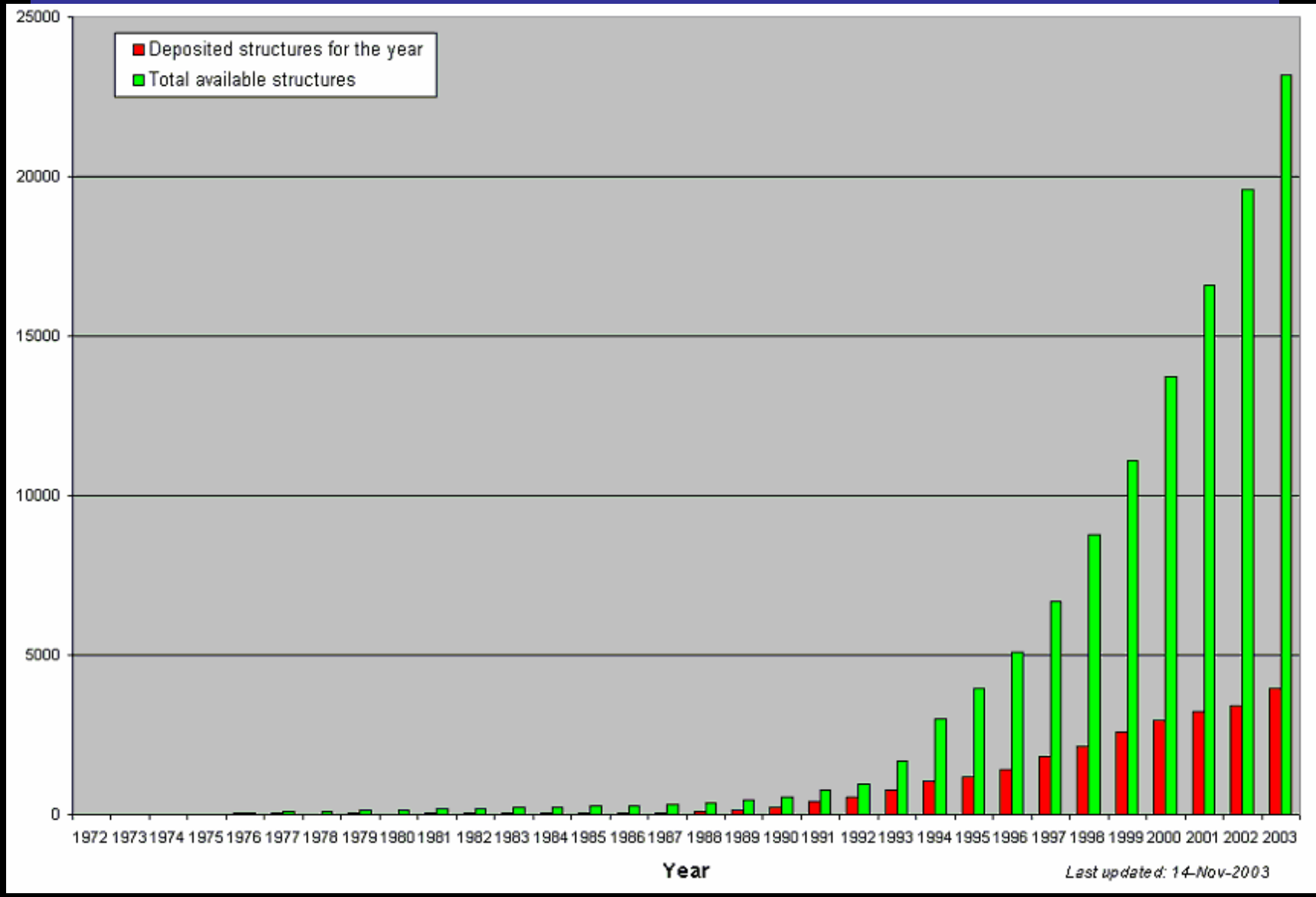
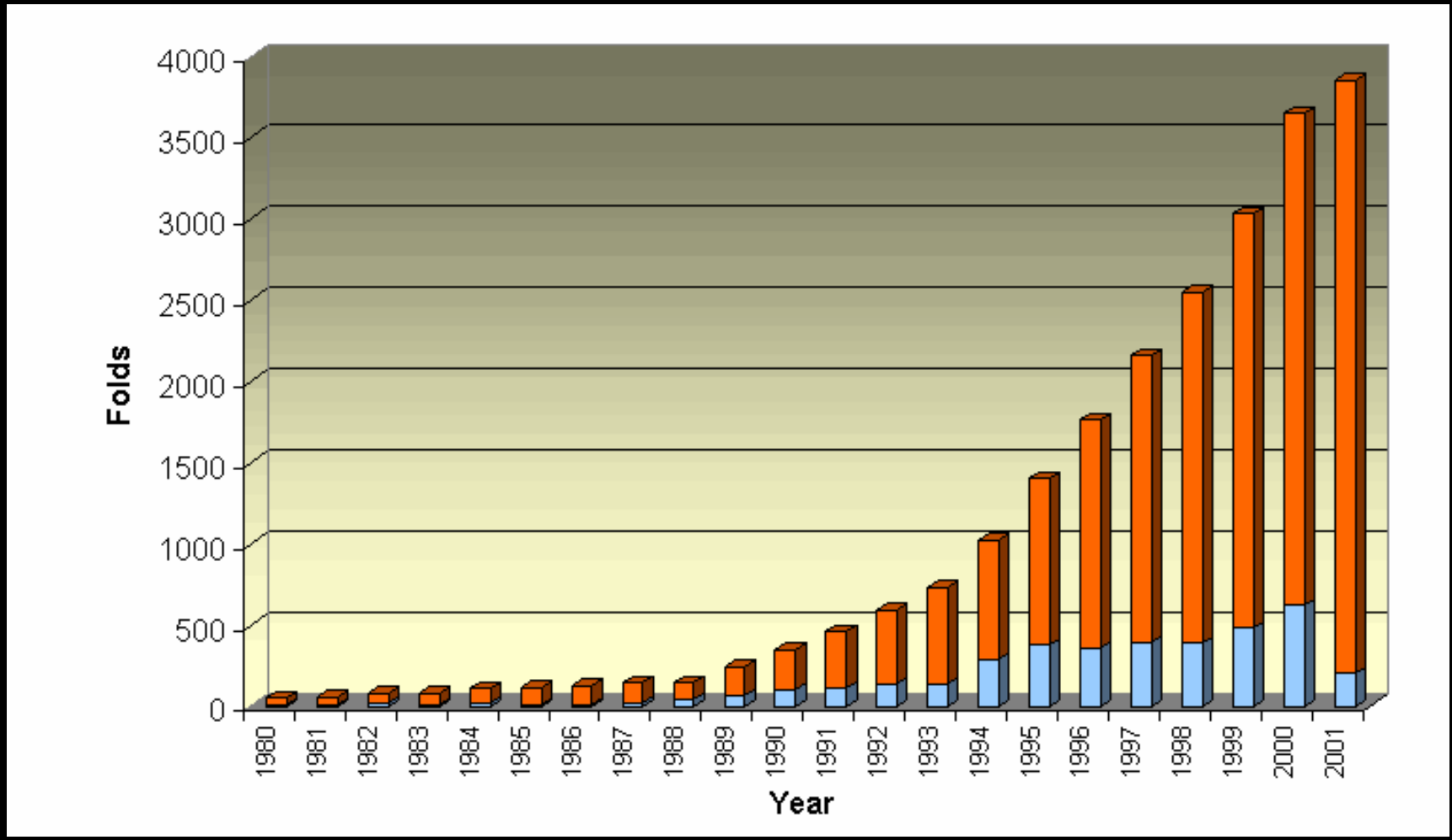
