I believe that the right way to direct science is almost not to direct it at all. Attention must certainly be paid to what the public wants and what the political system can be persuaded to support, but the notion that bureaucrats—even those who were once scientists—know what our scientific priorities should be and can steer us in the appropriate direction strikes me as a recipe for disaster. Scientific priorities must, for the most part, be set by the free exchange of ideas in the scientific literature, at meetings and in review panels. They must be set from the bottom up, from the community of scientists, not by the people who control the purse strings.

GDCh
Essays
Angewandte Chemie
International Edition

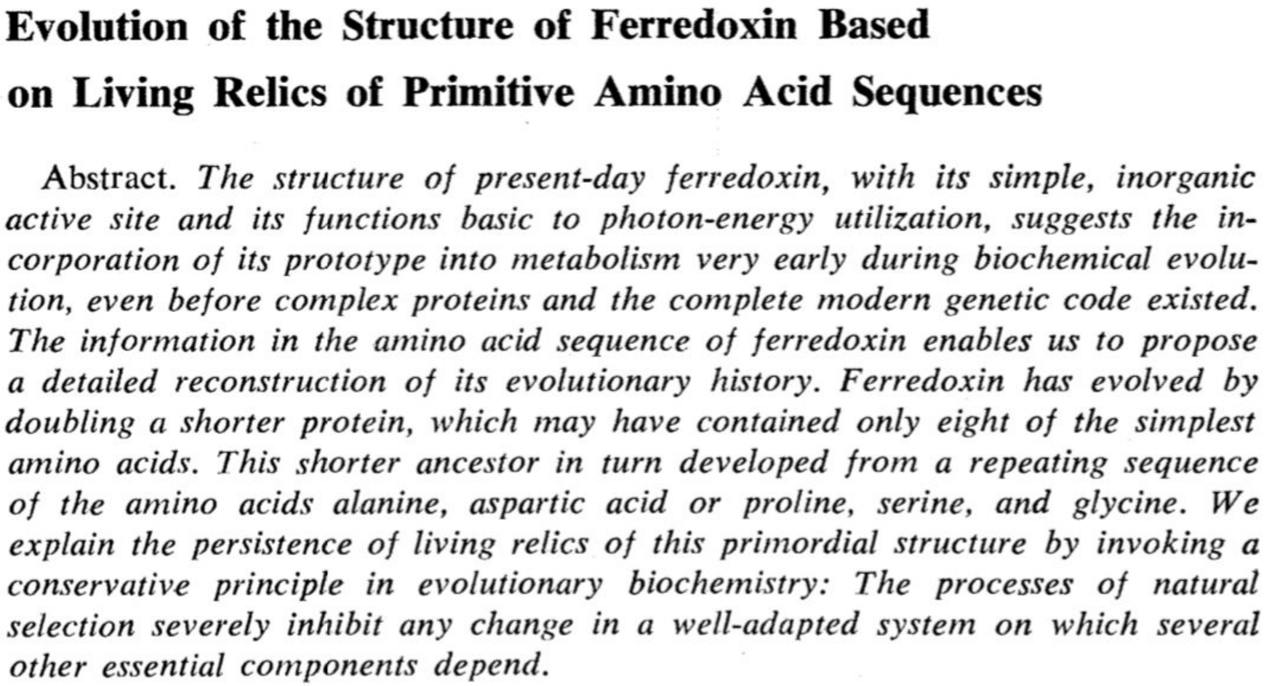
Origin of Proteins

International Edition: DOI: 10.1002/anie.201609977
German Edition: DOI: 10.1002/ange.201609977

Functional Proteins from Short Peptides: Dayhoff's Hypothesis Turns 50

*M. Luisa Romero Romero, Avigayel Rabin, and Dan S. Tawfik**

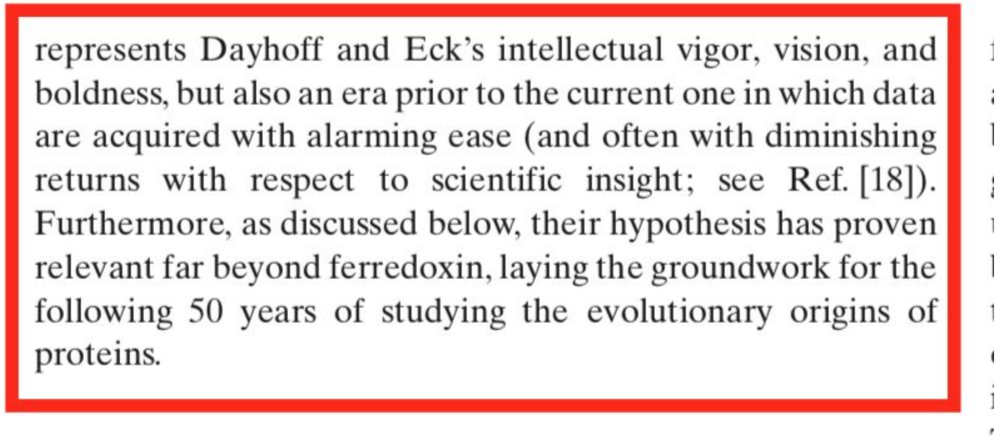
Margaret O. Dayhoff · peptides · proteins ·
protein evolution · origin of life



Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences

Abstract. *The structure of present-day ferredoxin, with its simple, inorganic active site and its functions basic to photon-energy utilization, suggests the incorporation of its prototype into metabolism very early during biochemical evolution, even before complex proteins and the complete modern genetic code existed. The information in the amino acid sequence of ferredoxin enables us to propose a detailed reconstruction of its evolutionary history. Ferredoxin has evolved by doubling a shorter protein, which may have contained only eight of the simplest amino acids. This shorter ancestor in turn developed from a repeating sequence of the amino acids alanine, aspartic acid or proline, serine, and glycine. We explain the persistence of living relics of this primordial structure by invoking a conservative principle in evolutionary biochemistry: The processes of natural selection severely inhibit any change in a well-adapted system on which several other essential components depend.*

1966 ----> 2016



represents Dayhoff and Eck's intellectual vigor, vision, and boldness, but also an era prior to the current one in which data are acquired with alarming ease (and often with diminishing returns with respect to scientific insight; see Ref. [18]). Furthermore, as discussed below, their hypothesis has proven relevant far beyond ferredoxin, laying the groundwork for the following 50 years of studying the evolutionary origins of proteins.

