V.A.F 2022.09.20

Proteínas y su dualidad discreta-continua

Martín Graña – Unidad de Bioinformática – Institut Pasteur de Montevideo

Harpers, issue 179, June/November 1939

THE USEFULNESS OF USELESS KNOWLEDGE

BY ABRAHAM FLEXNER

Natural Selection and the Concept of a Protein Space

form a
ossible.
asses of
ls, and
)-acids,
nd (ii)
no-acid
inver-

JOHN MAYNARD SMITH

School of Biological Sciences,
University of Sussex.

Received November 7, 1969.

[1] Salisbury, F. B., *Nature*, **224**, 342 (1969).
[2] Maynard Smith, J., in *The Scientist Speculates* (edit. by Good, I. J.) (Heinemann, London, 1961).
[3] King, J. L., and Jukes, T. H., *Science*, **164**, 788 (1969).

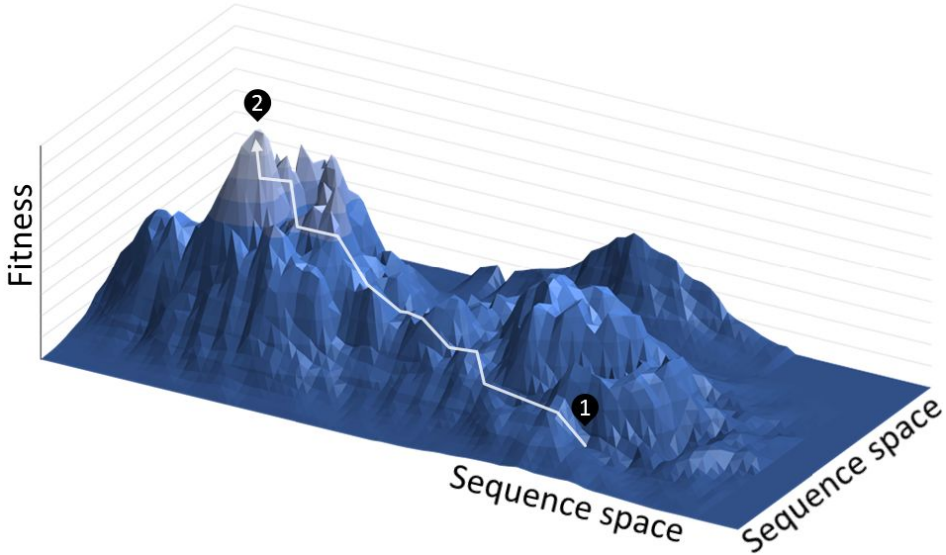
PROTEINS

One thousand families for the molecular biologist

Cyrus Chothia

Sequence space

Natural Selection and the Concept of a Protein Space



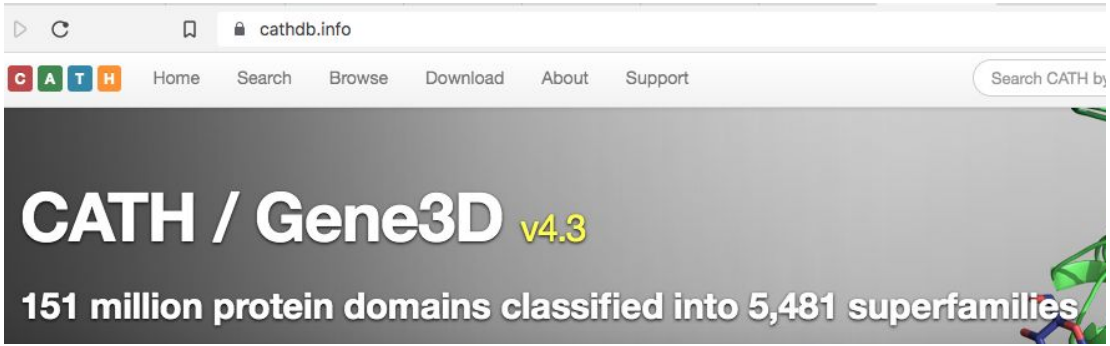
SCOPe Browse Stats & History Downloads Help

