

Poster presentation

Open Access

Protein evolution driven by symmetric structural repeats

Anne-Laure Abraham*^{1,2}, Joël Pothier¹ and Eduardo PC Rocha^{1,2}

Address: ¹UPMC Univ Paris 06, Atelier de BioInformatique, F75005, Paris, France and ²Institut Pasteur, Microbial Evolutionary Genomics ; CNRS, URA2171, F75015 Paris, France

Email: Anne-Laure Abraham* - annela@abi.snv.jussieu.fr

* Corresponding author

from Fourth International Society for Computational Biology (ISCB) Student Council Symposium Toronto, Canada. 18 July 2008

Published: 30 October 2008

BMC Bioinformatics 2008, 9(Suppl 10):P3 doi:10.1186/1471-2105-9-S10-P3

This abstract is available from: <http://www.biomedcentral.com/1471-2105/9/S10/P3>

© 2008 Abraham et al; licensee BioMed Central Ltd

Background

Duplications play a major role in the evolution of genomes by creating and modifying molecular functions. Repeats are created at the DNA level, but when they are intragenic they affect protein sequence and structure. Moreover, as sequences evolve faster than 3D structures, some ancient repeats cannot be detected at the sequence level. Therefore, an understanding of the evolutionary and functional role of repeats requires tracking them at the three levels.

Materials and methods

For this task we have developed a program (SWELFE) to find repeats in DNA, amino acid sequences and 3D protein structures (available at <http://bioserv.rpbs.jussieu.fr/swelfe>) [1]. Repeats were searched independently at all three levels using dynamic programming (SIM algorithm) and we adapted scoring matrices for each level. 3D structures were encoded as strings of α angles (dihedral angle between four consecutive $C\alpha$) so that the same algorithm can be used at the three levels. We verified the validity of our approach by superimposing the 3D repeats. We made a statistical evaluation of the repeats we found.

As there is currently no connection between 3D structures and the corresponding DNA sequences, we built a databank where we associate a PDB structure with its best match among the genes in TREMBL. The databank has 85 845 entries corresponding to ~85% of the unique chains of the PDB databank after removing entries with very small peptides or many undefined residues. We could

then compare directly the results of repeat identification at the three levels.

Results

In our analyses we were surprised to find that many of the large structural repeats are symmetrical (2-fold symmetry): one copy of the repeat can be superimposed with the other copy by a simple rotation of 180° around an axis (some repeats have 3-fold symmetry or higher). An example of a symmetrical repeat is shown on Figure 1. Then structures containing such repeats resemble to homodimers combined into one unit.

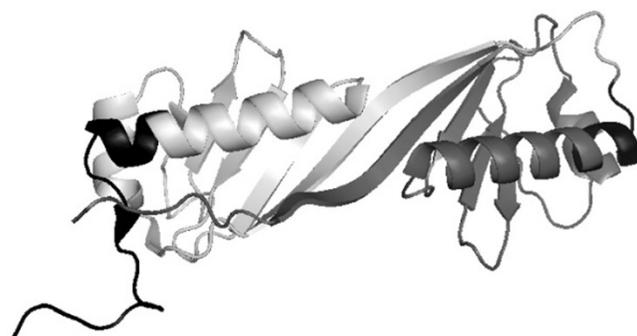


Figure 1
Example of repeat found in the 3D structure of Tata-box Binding Protein (TBP) of *Sulfolobus acidocaldarius* (IMP9). Repeats are shown in light grey and non-repeated regions are shown in black. The repeat is 83 amino acids long.

Acknowledgements

This work was supported by grants from Region Ile-de-France to ALA, ACI IMPBIO to EvolRep and ANR-06-CIS to project PROTEUS.

References

1. Abraham A-L, Rocha EPC, Pothier J: **Swelfe : a detector of internal repeats in sequences and structures.** *Bioinformatics* 2008. doi: 10.1093/bioinformatics/btn234

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