A galaxy of folds

Vikram Alva,[1] Michael Remmert,[1,2] Andreas Biegert,[1,2] Andrei N. Lupas,[1]* and Johannes Söding[1,2]*

[1]Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Tübingen 72076, Germany
[2]Gene Center and Center for Integrated Protein Science, Ludwig-Maximilians-University Munich, Munich 81377, Germany

Received 28 September 2009; Accepted 4 November 2009
DOI: 10.1002/pro.297

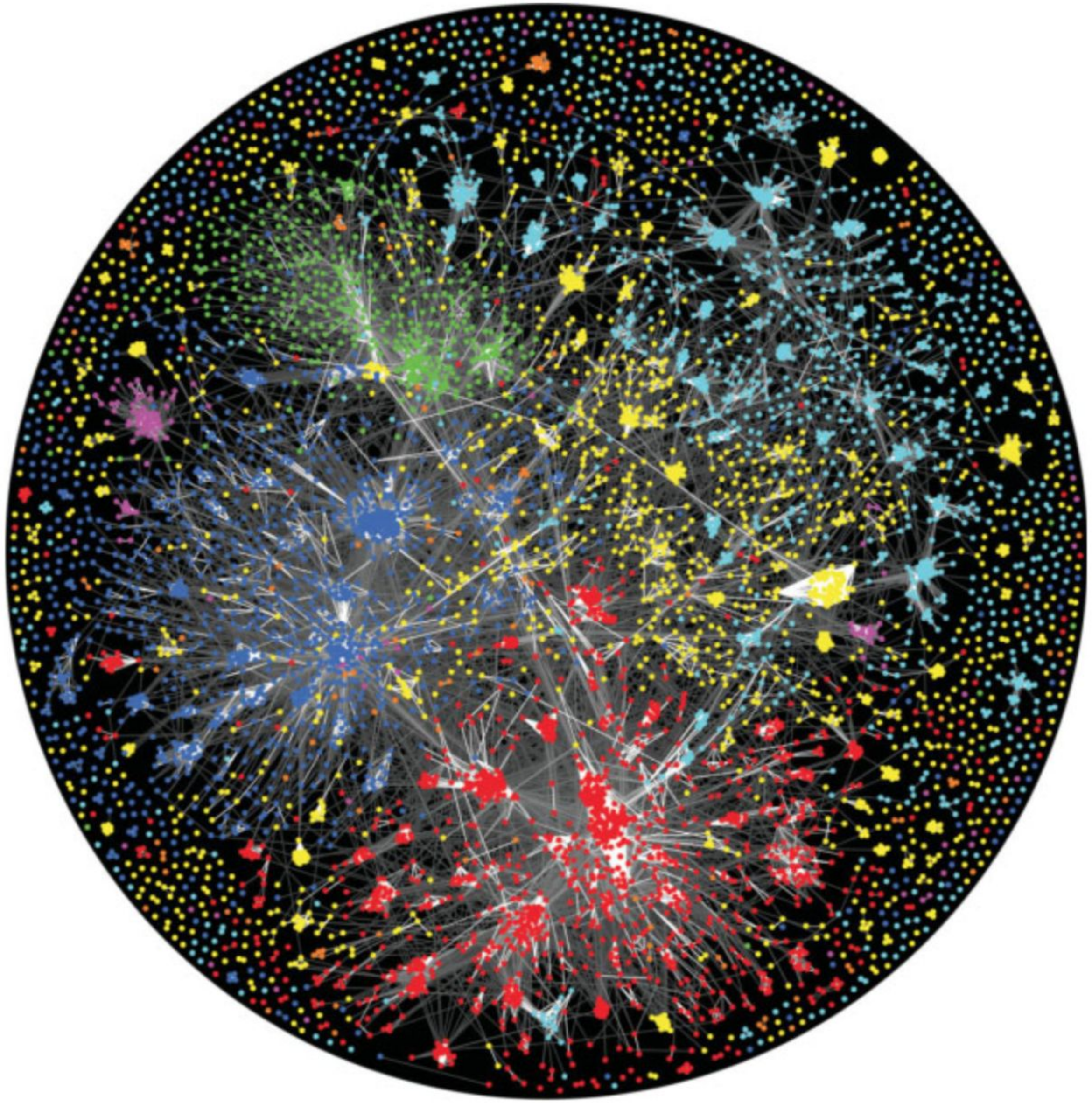
all alpha

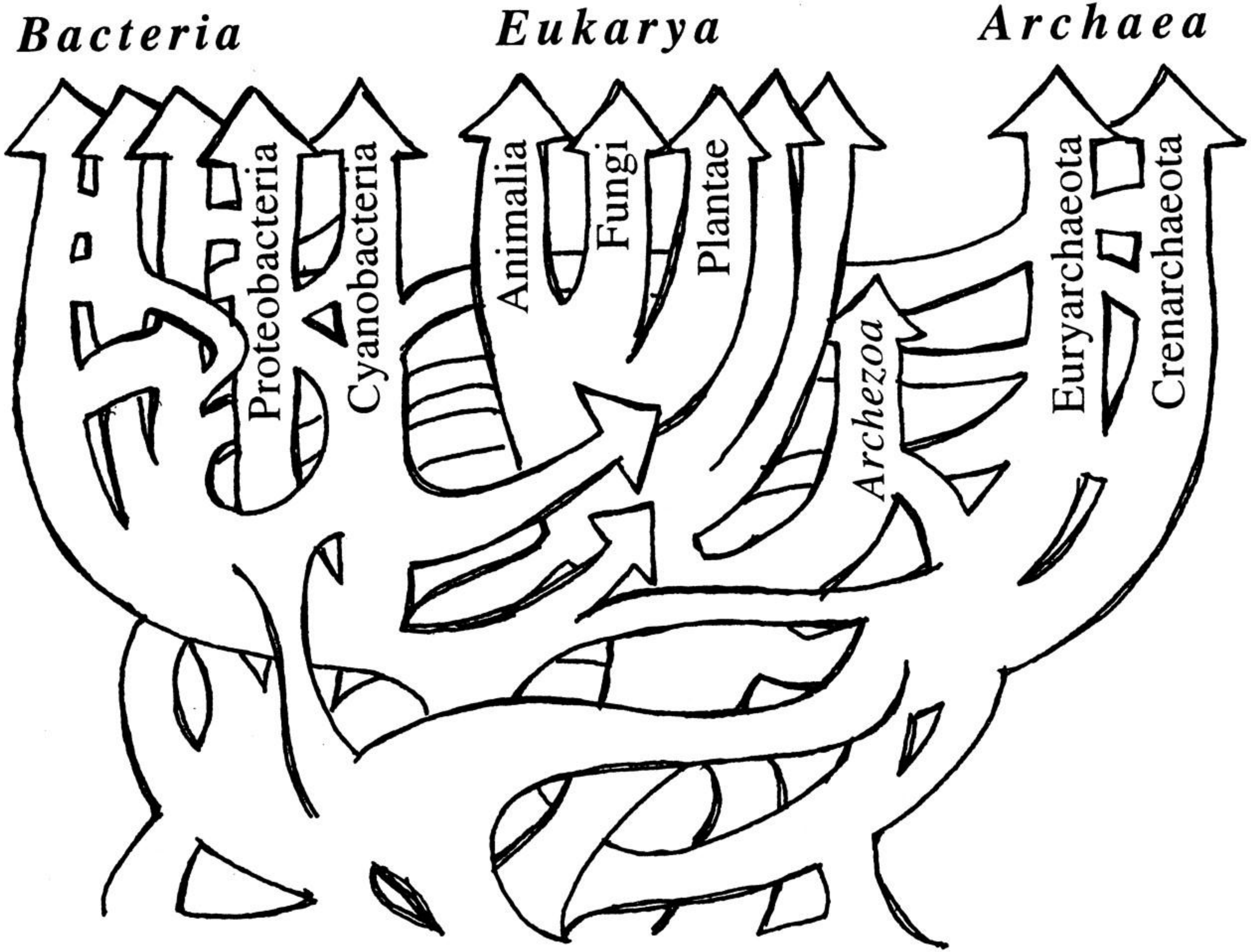
all beta