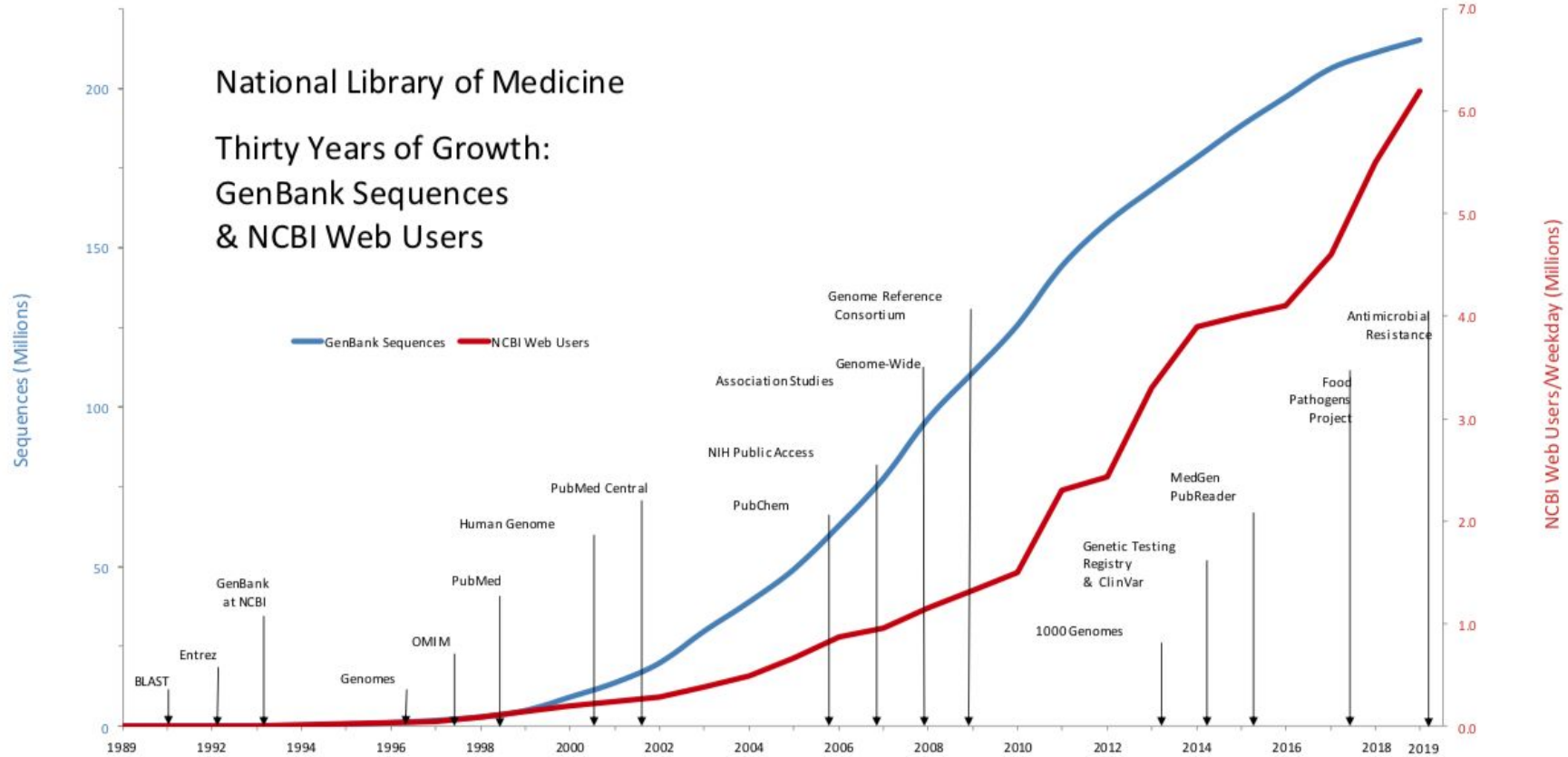
Statistics

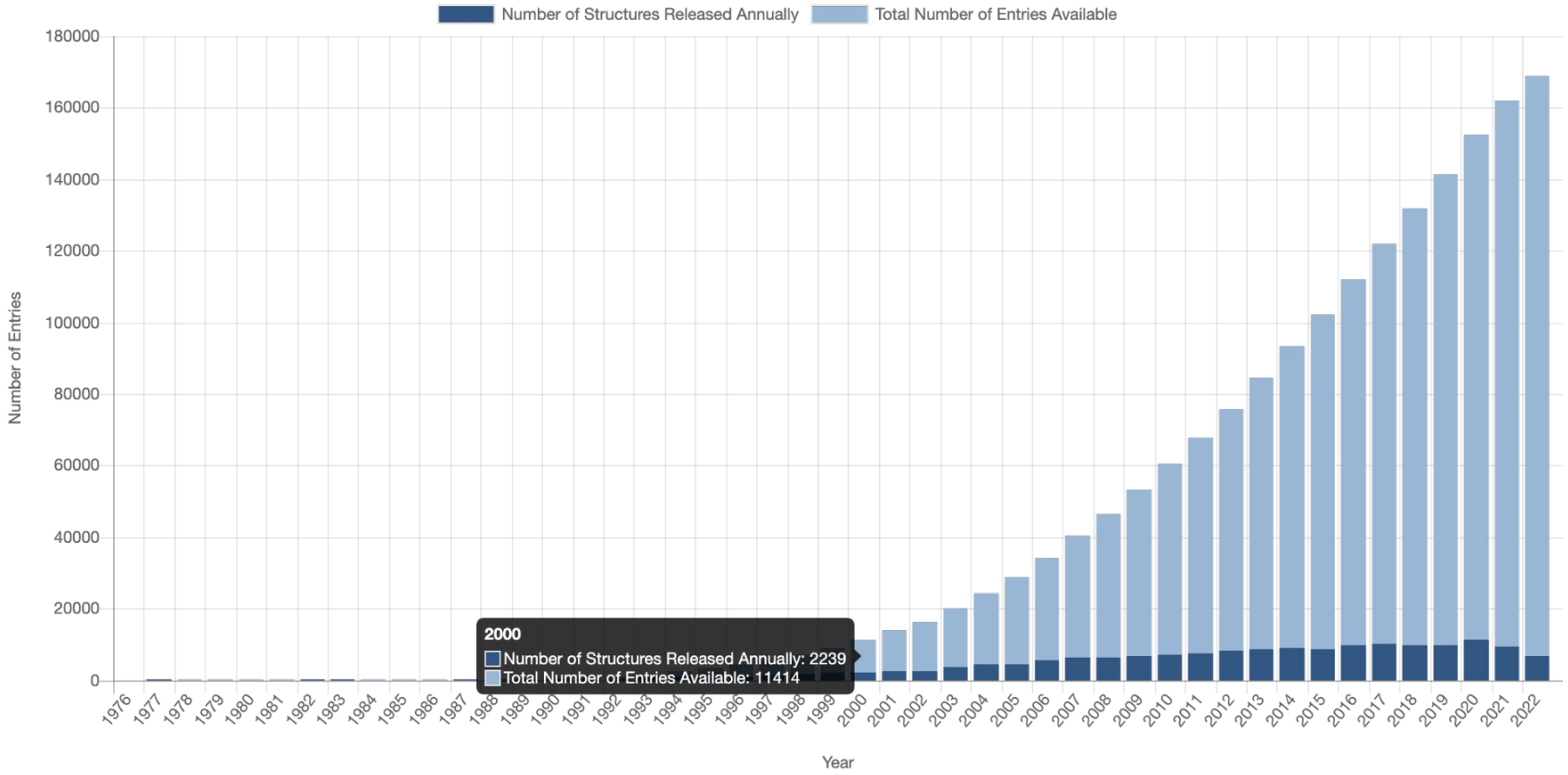
SCOPe 2.08-stable statistics:
106976 PDB entries (released/updated prior to 2021-07-28). 344851 Domains. 1 Literature reference.