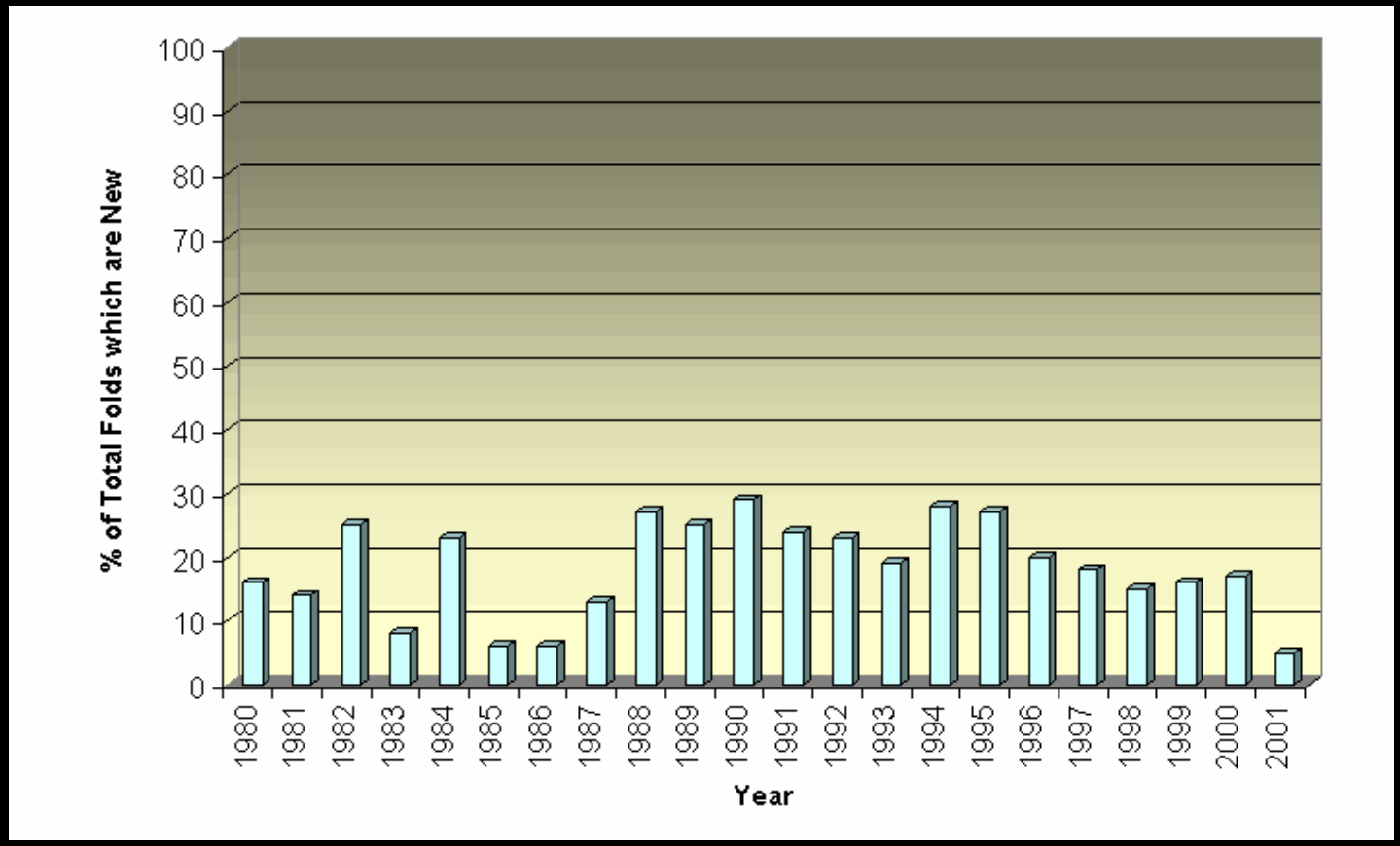
Growth of the Protein Data Bank



