

2 Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths.

Nepomnyachiy S, Ben-Tal N, Kolodny R

[show author affiliations](#)

Proc Natl Acad Sci U S A. 2017 Oct 31; 114(44):11703-11708

Save/Follow

Export

Get Article

FIND IT @ HARVARD

G+

RECOMMENDATIONS 1 | ABSTRACT | COMMENTS

[collapse all](#)

Recommendations:

Very Good

21 Nov 2017



Patrice Koehl

F1000 Structural Biology

University of California, Davis, Davis, CA, USA.

REFUTATION

DOI: 10.3410/f.732110432.793538904

There is a common belief in molecular biology, which states that domains are the basic units that define protein sequence evolution (although we note that the definition of domains is not universal). In this paper, the authors question this belief through a set of statistical analyses of the repeated use of 'short' fragments (at least 35 residues long) among a large collection of protein chains. They have shown that the repeated use of such fragments does not map with domain boundaries. They derive from these data a recursive model for protein evolution that goes beyond the concept of domains. This model is expected to prove useful for protein sequence design.

Disclosures

None declared

[Add a comment](#)

Abstract:

ABSTRACT

Proteins share similar segments with one another. Such "reused parts"-which have been successfully incorporated into other proteins-are likely to offer an evolutionary advantage over de novo evolved segments, as most of the latter will not even have the capacity to fold. To systematically explore the evolutionary traces of segment "reuse" across proteins, we developed an automated methodology that identifies reused segments from protein alignments. We search for "themes"-segments of at least 35 residues of similar sequence and structure-reused within representative sets of 15,016 domains [Evolutionary Classification of Protein Domains (ECOD) database] or 20,398 chains [Protein Data Bank (PDB)]. We observe that theme reuse is highly prevalent and that reuse is more extensive when the length threshold for identifying a theme is lower. Structural domains, the best characterized form of reuse in proteins, are just one of many complex and intertwined evolutionary traces. Others include long themes shared among a few proteins, which encompass and overlap with shorter themes that recur in numerous proteins. The observed complexity is consistent with evolution by duplication and divergence, and some of the themes might include descendants of ancestral segments. The observed recursive footprints, where the same amino acid can simultaneously participate in several intertwined themes, could be a useful concept for protein design. Data are available at <http://trachel-srv.cs.haifa.ac.il/rachel/ppi/themes/>.

Copyright © 2017 the Author(s). Published by PNAS.

DOI: [10.1073/pnas.1707642114](https://doi.org/10.1073/pnas.1707642114)

PMID: [29078314](https://pubmed.ncbi.nlm.nih.gov/29078314/)



Abstract courtesy of PubMed: A service of the National Library of Medicine and the National Institutes of Health.

Comments:

COMMENTS

[add a comment](#)

Librarian Resources	Article Recommendations	Work	Articles
Press Office	F1000Prime Reports	FAQ	Advisory Panel
F1000 Specialists	F1000Prime Faculty	Contact	Blog
F1000 Updates	Blog		Submit
Subscribe	Subscribe		Author Guidelines
Careers	About		Register
About/Contact	FAQs		About
	Contact		Contact

The F1000.com website uses cookies. By continuing to browse the site, you are agreeing to our use of cookies. [Find out more »](#)

© 2000-2017 Faculty of 1000 Ltd. ISSN 2051-9796 | Legal | Partner of HINARI • CrossRef • ORCID
F1000 is a registered trademark of Faculty of 1000 Limited