Class	Number of folds	Number of superfamilies	Number of families
a: All alpha proteins	290	519	1089
b: All beta proteins	180	375	993
c: Alpha and beta proteins (a/b)	148	247	1003
d: Alpha and beta proteins (a+b)	396	580	1387
e: Multi-domain proteins (alpha and beta)	74	74	128
f: Membrane and cell surface proteins and peptides	69	131	204
g: Small proteins	100	141	280
Totals	**1257** (26 new)	**2067** (42 new)	**5084** (88 new)

Stats for the latest *periodic* (info) release, **SCOPe** 2.08-2022-02-10 (adding PDB entries released/updated prior to 2022-02-10):
107643 PDB entries. 346905 Domains. 1 Literature reference.



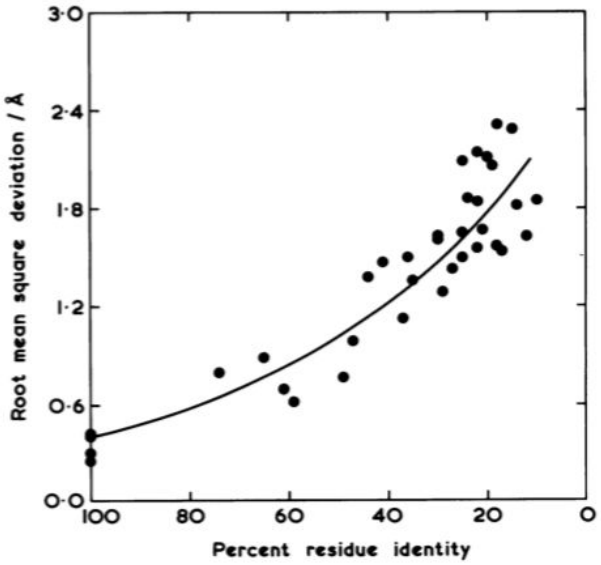

National Library of Medicine
Thirty Years of Growth:
GenBank Sequences
& NCBI Web Users
Sequences (Millions)
NCBI Web Users/Weekday (Millions)
GenBank Sequences
NCBI Web Users
BLAST
Entrez
GenBank at NCBI
Genomes
OMIM
PubMed
Human Genome
PubMed Central
PubChem
NIH Public Access
Association Studies
Genome-Wide
Genome Reference Consortium
1000 Genomes
Genetic Testing Registry & ClinVar
MedGen PubReader
Food Pathogens Project
Antimicrobial Resistance
0
50
100
150
200
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
1989
1992
1994
1996
1998
2000
2002
2004
2006
2008
2010
2012
2014
2016
2018
2019


Number of Structures Released Annually
Total Number of Entries Available
Number of Entries
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
2000
Number of Structures Released Annually: 2239
Total Number of Entries Available: 11414
1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022
Year