Newly Determined Structures- Number of New Folds



Newly Determined Structures- Fraction of New Folds



Classification of structures

SCOP: http://scop.mrc-lmb.cam.ac.uk/scop/

(domains, good annotation)

CATH: http://www.biochem.ucl.ac.uk/bsm/cath/

CE: http://cl.sdsc.edu/ce.html

Dali Domain Dictionary:

http://columba.ebi.ac.uk:8765/holm/ddd2.cgi

FSSP: http://www2.ebi.ac.uk/dali/fssp/

(chains, updated weekly)

HOMSTRAD:

http://www-cryst.bioc.cam.ac.uk/~homstrad/

HSSP: http://swift.embl-heidelberg.de/hssp/

SCOP Hierarchy of Structures

Fanily: evolutionarily related with a significant sequence identity - 2327 in SCOP

Superfamily: different families whose structural and functional features suggest common evolutionary origin -1294 in SCOP

Fold: different superfamilies having same major secondary structures in same arrangement and with same topological connections - 800 in SCOP

Scop Classification Statistics

. **1.65** release 20619 PDB Entries (1 August 2003).

54745 Domains

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	179	299	480
All beta proteins	126	248	462
Alpha and beta proteins (a/b)	121	199	542
Alpha and beta proteins (a+b)	234	349	567
Multi-domain proteins	38	38	53
Membrane and cell surface proteins	36	66	73
Small proteins	66	95	150
Total	800	1294	2327

Scop Classification Statistics

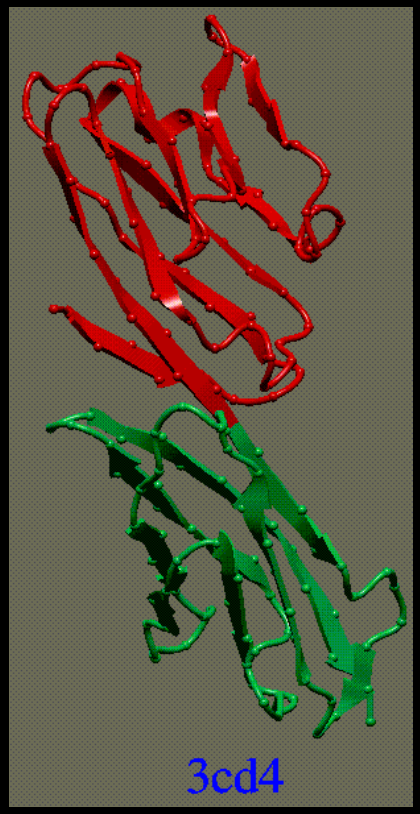
1.69 release

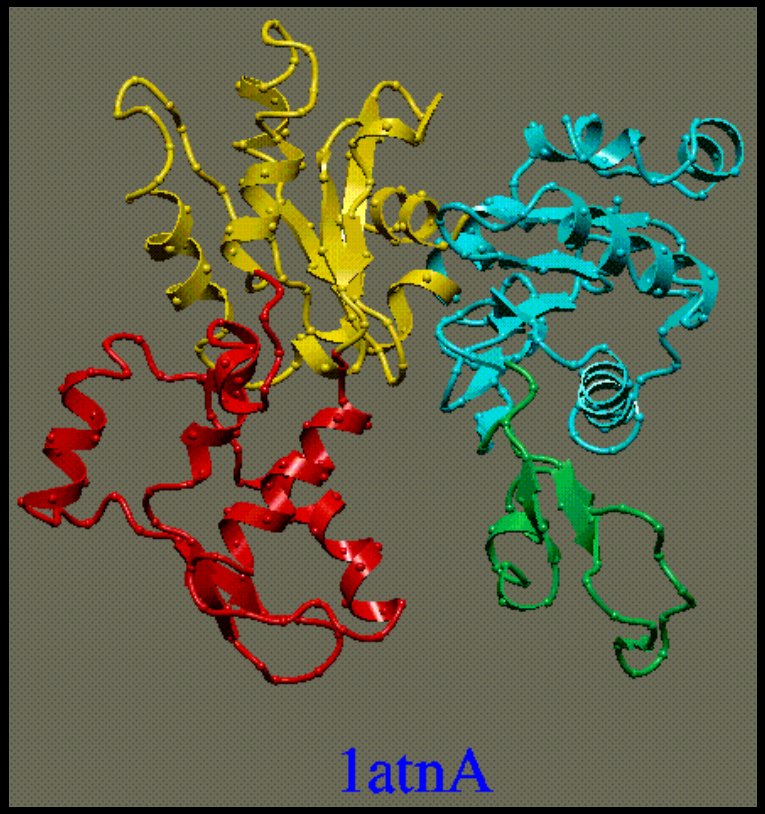
25973 PDB Entries (1 Oct 2004). 70859 Domains

(excluding nucleic acids and theoretical models)

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	218	376	608
All beta proteins	144	290	560
Alpha & beta proteins (α/β)	136	222	629
Alpha & beta proteins (α+β)	279	409	717
Multi-domain proteins	46	46	61
Membrane and cell surface proteins	47	88	99
Small proteins	75	108	171
Total	945	1539	2845

Chain/Domain Library





Domain may be more sensitive but depends on correct partition

Hundreds of thousands of gene sequences translated to proteins (SwissProt, PIR)

27,112 solved structures (PDB)

as of Sept 7, 2004

Goals:

- **Predict structure from sequence**
- **Predict function based on sequence**
- **Predict function based on structure**

PDB Current Holdings Breakdown

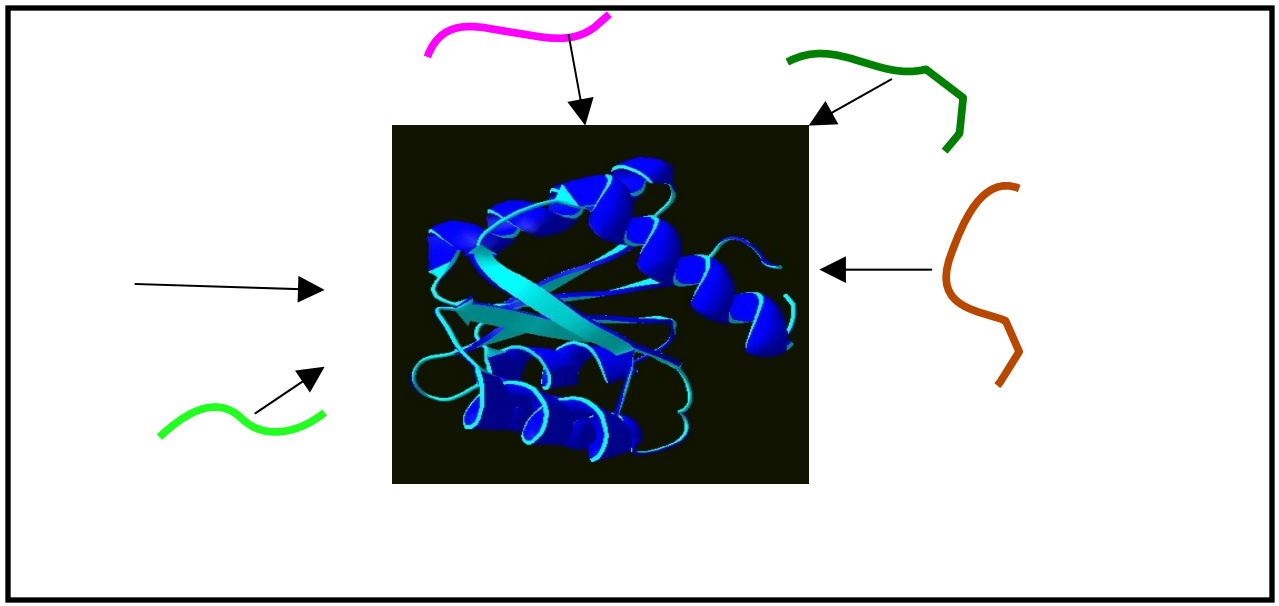
Tuesday May 30, 2006

Molecule Type

		Proteins	Nucleic Acids	Protein/NA Complexes	Other	**Total**
Exp. Method	**X-ray**	28945	900	1350	28	**31223**
	NMR	4583	699	121	6	**5409**
	Electron Microscopy	88	9	28	0	**125**
	Other	73	4	3	0	**80**
	Total	**33689**	**1612**	**1502**	**34**	**36837**

