3D-PP: A tool for discovering conserved 3D protein patterns

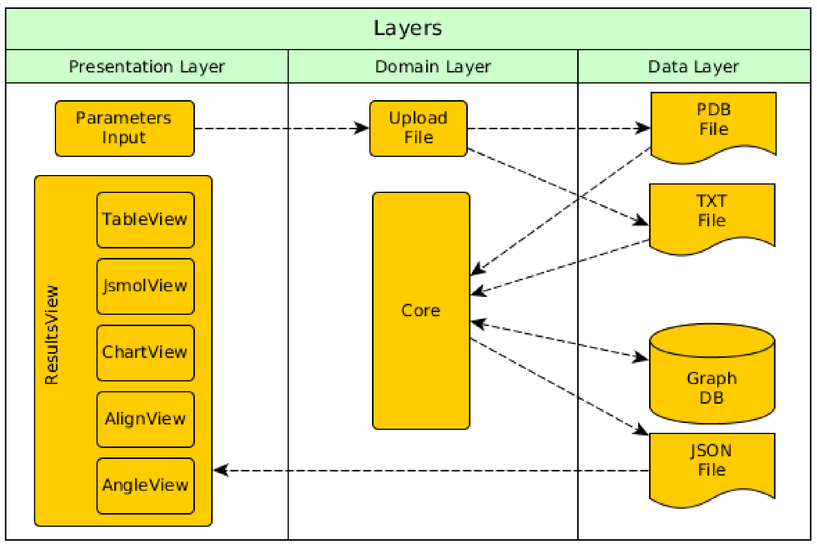
Valdés-Jiménez A[1]
Larriba-Pey JL[2]
Reyes-Paredes M[3,4]
Núñez-Vivanco G[1]

[1]CBSM, Universidad de Talca, Talca, Chile.
[2]DAMA-UPC, Data Management Universitat Politecnica de Catalunya. Spain.
[3]School of Medicine, Faculty of Medical Sciences, USACH, Chile
[4]Facultad Ciencias de la Salud, Universidad Autónoma, Chile
avaldes@utalca.cl

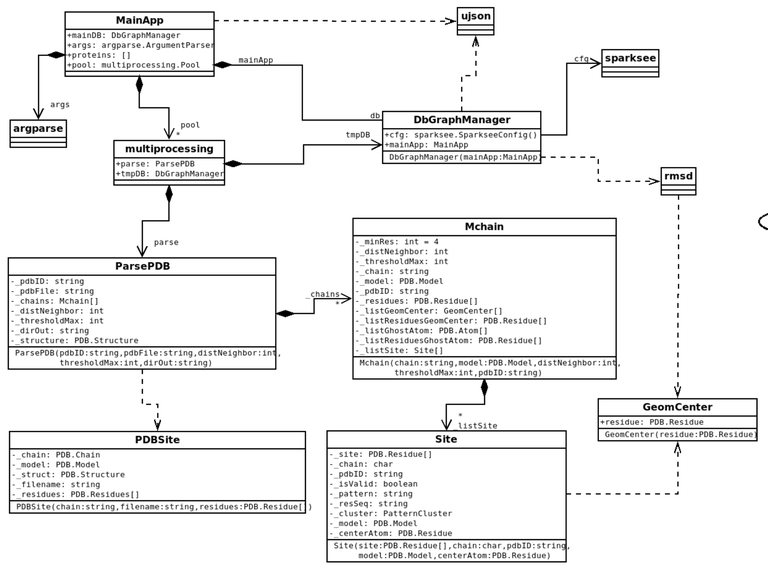
MOTIVATION

Several tools have been developed for the identification of similar 3D patterns, however, usually demand a known query or only consider the observed data (e.g. orthosteric binding sites in PDB, annotated motif, known ligands, etc.). Nevertheless, some approaches shows that 3D amino acids conservation is a enough prove for consider these residues as part of an active site or a binding site of a protein structure, even when no prior knowledge of functional residues are available. Thus, considering all unknown or unobserved 3D patterns (e.g. allosteric binding sites), for the discovery, search and characterization of putative common binding sites between a set of protein structures, cold be more informative than explore only known sites. Here, we present 3D-PP, a new free access web server to discover all conserved 3D amino acid patterns among a set of protein structures including those coming from both, X-ray crystallographic experiments and *in silico* comparative modelling. The preprocessing modules of 3D-PP were developed in Python and all data generated are processed and organized automatically in a scalable high-performance graph database.