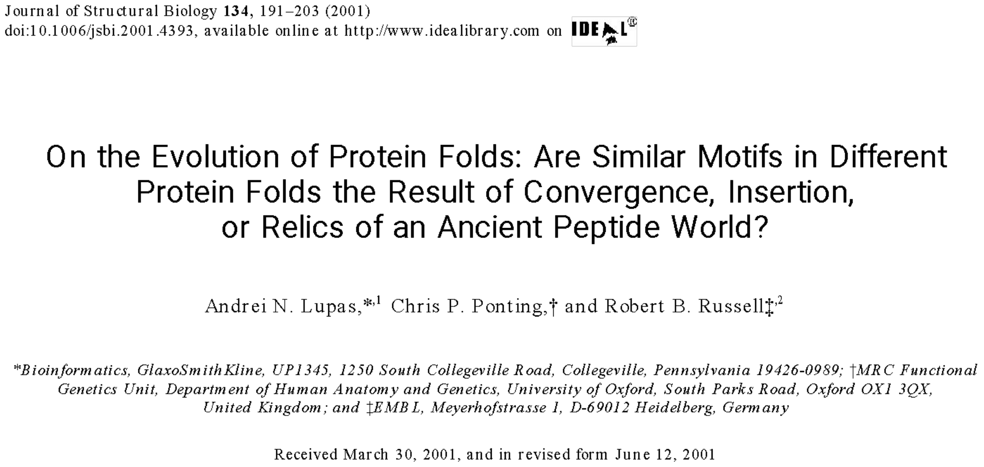
alpha/beta

alpha + beta

small proteins




Bacteria
Eukarya
Archaea
Proteobacteria
Cyanobacteria
Animalia
Fungi
Plantae
Archezoa
Euryarchaeota
Crenarchaeota



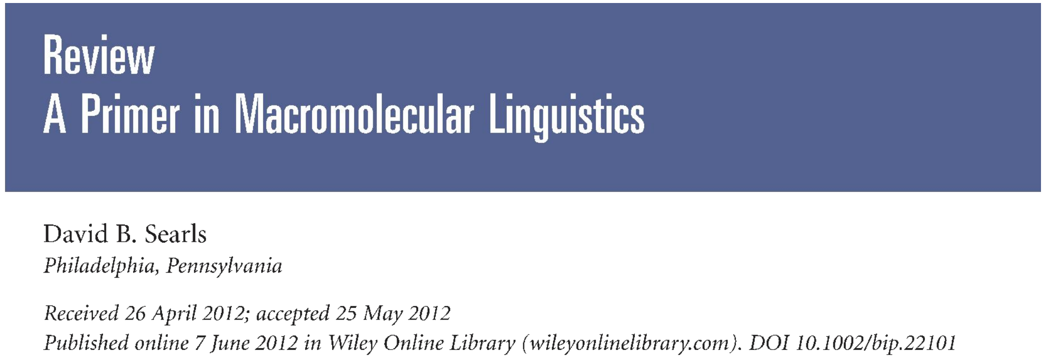
Journal of Structural Biology **134**, 191–203 (2001)
doi:10.1006/jsbi.2001.4393, available online at http://www.idealibrary.com on

On the Evolution of Protein Folds: Are Similar Motifs in Different Protein Folds the Result of Convergence, Insertion, or Relics of an Ancient Peptide World?