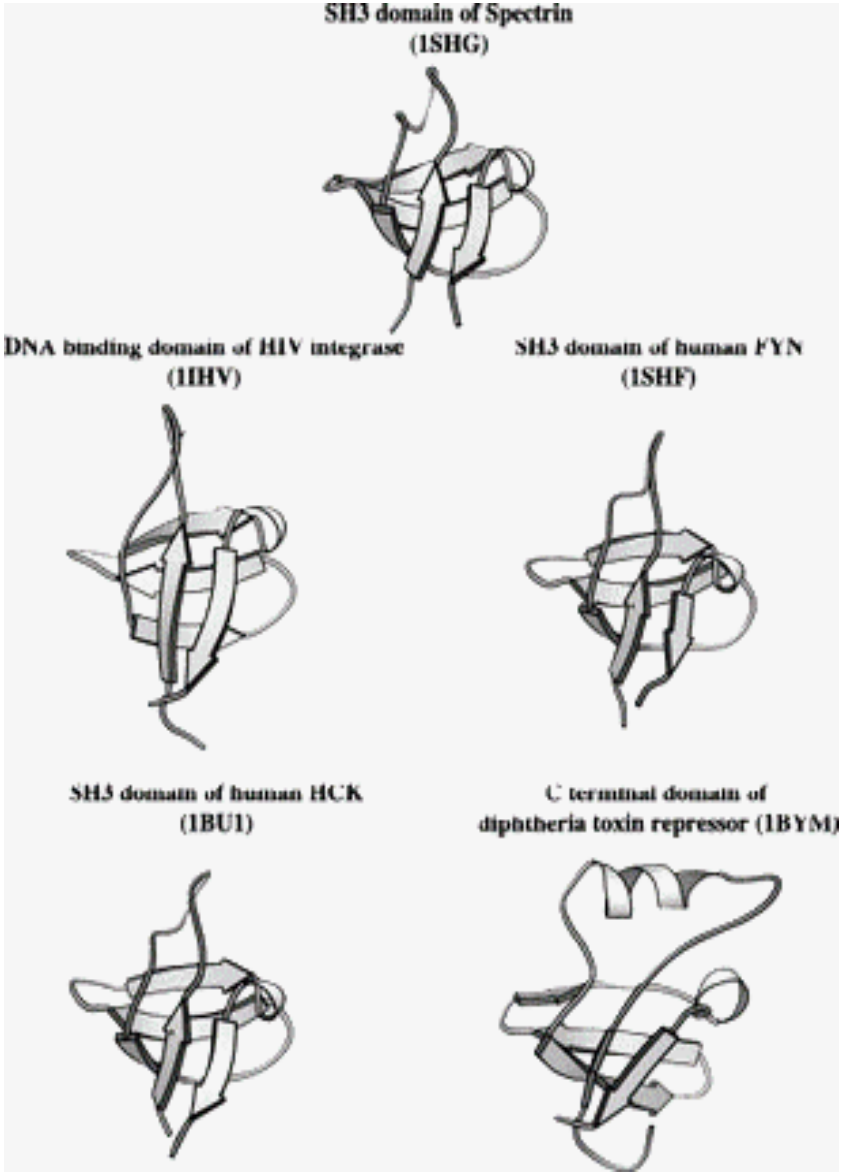
Many proteins with dissimilar sequences fold into similar structures

- Estimated number of folds: 600-5000



Protein Folds: sequential and spatial arrangement of secondary structures

Sequence diverges faster than structure



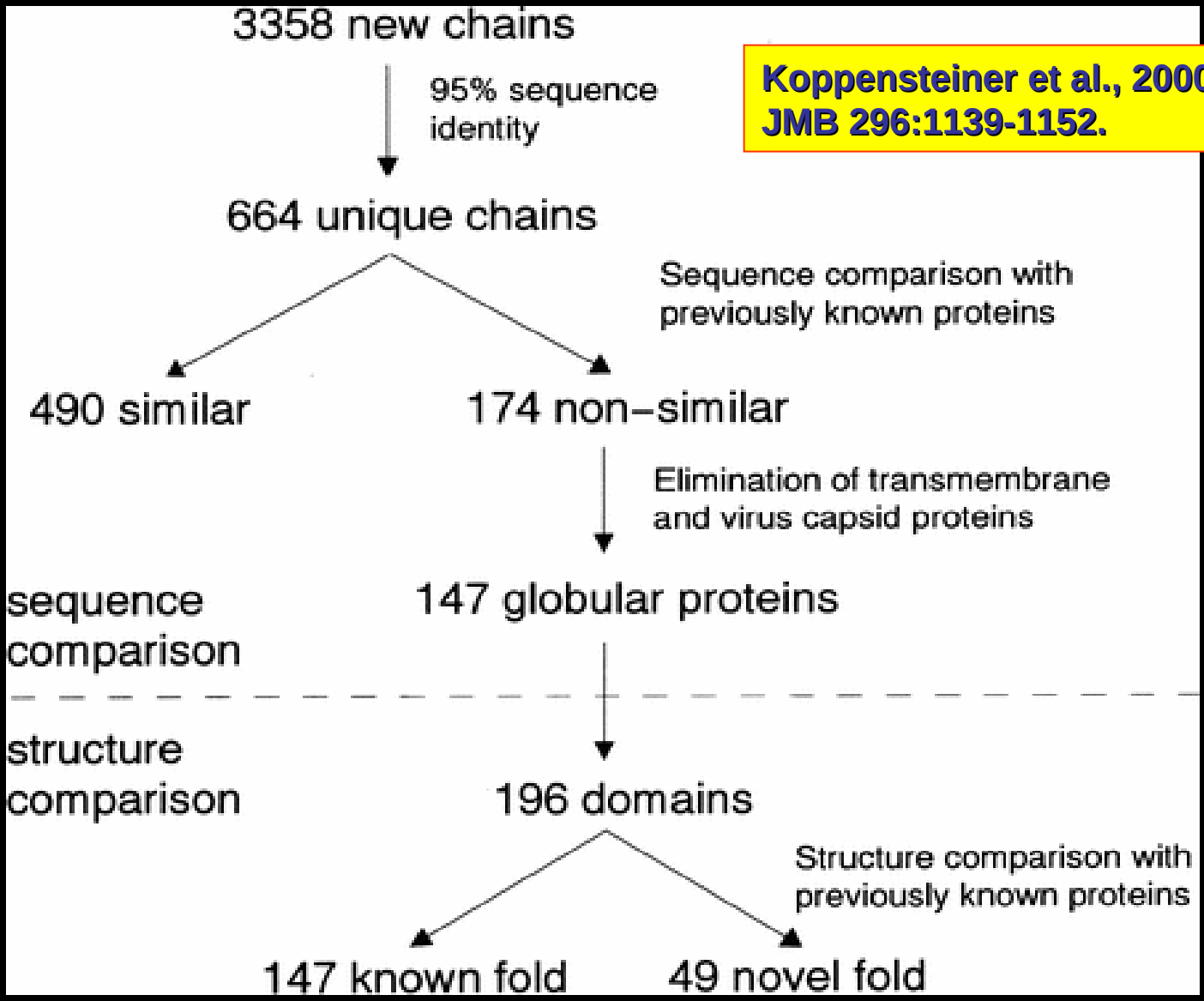
SH3 domains are small protein modules. Five **SH3 domains** are shown.