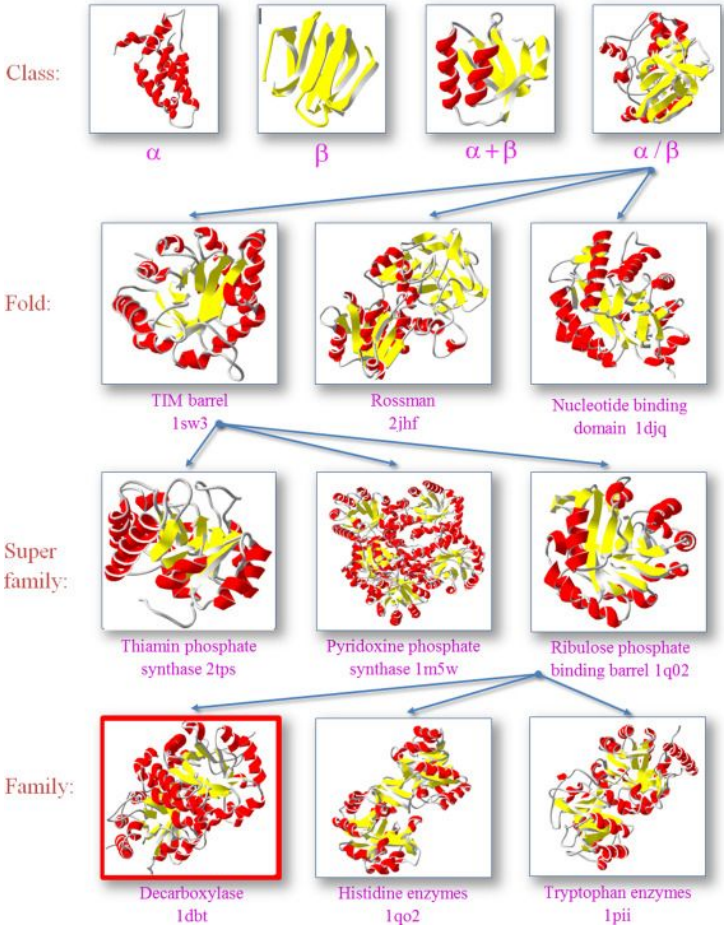
The EMBO Journal vol.5 no.4 pp.823–826, 1986

The relation between the divergence of sequence and structure in proteins

Cyrus Chothia[1] and Arthur M.Lesk[2]



Discrete folds



"Super folds"

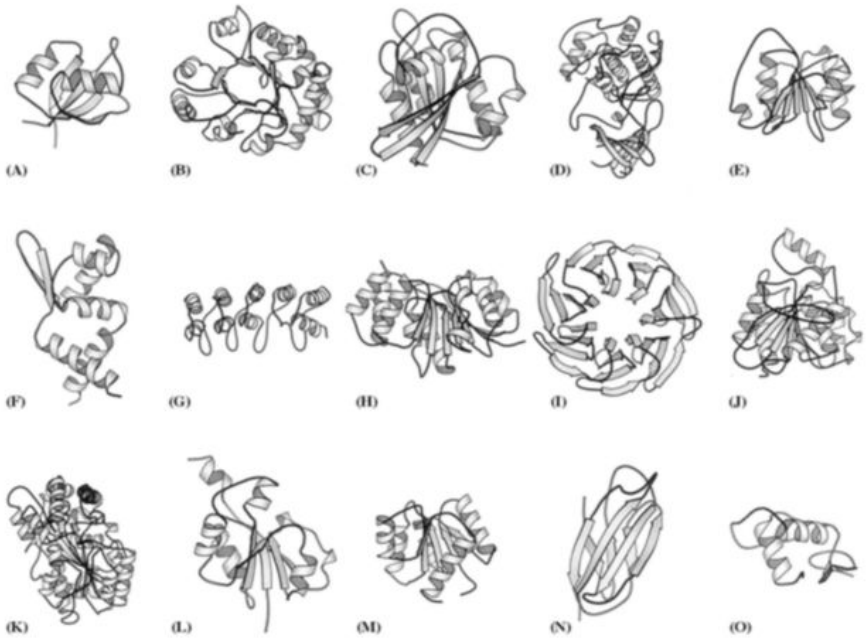


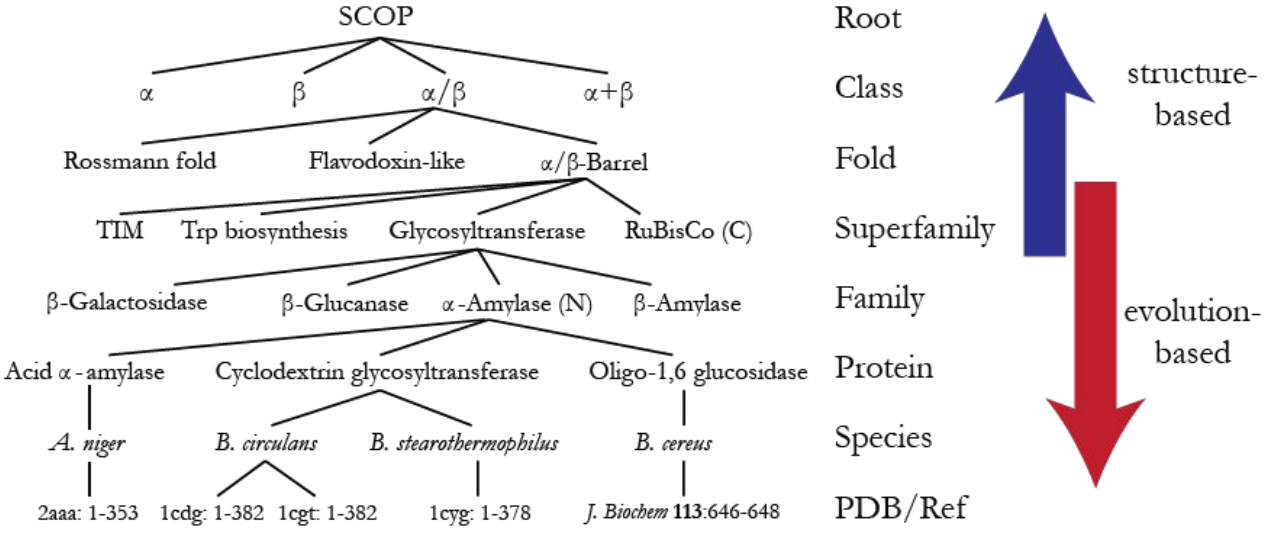
Figure 1. The 15 most populated folds. They were selected on the basis of a structural annotation of proteins from completely sequenced genomes of 20 bacteria, five Archaea, and three eukaryotes [C. Zhang, unpublished data]. From left to right and top to bottom, they are: ferredoxin-like (4.45%) (*A*), TIM-barrel (3.94%) (*B*), P-loop containing nucleotide triphosphate hydrolase (3.71%) (*C*), protein kinases (PK) catalytic domain (3.14%) (*D*), NAD(P)-binding Rossmann-fold domains (2.80%) (*E*), DNA/RNA-binding 3-helical bundle (2.60%) (*F*), α-α superhelix (1.95%) (*G*), S-adenosyl-L-methionine-dependent methyltransferase (1.92%) (*H*), 7-bladed beta-propeller (1.85%) (*I*), α/β-hydrolases (1.84%) (*J*), PLP-dependent transferase (1.61%) (*K*), adenine nucleotide α-hydrolase (1.59%) (*L*), flavodoxin-like (1.49%) (*M*), immunoglobulin-like β-sandwich (1.38%) (*N*), and glucocorticoid receptor-like (0.97%) (*O*), where the values in parentheses are the percentages of annotated proteins adopting the respective folds.

