

Domain definitions

"The definition of protein domains varies widely across the discipline of biology. Domains are defined simultaneously as:

- (1) regions that display a significant level of sequence homology;
- (2) a minimal part of the gene that is capable of performing a function;
- (3) a region of the protein with an experimentally assigned function;
- (4) parts of structures that have significant structural similarity; and
- (5) compact spatially distinct units of protein structure."

[Veretnik S, Bourne PE, Alexandrov NN, Shindyalov IN. Toward consistent assignment of structural domains in proteins. J Mol Biol. 2004 Jun 4;339(3):647-678.]

STRUDL (STRUctural Domain Limits)

[Wernisch L, Hunting M, Wodak SJ. Identification of structural domains in proteins by a graph heuristic. Proteins. 1999 May 15;35(3):338-352.]

Algorithm designed to identify domains with any number of non-contiguous chain segments.

Uses the Kernighan-Lin graph heuristic to partition the protein into residue sets which display minimum interactions between them.

Interactions are deduced from contact areas between atoms in the weighted Voronoi diagram.

The radius of the "accessible sphere" around each atom is the van der Waals radius of the atom increased by 1.4 *angstrom*.

DOMAK (DOmain MAKER)

[Siddiqui AS, Barton GJ. Continuous and discontinuous domains: an algorithm for the automatic generation of reliable protein domain definitions. Protein Sci. 1995 May;4(5):872-884.]

Based on the principle that the residues comprising a domain make more contacts between themselves (internal contacts) than they do with the rest of the protein (external contacts).

Two residues make contact if a heavy atom of one is within 5 *angstrom* of a heavy atom of the other.

$$\frac{int_A}{ext_{AB}} \times \frac{int_B}{ext_{AB}}$$

Can deal with domains consisting of two segments.

Conflicting domain assignments

"The major factors responsible for conflicting domain assignments between methods, both experts and automatic, are:

- (1) the definition of very small domains;
- (2) splitting secondary structures between domains;
- (3) the size and number of discontinuous domains;
- (4) closely packed or convoluted domain-domain interfaces;
- (5) structures with large and complex architectures; and
- (6) the level of significance placed upon structural, functional and evolutionary concepts in considering structural domain definitions."

[Veretnik S, Bourne PE, Alexandrov NN, Shindyalov IN. Toward consistent assignment of structural domains in proteins. J Mol Biol. 2004 Jun 4;339(3):647-678.]