Andrei N. Lupas,*[1] Chris P. Ponting,† and Robert B. Russell‡[2]

*Bioinformatics, GlaxoSmithKline, UP1345, 1250 South Collegeville Road, Collegeville, Pennsylvania 19426-0989; †MRC Functional Genetics Unit, Department of Human Anatomy and Genetics, University of Oxford, South Parks Road, Oxford OX1 3QX, United Kingdom; and ‡EMBL, Meyerhofstrasse 1, D-69012 Heidelberg, Germany

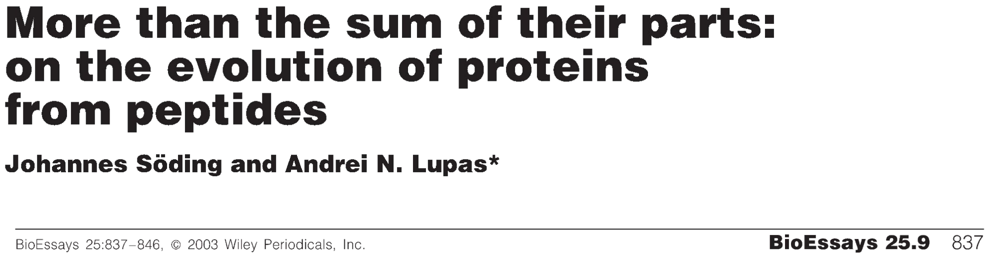
Received March 30, 2001, and in revised form June 12, 2001



Review
A Primer in Macromolecular Linguistics

David B. Searls
Philadelphia, Pennsylvania

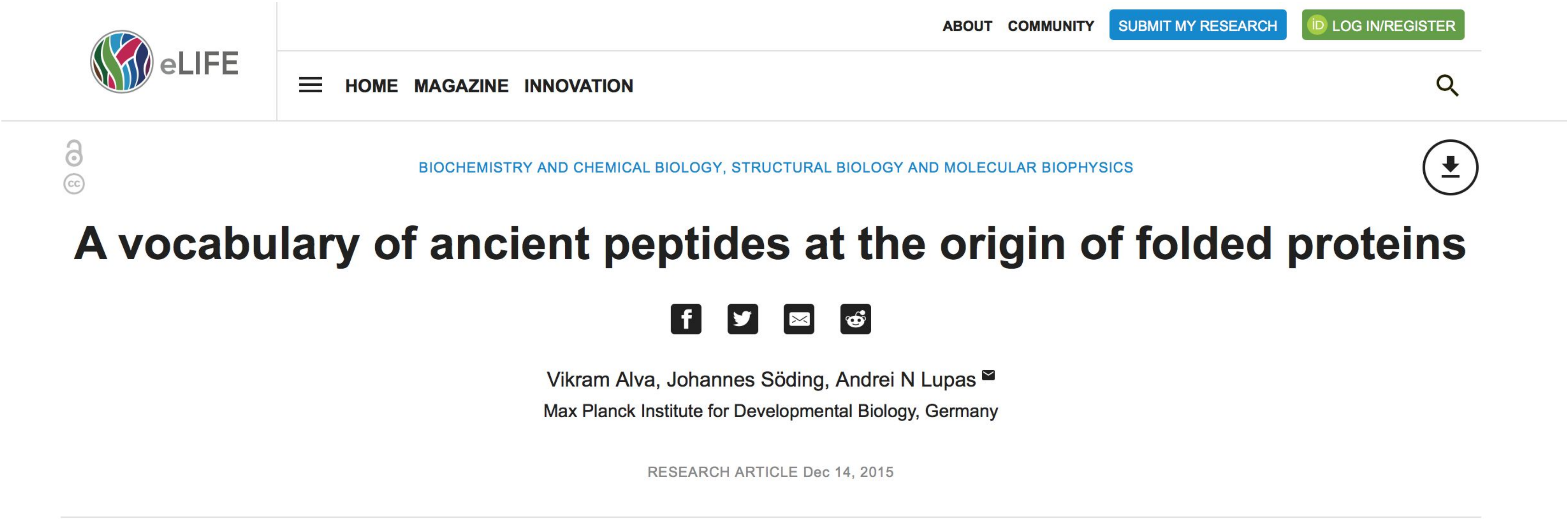
Received 26 April 2012; accepted 25 May 2012
Published online 7 June 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.22101