Structural similarities are observed among these different protein families

but

no detectable sequence similarities are found.

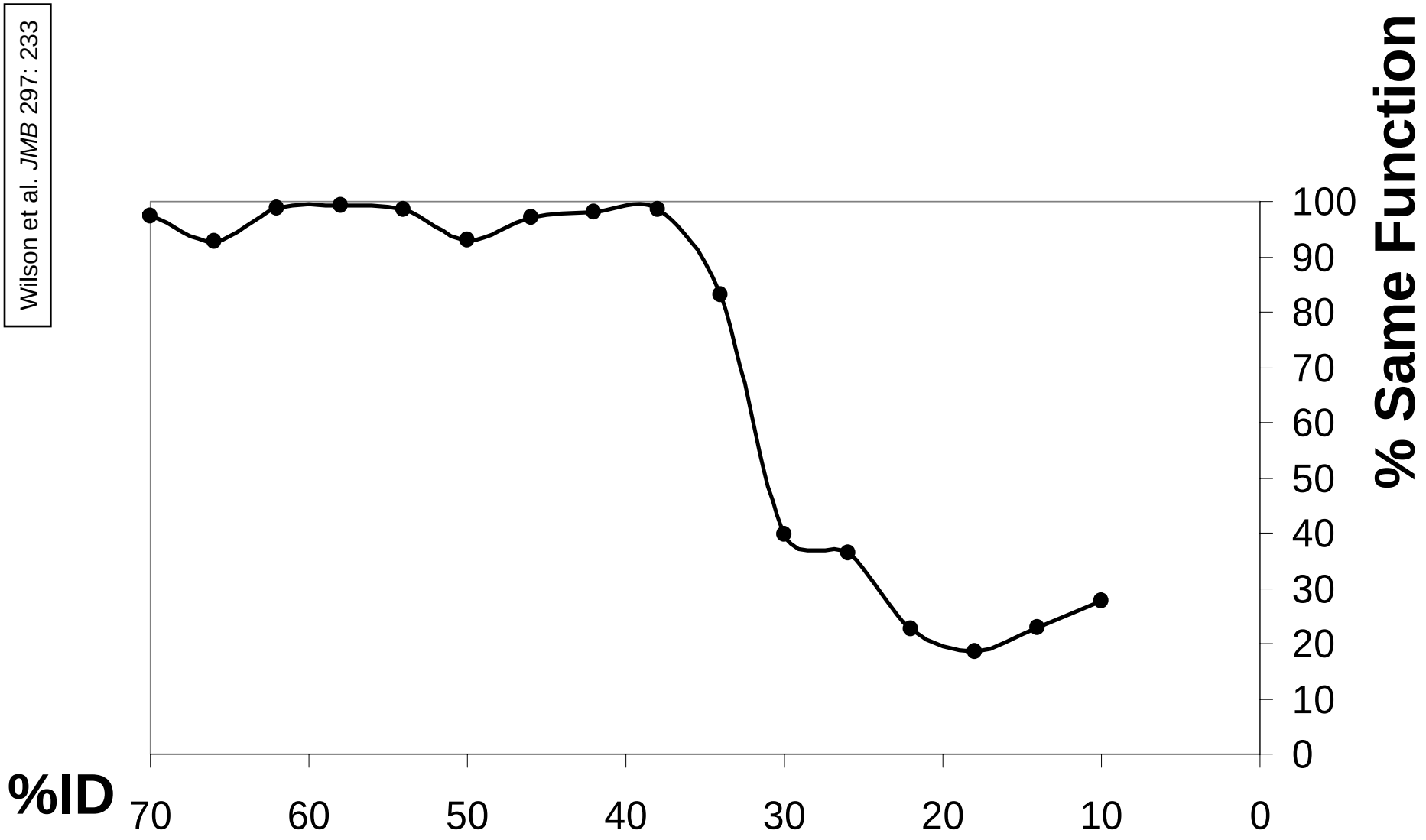
Fraction of new folds (PDB new entries in 1998)



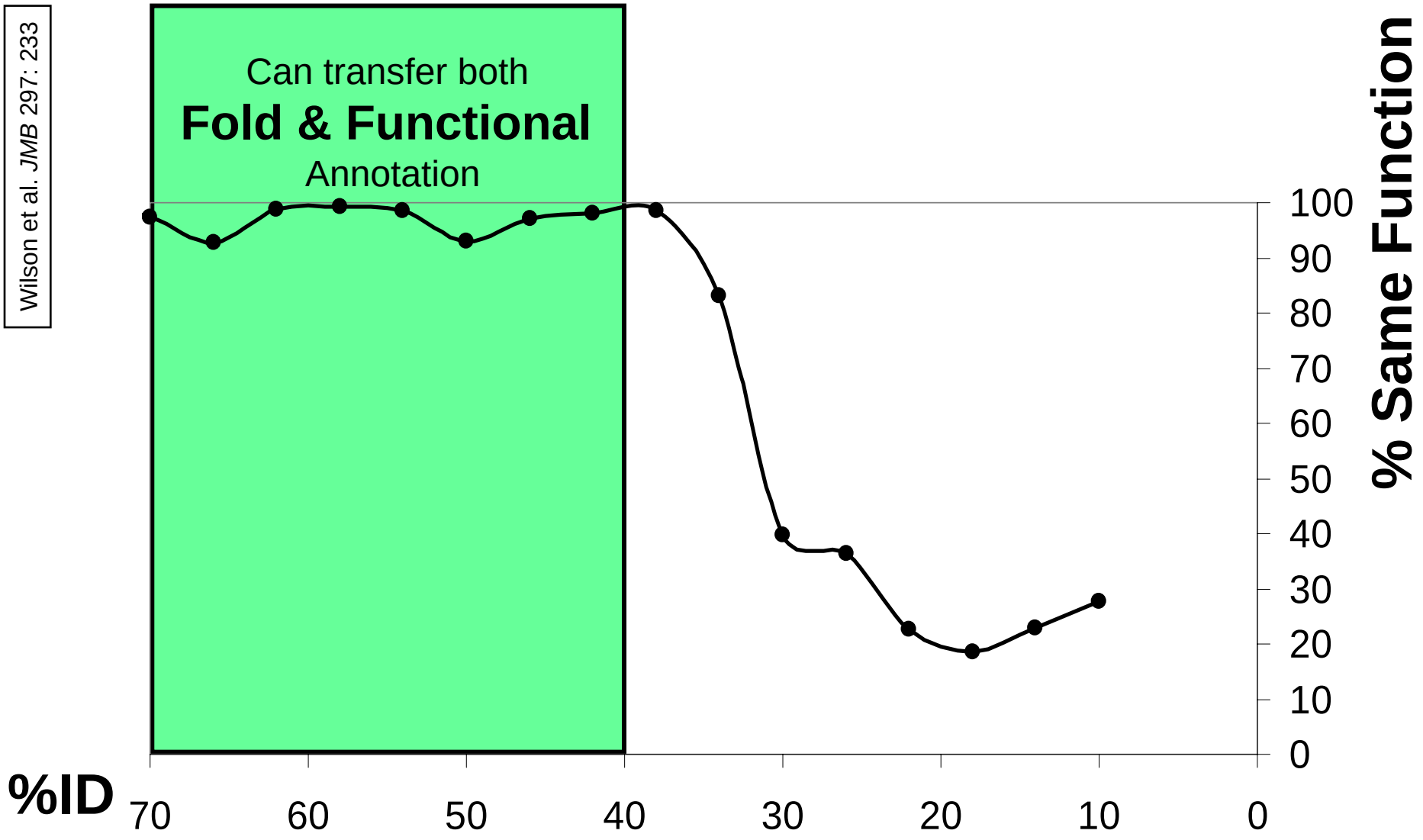
Koppensteiner et al., 2000, JMB 296:1139-1152.