On the Universe of Protein Folds

Rachel Kolodny,[1] Leonid Pereyaslavets,[2]
Abraham O. Samson,[3] and Michael Levitt[2]

Annu. Rev. Biophys. 2013. 42:559–82

Discrete folds



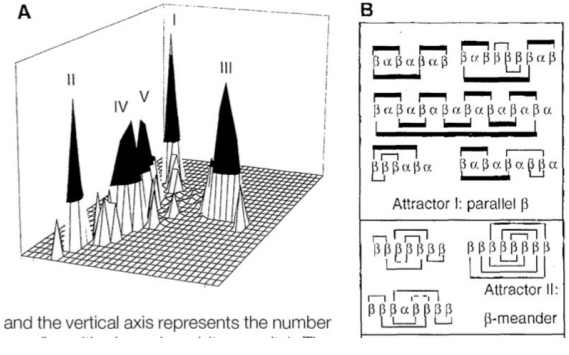
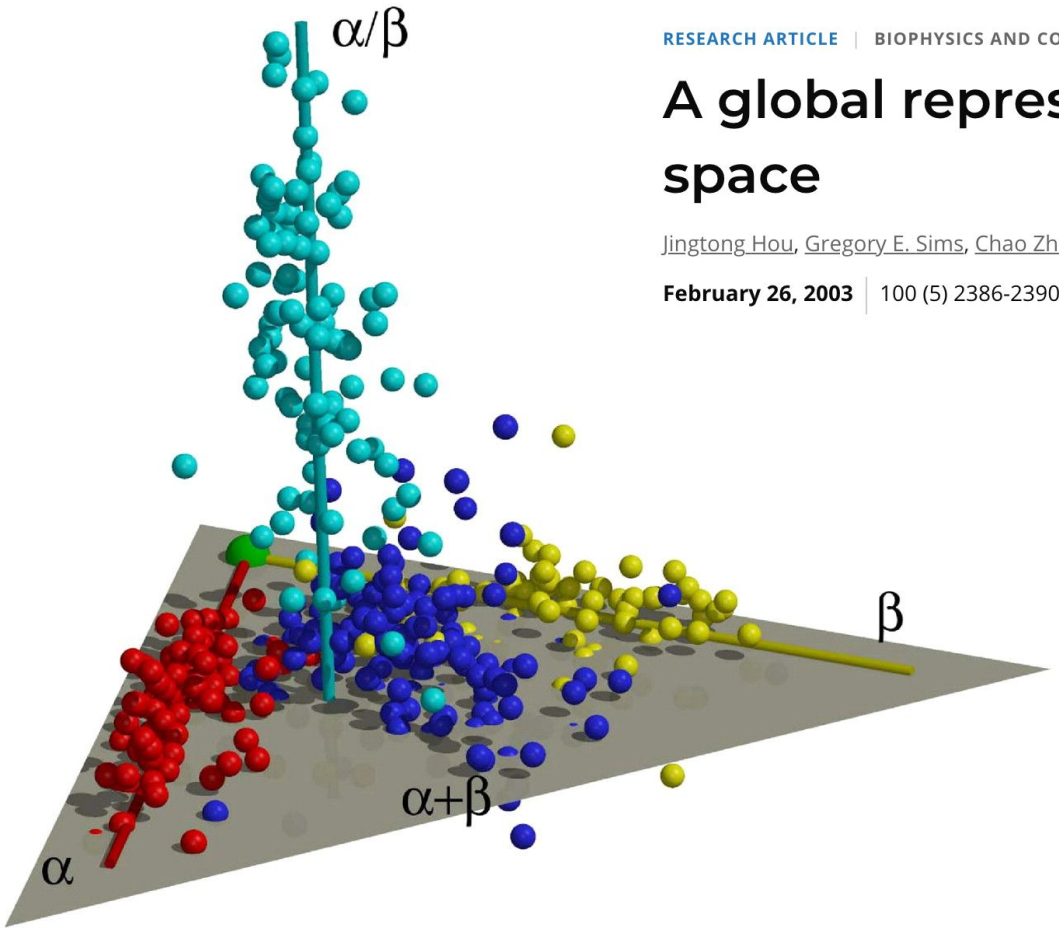


