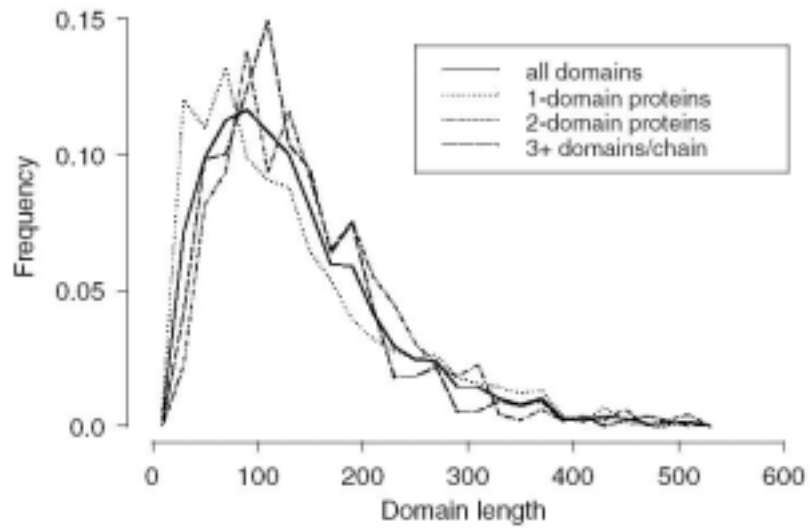# (domain boundary)

1.

100

. 200

(

200)

(1).

(domain) (2).

.

Size of protein domains
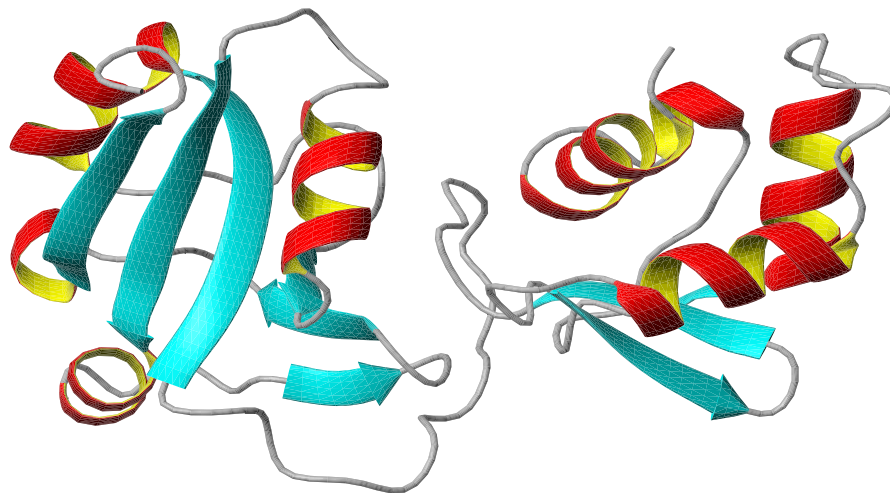
1:
(x- :) [1].

(folding) .

(fold)

.

(fold) .

SCOP(Structure
Classification   Of   Proteins;   http://scop.mrc-
lmb.cam.ac.uk/scop/index.html)   CATH(Class, Architecture, Topology,
and Homology; http://www.cathdb.info/latest/index.html)   .

(linker)

2:

SCOP(1.67 release)   PDB(Protein Data Bank)   24037
    65122   (fold)
        (1   ).   1   class   Multi-
domain proteins   ,
                .   1   fold
                    , superfamily
    , family   25%   .
    , SCOP   alpha-helix   (
)   202   (fold)   ,
    globin-like fold   .   fold   6
alpha-helix   . globin-like fold   globin-like
superfamily   alpha-helical ferredoxin superfamily   .
    4   family   2   family   .

globin- like superfamily   family   globins
family   hemoglobin   myoglobin   .

1: SCOP

Class	Number of folds	Number of superfamilies	Number of families
All alpha proteins	202	342	550
All beta proteins	141	280	529
Alpha/beta proteins	130	213	593
Alpha+beta proteins	260	386	650
Multi- domain proteins	40	40	55
Membrane and cell surface proteins	42	82	91
Small proteins	72	104	162
Total	887	1447	2630

2.

(sequence)

(domain   boundary)   .

(bioinformatics)

.

'   .

.

.

.

.

SnapDRAGON   [2]   RosettaDom   .

.

'

.   (1
)   DGS(Domain Guess by Size)   [1]. DGS

.   PSIPRED
DomSSEA   [3]
linker(2   )   DomCut   [4]   .
PSI- BLAST
"   (   )   "(position- specific scoring
matrix;   PSSM   )   . PSSM
DOMAINATION   [5]   ,
.   PSSM   (neural
network)   PPRODO(Prediction   of   PROtein   DOmain
boundaries)   [6]   .   (homology)
.

[   ]

[1] S. J. Wheelan, A. Marchler- Bauer, and S. H. Bryant, Domain size
    distributions can predict domain boundaries, Bioinformatics **16** (2000)
    613- 618.

[2] R. A. George and J. Heringa, SnapDRAGON: a method to delineate
    protein structural domains from sequence data, Journal of Molecular
    Biology **316** (2002) 839- 851.

[3] R. L. Marsden, L. J. McGuffin, and D. T. Jones, Rapid protein domain
    assignment from amino acid sequence using predicted secondary
    structure, Protein Science **11** (2002) 2814- 2824.

[4] M. Suyama and O. Ohara, DomCut: prediction of inter- domain linker
    regions in amino acid sequences, Bioinformatics **19** (2003) 673- 674.

[5] R. A. George and J. Heringa, Protein domain identification and

improved sequence similarity searching using PSI- BLAST, Proteins: Structure, Function, and Bioinformatics **48** (2002) 672- 681.

[6] J. Sim, S.- Y. Kim, and J. Lee, PPRODO: prediction of protein domain boundaries using neural network, Proteins: Structure, Function, and Bioinformatics **59** (2005) 627- 632.