometeR, an R package that easily computes local energetic frustration on a personal computer or a cluster. This package facilitates large scale analysis of local frustration, point mutants and MD trajectories, allowing straightforward integration of local frustration analysis in to pipelines for protein structural analysis.
Availability and implementation:
https://github.com/proteinphysiologylab/frustratometeR

R package | protein structures | local frustration | molecular dynamics
Correspondence: ***gonzalo.parra@embl.de***

Introduction

Proteins are evolved biological molecules that adopt a defined set of structures constituting their 'native' state. Built as linear polymers, proteins find their native state easily since evolution has minimized the internal energetic conflicts within their polypeptide chain, following the "principle of minimal frustration", (1)). Proteins are not only biologically optimized to fold or to be stable but also to 'function' (2, 3) and therefore it's not surprising to find that about 10-15% of the internal interactions in proteins are in energetic conflict within their local structure (4). These conflicts, kept in place over evolutionary and physiological time scales, allow proteins to explore different conformations within their native ensemble and thus enable the emergence of 'function'. Over the last years, the concept of local frustration has given insights into a diver set of functional phenomena: protein-protein interactions, ligand recognition, allosteric sites (5), enzymatic active sites and co-factors binding (6), evolutionary patterns in protein families (7) and disease associated mutations (8). Frustration and its role in functional dynamics has been recently reviewed in (9). Up to now, facile location and quantification of energetic frustration has been made possible via the Frustratometer web server (10, 11). Unfortunately high-throughput analysis using the server is not feasible as the flexibility of the algorithm was reduced in order to maximise usability by non computational scientists. Here we present FrustratometeR, an R package that retains all the capabilities present in the web server but that also includes brand new modules to evaluate how frustration varies upon point mutations as well as to analyse how frustration varies as a function of time during molecular dynamics simulations. FrustratometeR facilitates the analysis both at small and high-throughput scales and thus allows one to integrate local frustration analysis into other protein structural bioinformatics pipelines.

Description: The FrustratometeR package

Full description of methods and code examples can be found in the supplementary data and github repository.

Calculate local frustration: A single function, *calculate_frustration()*, the PDB structure or PDBID and the desired frustration index (i.e. *configurational, mutational or singleresidue*) specification are enough to calculate the local frustration. Different visualization functions are available to explore results (Fig1 and supplementary material).
Frustration upon mutations: FrustratometeR provides the *mutate_res()* function that allows users to analyse how point mutations affect local frustration (Fig1B). It will calculate the frustration for all the 20 possible amino acid variants, including the one naturally found in the structure for a specific residue in a specific chain. Mutations can be generated in two different ways: 1) threading mode: changes the amino acid identity but does not apply any change to the backbone structure. 2) modeller mode: generates a homology model based on the PDB native structure which is then energetically minimised before calculating local frustration.
Frustration along molecular dynamics simulations: The *dynamic_frustration()* function can analyze frustration along a molecular dynamics trajectory (Fig1C). As an input, N frames must be extracted as individual PDB files and placed in a directory. The *detect_dynamic_clusters()* function can group protein residues based on their temporal frustration profiles and represent them in a graph to find dynamic modules (Fig1C). Individual residues temporal dynamics can be visualized as well (Fig1D). To illustrate frustratometeR us-

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400432; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