More than the sum of their parts: on the evolution of proteins from peptides

Johannes Söding and Andrei N. Lupas*

BioEssays 25:837–846, © 2003 Wiley Periodicals, Inc. **BioEssays 25.9** 837



BIOCHEMISTRY AND CHEMICAL BIOLOGY, STRUCTURAL BIOLOGY AND MOLECULAR BIOPHYSICS

A vocabulary of ancient peptides at the origin of folded proteins

Vikram Alva, Johannes Söding, Andrei N Lupas ✉

Max Planck Institute for Developmental Biology, Germany

RESEARCH ARTICLE Dec 14, 2015

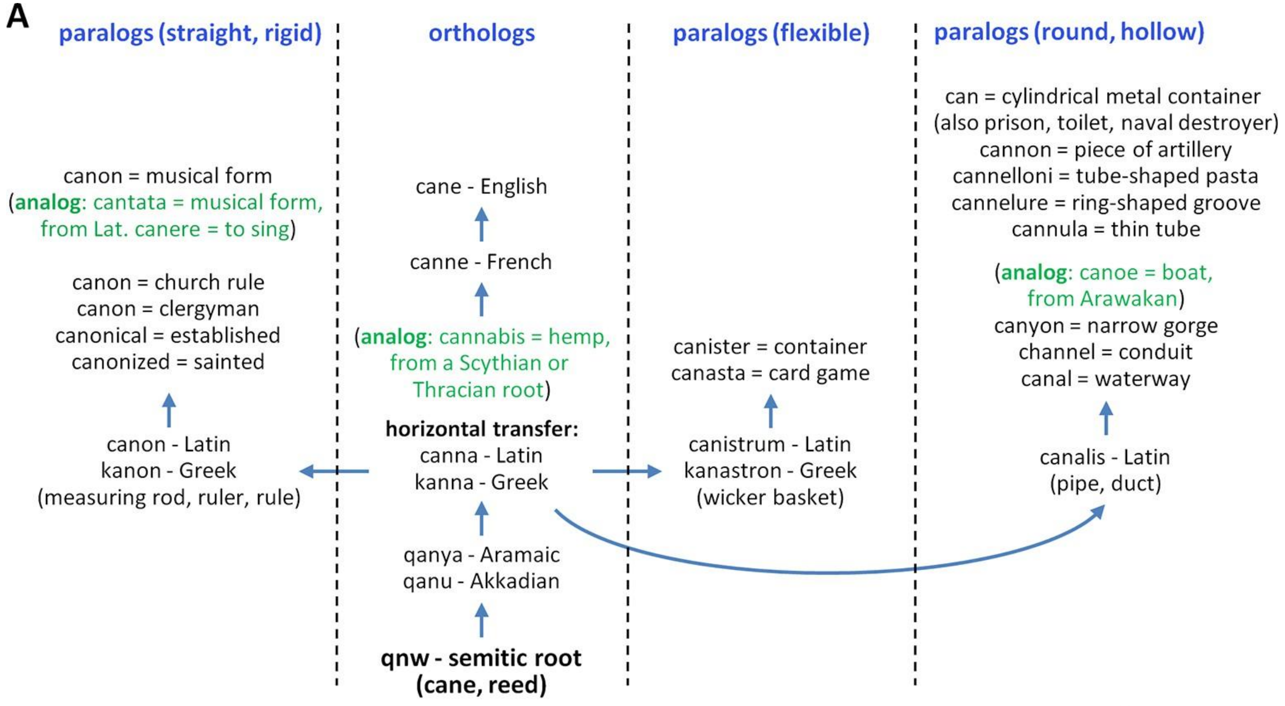
FÚTBOL

La Comisión Social de Peñarol organizó una 5 k con reclusos del ex Comcar

El colectivo lleva adelante diversas iniciativas a nivel social, como las ferias de emprendedores y actividades con niños y adultos mayores

Fiorella Rodríguez

A vocabulary of ancient peptides at the origin of folded proteins



A

paralogs (straight, rigid)

canon = musical form
(**analog**: cantata = musical form, from Lat. canere = to sing)

canon = church rule
canon = clergyman
canonical = established
canonized = sainted

canon - Latin
kanon - Greek
(measuring rod, ruler, rule)

orthologs

cane - English

canne - French

(**analog**: cannabis = hemp, from a Scythian or Thracian root)

horizontal transfer:
canna - Latin
kanna - Greek

qanya - Aramaic
qanu - Akkadian

qnw - semitic root
(cane, reed)

paralogs (flexible)

canister = container
canasta = card game

canistrum - Latin
kanastron - Greek
(wicker basket)

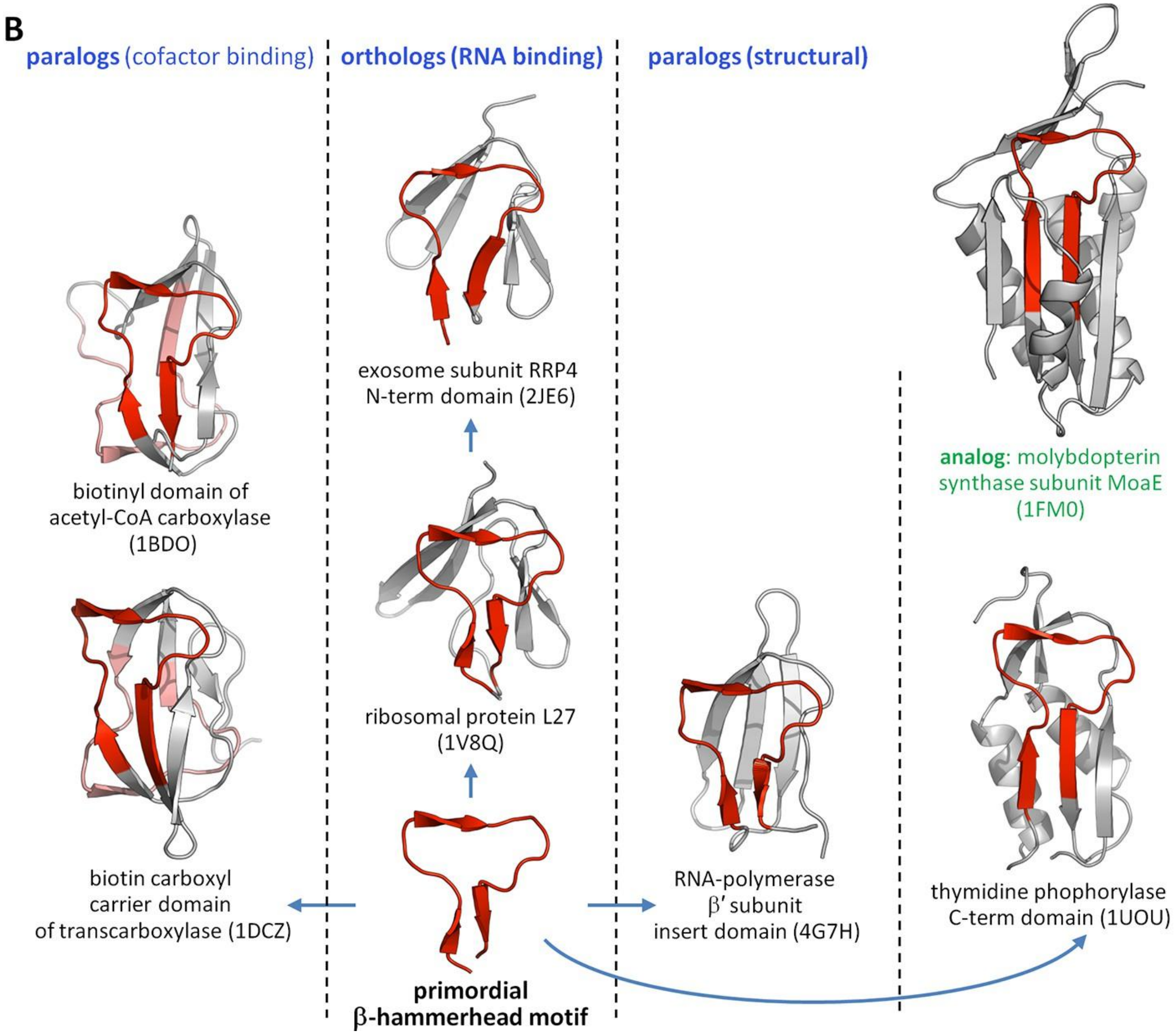
paralogs (round, hollow)

can = cylindrical metal container
(also prison, toilet, naval destroyer)
cannon = piece of artillery
cannelloni = tube-shaped pasta
cannelure = ring-shaped groove
cannula = thin tube

(**analog**: canoe = boat, from Arawakan)
canyon = narrow gorge
channel = conduit
canal = waterway

canalis - Latin
(pipe, duct)