Relationship of Similarity in Sequence to that in Function

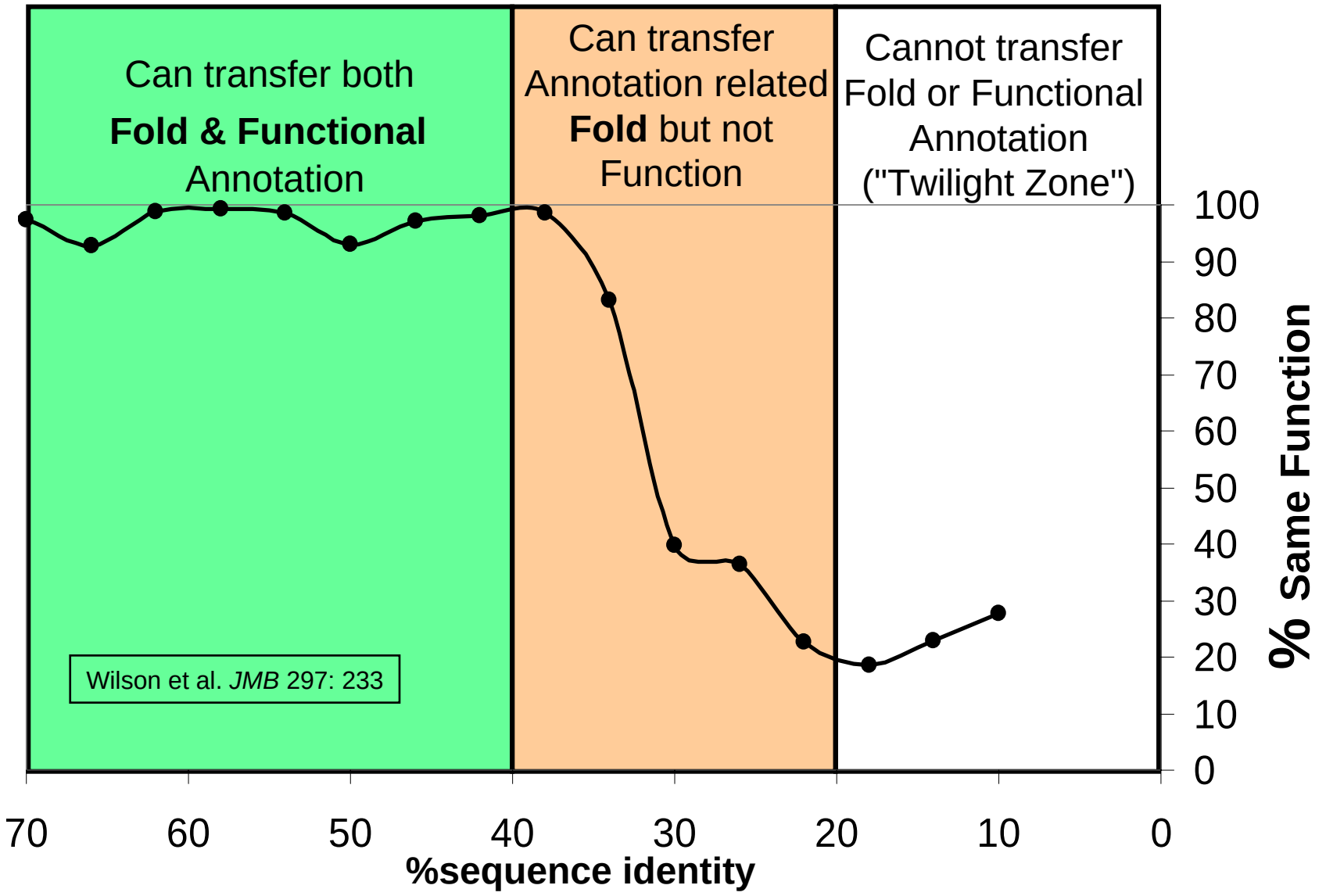
Wilson et al. *JMB* 297: 233



Relationship of Similarity in Sequence to that in Function

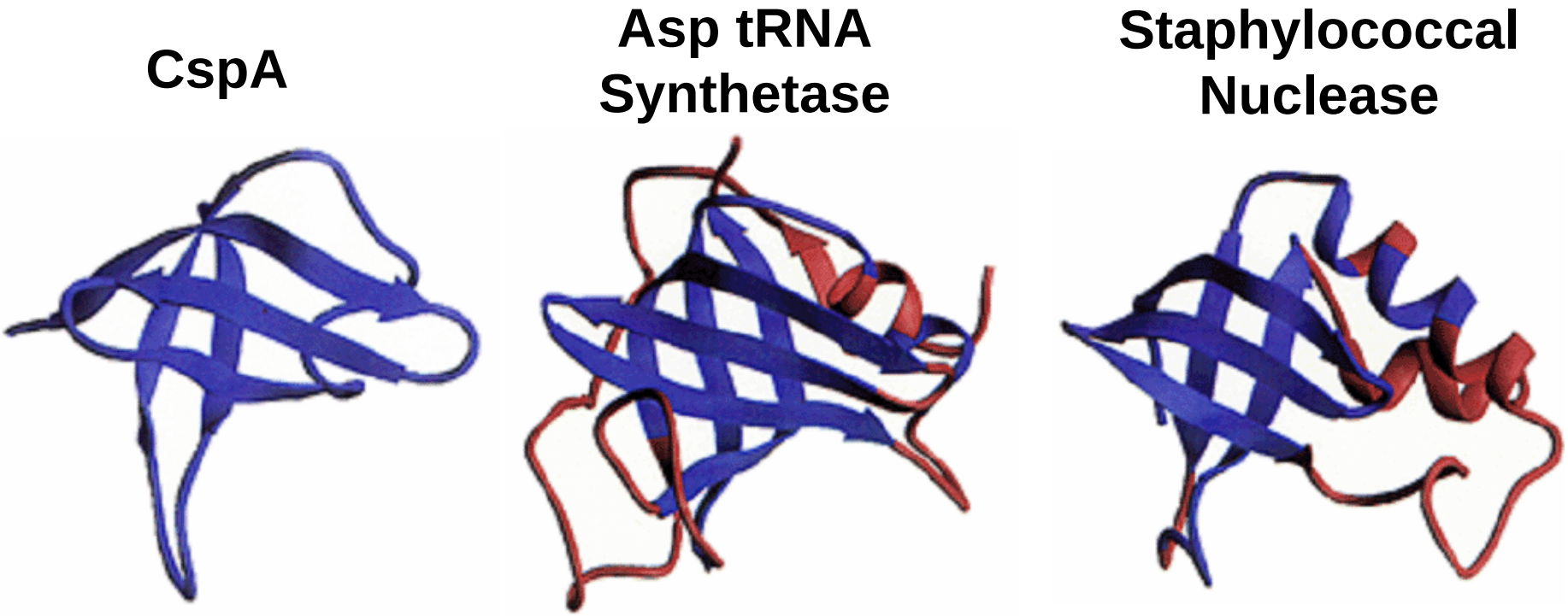