Fig. 5. Fold space attractors. (**A**) Quantification of the pairwise structural similarities in an all-on-all comparison of protein structures allows one to position each structure relative to the others in an abstract, high-dimensional fold space (shape space). The height of the peaks reflects population density (of folds in fold space). The horizontal axes are the two dominant eigenvalues (*21*), and the vertical axis represents the number

SSM: sequence structure maps



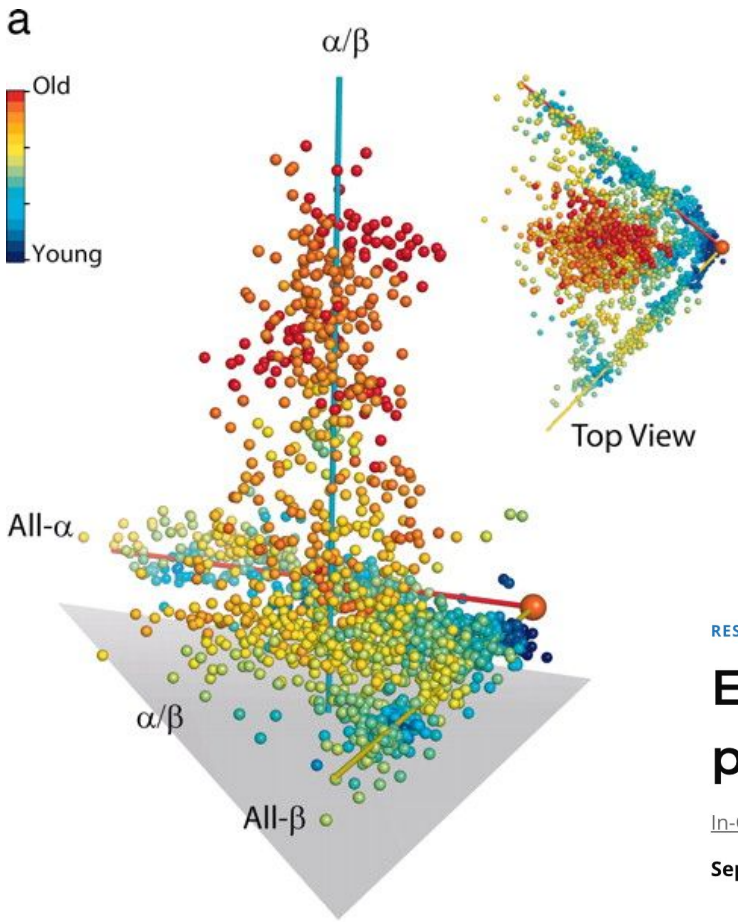
RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY |

A global representation of the protein fold space

Jingtong Hou, Gregory E. Sims, Chao Zhang, and Sung-Hou Kim Authors Info & Affiliations

February 26, 2003 | 100 (5) 2386-2390 | https://doi.org/10.1073/pnas.2628030100

SSM: sequence structure maps



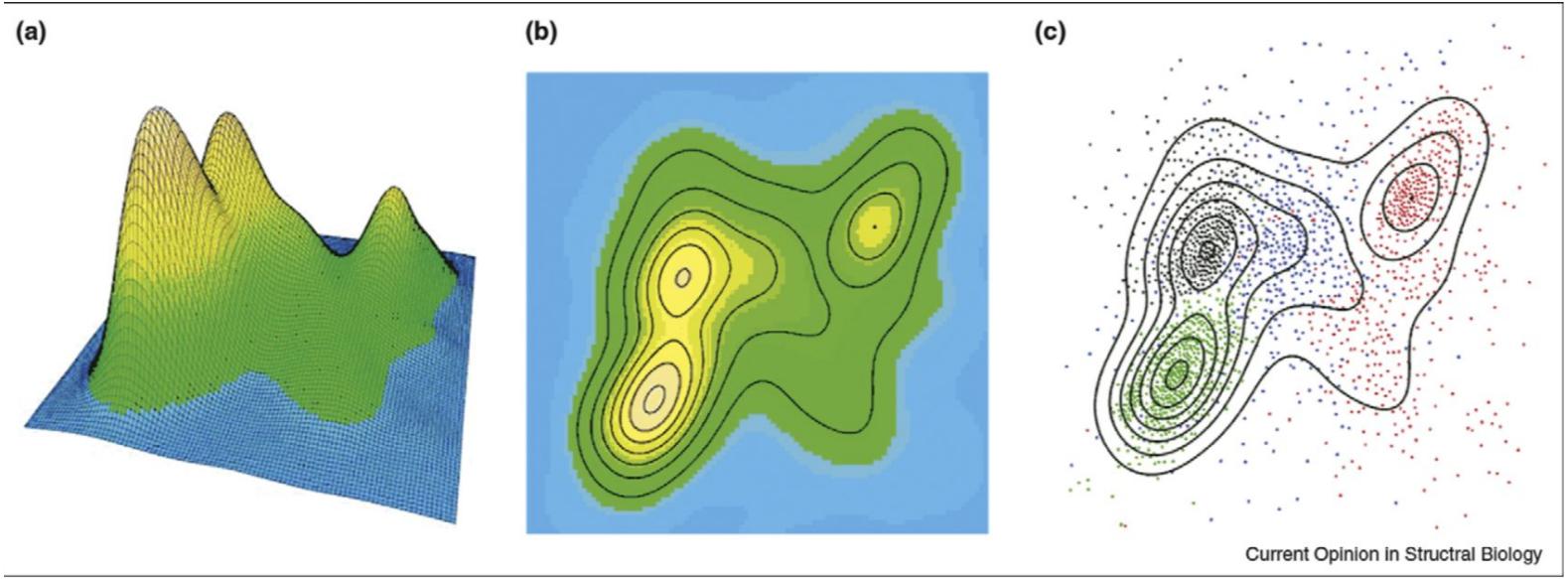
RESEARCH ARTICLE | BIOLOGICAL SCIENCES |