A vocabulary of ancient peptides at the origin of folded proteins



B

paralogs (cofactor binding)

orthologs (RNA binding)

paralogs (structural)

exosome subunit RRP4
N-term domain (2JE6)

ribosomal protein L27
(1V8Q)

primordial
β-hammerhead motif

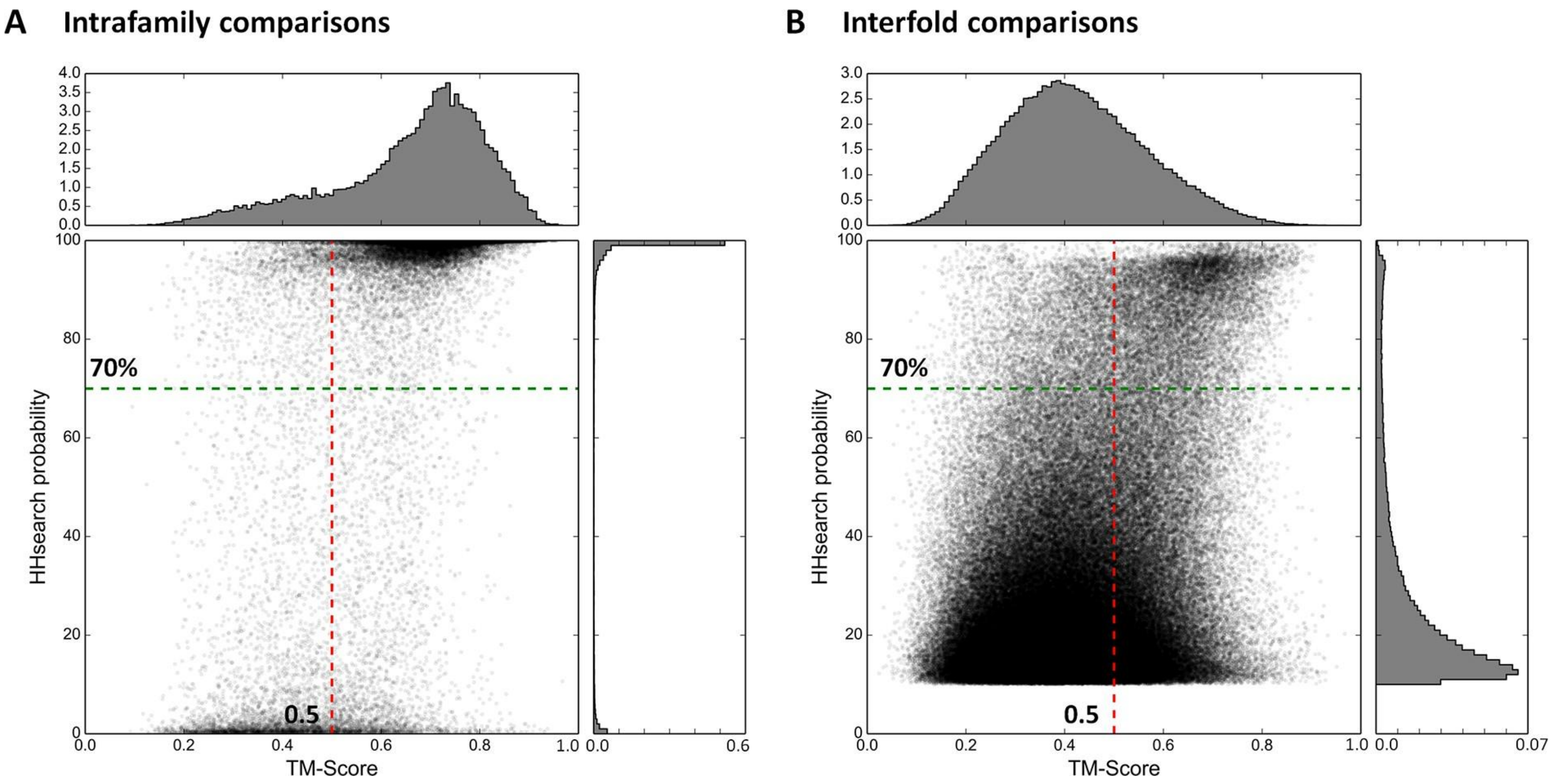
biotinyl domain of
acetyl-CoA carboxylase
(1BDO)

biotin carboxyl
carrier domain
of transcarboxylase (1DCZ)

RNA-polymerase
β′ subunit
insert domain (4G7H)