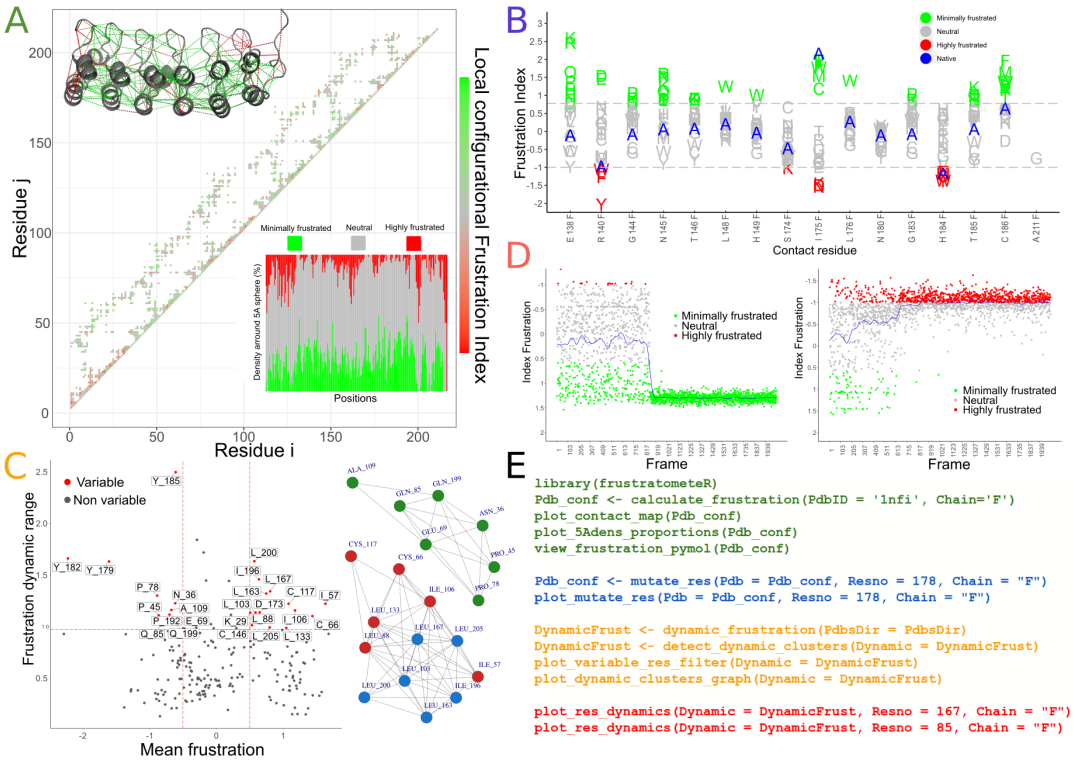

Fig. 1. Frustration in IκBα (PDB ID 1nfi,F). (A) Contact map, 5Adens plot and pymol representation. (B) Frustration changes when mutating a specific residue to all canonical amino acids alternatives. X axis: all residues that interact with the residue of interest for all mutant are displayed. Y axis: frustration values are shown and coloured based on their frustration classification. Native variant appears in blue. Each mutant is represented by its 1-letter amino acid code to identify to which variant it corresponds. (C) Dynamic frustration modules: Residues that vary their frustration across frames in molecular dynamic simulations are identified based on their frustration average and dynamic range values (left). Variable residues are connected to each other in a correlation network that is then clustered to find modules with similar dynamic behaviour (details in supplementary data). (D) Frustration values for a given residue across molecular dynamics frames. (E) Minimum code to generate the previous panels

ability, Fig1E shows the minimum code to reproduce the rest of Fig1 panels.

Application

As an example, we have applied FrustratometeR to IκBα, an inhibitor of NfκB (PDBID 1nfi,F). Fig1A shows a composite figure of typical Frustratometer web-server-like visualisations (11). FrustratometeR introduces a new functionality to evaluate the change in frustration for amino acid variants. Fig1B shows the change in the configurational frustration index for residue ALA178 when it is mutated to all the other amino acids in threading mode. FrustratometeR also includes a module to analyse molecular dynamics simulations and to identify residues that have similar dynamics. We extracted frames from an IκBα coarse-grained folding simulation (see supplementary data). First, m variable residues are selected based on their average and the dynamic range of frustration values across frames. A low dimensional representation of a matrix of m residues and n frames is obtained using Principal Components Analysis (PCA) and residues pairwise correlations are calculated to create a graph that is clustered to define dynamic modules Fig1C. The blue cluster contains residues that become minimally frustrated over time, represented by LEU167 (Fig1D) in contrast to the green cluster that groups the residues that increase their frustration values represented by GLN85 (Fig1D). FrustratometeR is designed to minimize the amount of code to perform the different analyses (Fig1E).

Discussion and conclusion:

We present a user-friendly R package to calculate energetic local frustration in protein structures. The package includes new features to assess the effect of mutations in local frustration as well as to analyse frustration along molecular dynamics trajectories. Its simple interface together with the newly implemented functionalities will facilitate frustration analysis at larger scales and can be used to include FrustratometeR as part of different pipelines for protein structural analysis.

Bibliography

1. JD Bryngelson and PG Wolynes. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. U. S. A.*, 84(21):7524–7528, November 1987.
2. DU Ferreiro et al. Frustration in biomolecules. *Quarterly reviews of biophysics*, 47(4):285–363, 2014.
3. R Espada et al. The design of repeat proteins: Stability conflicts with functionality, 2017.
4. DU Ferreiro et al. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A.*, 104(50):19819–19824, December 2007.
5. DU Ferreiro et al. On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. U. S. A.*, 108(9):3499–3503, March 2011.
6. MI Freiberger et al. Local frustration around enzyme active sites. *Proc. Natl. Acad. Sci. U. S. A.*, 116(10):4037–4043, March 2019.
7. RG Parra et al. Structural and energetic characterization of the ankyrin repeat protein family. *PLoS Comput. Biol.*, 11(12):e1004659, December 2015.
8. S Kumar et al. Localized structural frustration for evaluating the impact of sequence variants. *Nucleic Acids Res.*, 44(21):10062–10073, December 2016.
9. DU Ferreiro et al. Frustration, function and folding. *Curr. Opin. Struct. Biol.*, 48:68–73, February 2018.
10. M Jenik et al. Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Res.*, 40(Web Server issue):W348–51, July 2012.
11. RG Parra et al. Protein frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.*, 44(W1):W356–60, July 2016.