Relationship of Similarity in Sequence to that in Function



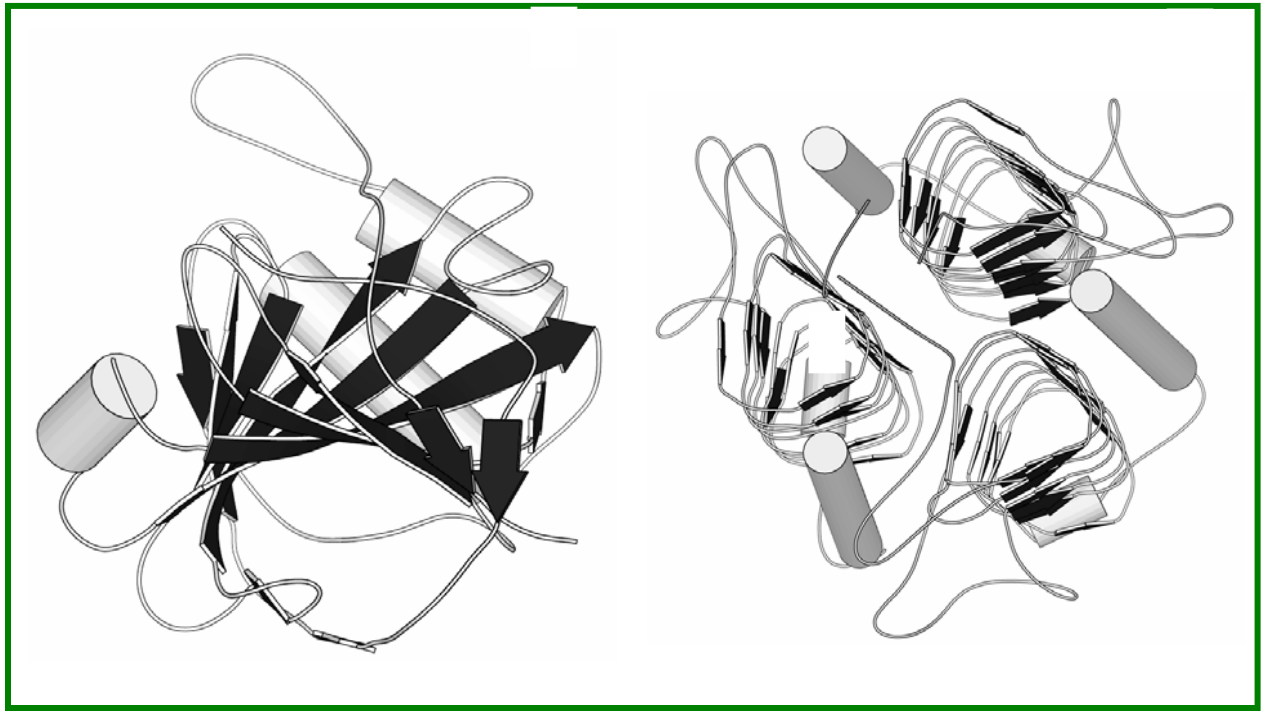
Can we predict function via fold similarity?

E.G. cspA OB fold suggests DNA binding [Montelione]