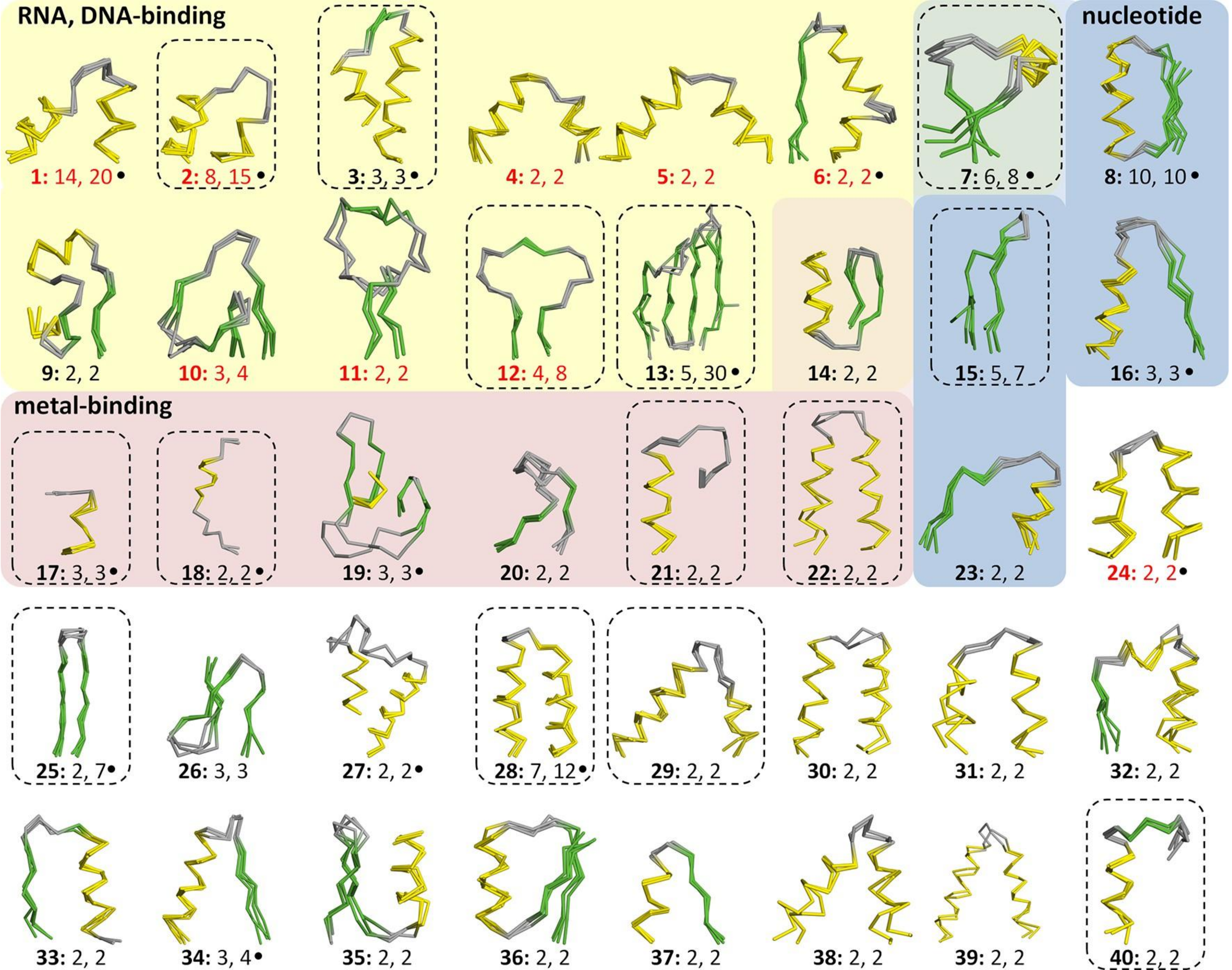
analog: molybdopterin
synthase subunit MoaE
(1FM0)

thymidine phophorylase
C-term domain (1UOU)