Evolution of protein structural classes and protein sequence families

In-Geol Choi and Sung-Hou Kim Authors Info & Affiliations

September 19, 2006 | 103 (38) 14056-14061 | https://doi.org/10.1073/pnas.0606239103

Traversing sequence space



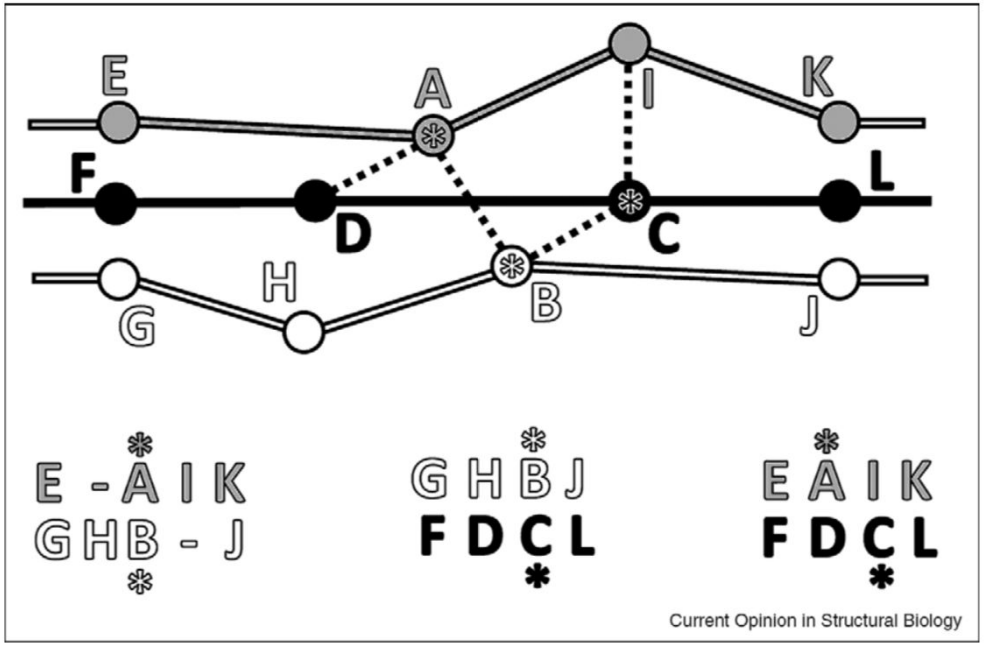
all alpha BLACK

all beta RED

alpha + beta BLUE

alpha/beta GREEN

Non transitivity of structure-based alignments



Nontransitivity of structure-based alignments. Unlike homology-based alignments, alignments guided purely by structure geometry do not necessarily have the property of transitivity. Structure-based alignment of fragments of three protein chains (colored white, black, and gray, with circles representing C-alpha atoms) is schematically shown. Residues are aligned based on the criterion of minimal distance (marked by dotted lines). In this example, pairwise alignments between residues A, B, and C (marked with asterisks both in the schema and in the alignments below) are not transitive: A is aligned to B and B to C, yet A and C are not aligned.

Biophysical Journal Volume 73 November 1997 2393–2403

How Are Model Protein Structures Distributed in Sequence Space?

Erich Bornberg-Bauer
Abteilung 0815 Theoretische Bioinformatik, Deutsches Krebsforschungszentrum Im Neuenheimer Feld 280, Heidelberg, D-69120, Germany

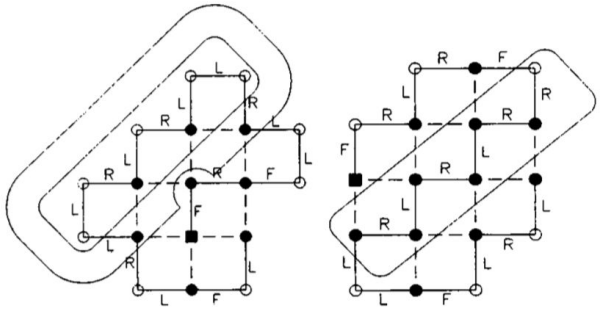


FIGURE 1 Examples of frequent structures. *Left:* The most frequent structure as formed by a typical sequence. Closed circles denote **H**'s, open circles **P**'s, solid lines correspond to peptide bonds connecting two subsequent residues, dashed lines are energy-contributing contacts between 2 **H**'s. Letters along the bonds [F (*forward*), L (*left*), and R (*right*)] denote the corresponding relative moves. Squares symbolize the first residue since the structure is not considered identical as the results from reverse sequences. The first move is F by definition, the first non-F move R. The structure can be encoded as FRFLLRLLRLRLLRLFL. Frequent motifs are boxed (see text). *Right:* The most frequent maximum compact structure. It can be encoded as FRLRFRRLRLRLLFLRL.