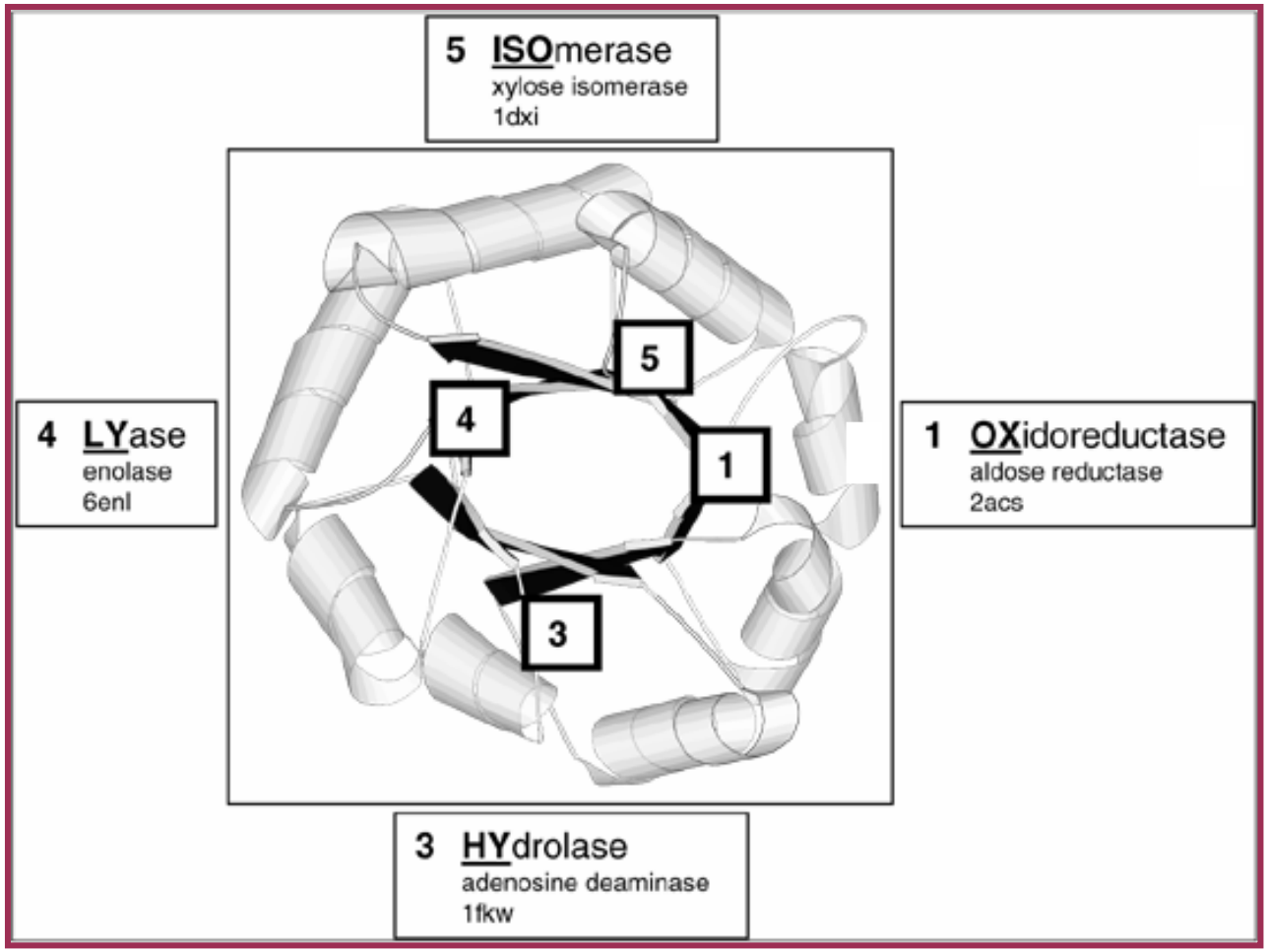


Issue: To what degree does fold determine function, globally?

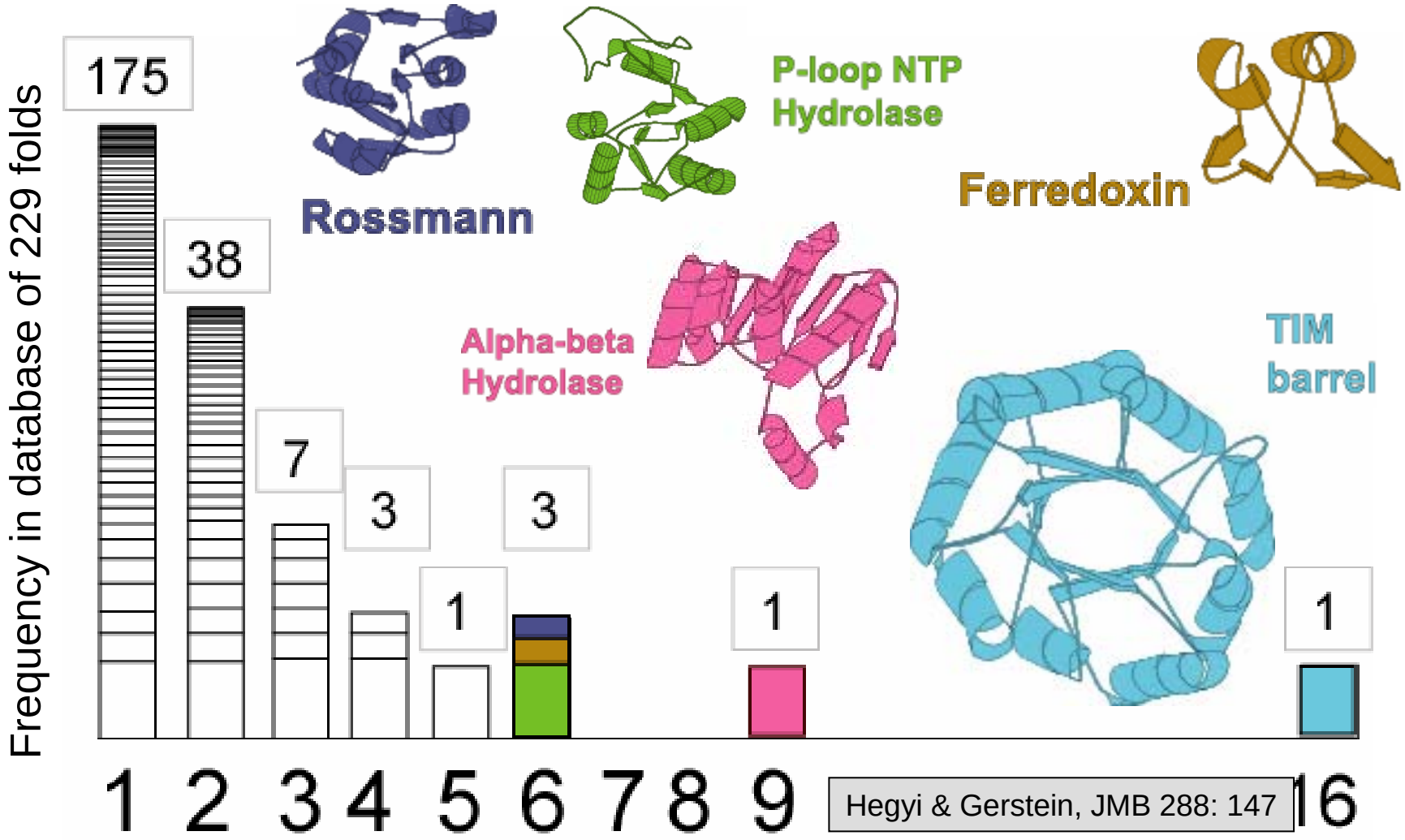
Different Folds with Same Function (Carbonic Anhydrases, 4.2.1.1)



Many Functions on Same Scaffold (TIM-barrel)



To what degree is fold associated with function? Folds with multiple functions



[Similar results by Thornton]