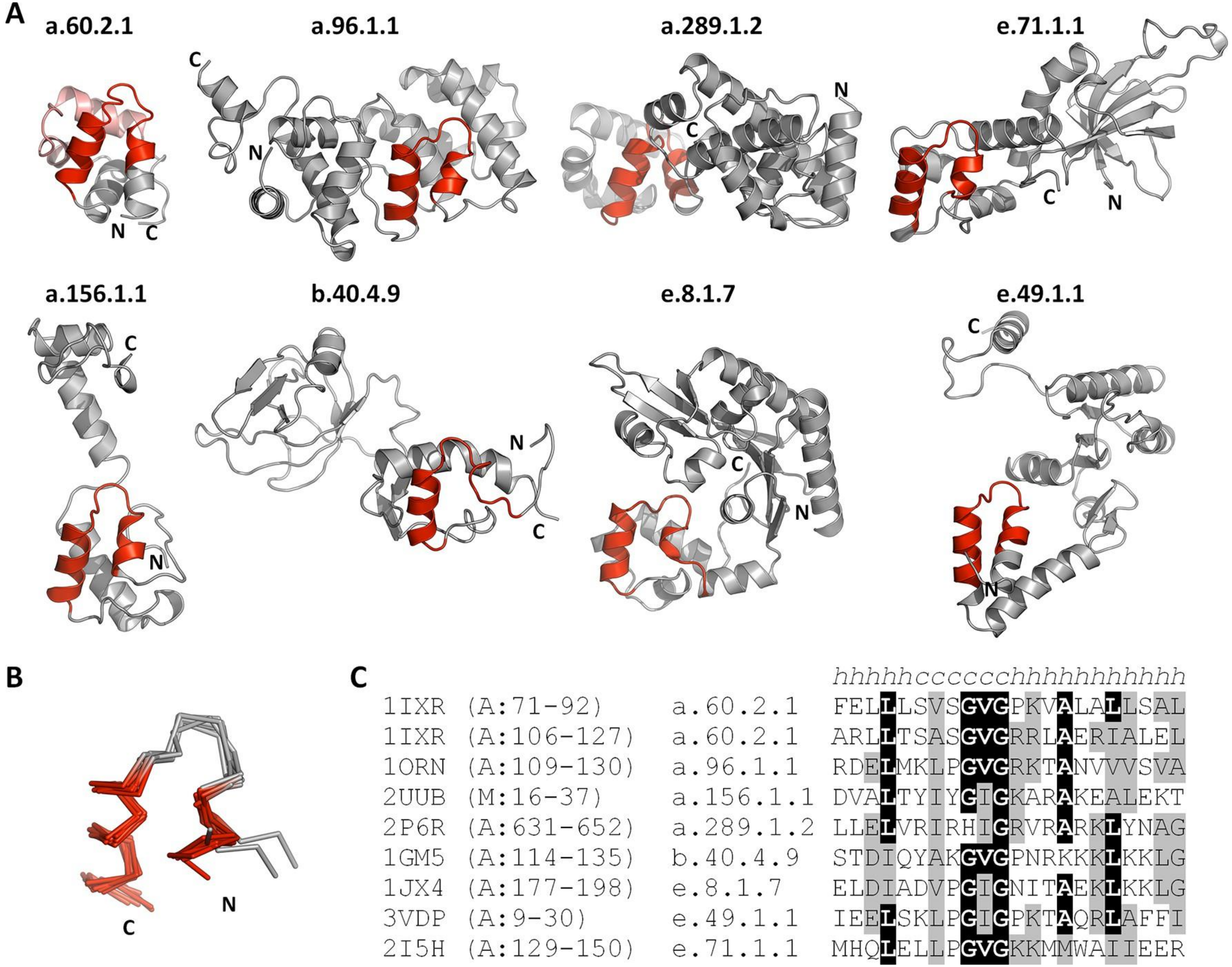


A Intrafamily comparisons

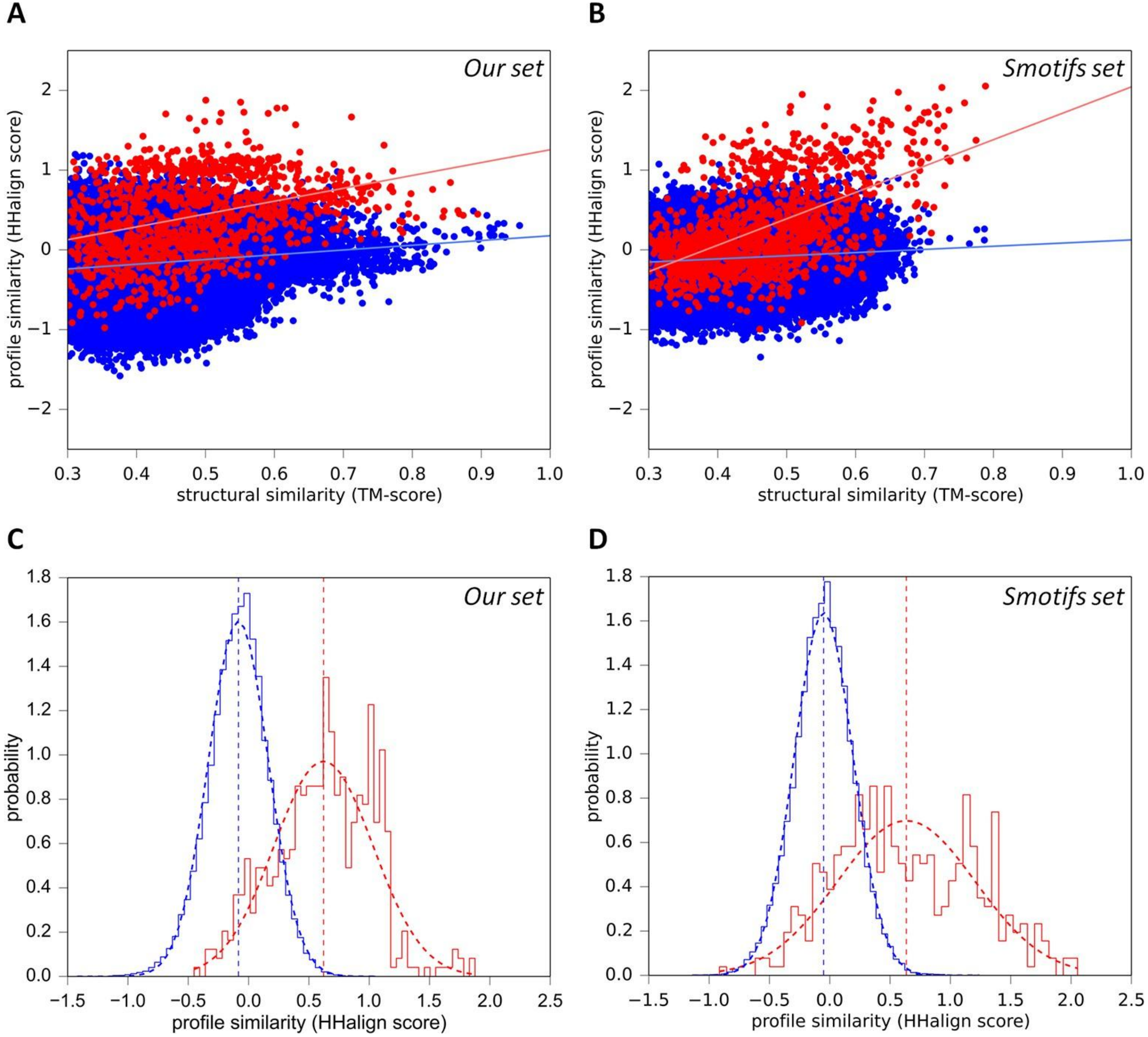
B Interfold comparisons


RNA, DNA-binding
nucleotide
metal-binding
1: 14, 20 •
2: 8, 15 •
3: 3, 3 •
4: 2, 2
5: 2, 2
6: 2, 2 •
7: 6, 8 •
8: 10, 10 •
9: 2, 2
10: 3, 4
11: 2, 2
12: 4, 8
13: 5, 30 •
14: 2, 2
15: 5, 7
16: 3, 3 •
17: 3, 3 •
18: 2, 2 •
19: 3, 3 •
20: 2, 2
21: 2, 2
22: 2, 2
23: 2, 2
24: 2, 2 •
25: 2, 7 •
26: 3, 3
27: 2, 2 •
28: 7, 12 •
29: 2, 2
30: 2, 2
31: 2, 2
32: 2, 2
33: 2, 2
34: 3, 4 •
35: 2, 2
36: 2, 2
37: 2, 2
38: 2, 2
39: 2, 2
40: 2, 2

The nucleic-acid binding helix-hairpin-helix motif is found in 8 different folds comprising 15 superfamilies.



			hhhhhcccccchhhhhhhhhhh
1IXR	(A:71-92)	a.60.2.1	FELLLSVSGVGPKVALALLSAL
1IXR	(A:106-127)	a.60.2.1	ARLLTSASGVGRRLAERIALEL
1ORN	(A:109-130)	a.96.1.1	RDELMKLPGVGRKTANVVVSVA
2UUB	(M:16-37)	a.156.1.1	DVALTYIYGIGKARAKEALEKT
2P6R	(A:631-652)	a.289.1.2	LLELVRIRHIGRVRARKLYNAG
1GM5	(A:114-135)	b.40.4.9	STDIQYAKGVGPNRKKKLKKLG
1JX4	(A:177-198)	e.8.1.7	ELDIADVPGIGNITAEKLKKLG
3VDP	(A:9-30)	e.49.1.1	IEELSKLPGIGPKTAQRLAFFI
2I5H	(A:129-150)	e.71.1.1	MHQLELLPGVGKKMMWAIIEER


A
Our set
profile similarity (HHalign score)
structural similarity (TM-score)
B
Smotifs set
profile similarity (HHalign score)
structural similarity (TM-score)
C
Our set
probability
profile similarity (HHalign score)
D
Smotifs set
probability
profile similarity (HHalign score)
2
1
0
−1
−2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5