Zipf distribution of structures

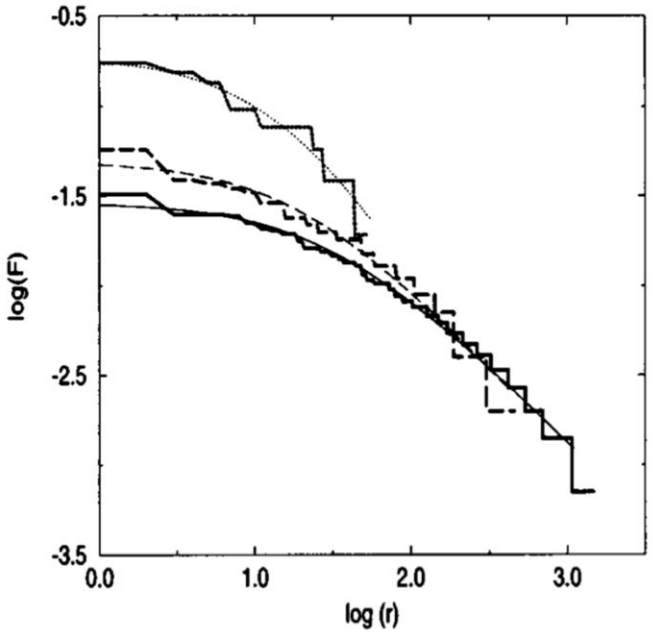


FIGURE 2 The frequency distribution of structures. Zipf plot showing the log of the frequence distribution of structures versus the log of their rank r. Results for ground states of uniquely folding sequences are shown for chain lengths $n = 13$ (*dotted*), 16 (*dashed*), and 18 (*solid line*). Corresponding fits to a generalized Zipf's law are drawn in thin lines.

Biophysical Journal Volume 73 November 1997 2393–2403

How Are Model Protein Structures Distributed in Sequence Space?

Erich Bornberg-Bauer
Abteilung 0815 Theoretische Bioinformatik, Deutsches Krebsforschungszentrum Im Neuenheimer Feld 280, Heidelberg, D-69120, Germany

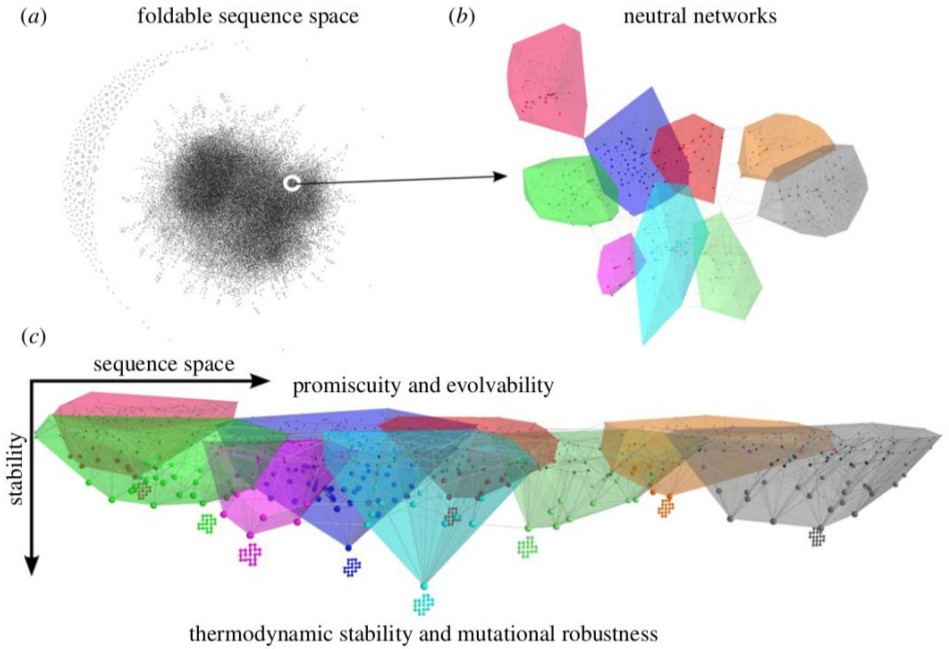
Mutational stability, prototype sequences and neutral nets

JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Headline review

Cite this article: Sikosek T, Chan HS. 2014 Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* **11**: 20140419.
http://dx.doi.org/10.1098/rsif.2014.0419



Mutational stability, prototype sequences and neutral nets

JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Headline review

Cite this article: Sikosek T, Chan HS. 2014 Biophysics of protein evolution and evolutionary protein biophysics. *J. R. Soc. Interface* **11**: 20140419.
http://dx.doi.org/10.1098/rsif.2014.0419