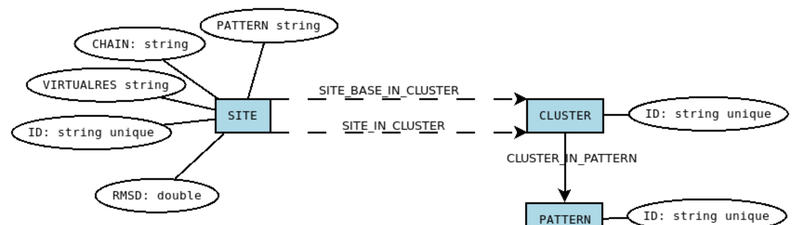
ARCHITECTURE



CLASS DIAGRAM



GRAPH DB SCHEMA



EXAMPLE

PROSITE Documentation PDOC50103
Zinc Finger C3H1-type profile

The distance to create Grid of Virtual Coordinates (GvC): 0.8Å
The maximum radius threshold (in Angstrom) for sites search, from the Virtual Coordinates (VC): 3Å
The maximum difference (in Angstrom) between the RMSDs of the sites (for clustering): 4.5Å
Proteins successfully processed: 21
List: 1m9o 1rgo 2cqe 2d9m 2d9n 2e5s 2fc6 2rhk 2rpp 3d2n 3d2q 3d2s 3tp2 3u1l 3u1m 3u9g 4c3b 4c3d 4c3e 4cyk 4ii1

Search Patterns (RegExp):

Show 10 entries

Pattern	In Prot	Not In	% Coverage	# Clusters	# Total sites
3C1H	21	0	100.0	1	83
2C1F1H	21	0	100.0	1	83
2C1H1Y	15	6	71.4	1	52
2C1G1H	14	7	66.7	1	35
1C1F1H1L	12	9	57.1	1	54
1K1L1T1V	10	11	47.6	1	83
1D1K1L1V	10	11	47.6	1	72
1E1K2L	10	11	47.6	1	49
1A1L2V	10	11	47.6	1	49
1E1L1R1T	10	11	47.6	1	24

82 of the detected sites have been previously confirmed by crystallographic studies where is posible to observe the respective Zinc Ion in coordination with 3CYS and 1HIS aminoacids.

Clusters in Pattern **3C1H**

Show 10 entries Search:

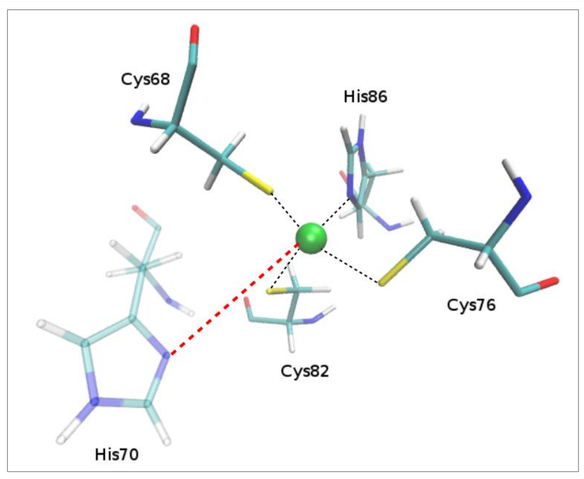
Cluster	# Sites	In Prot	% Coverage
3C1H-1	83	21	100

Sites in Cluster **3C1H-1**

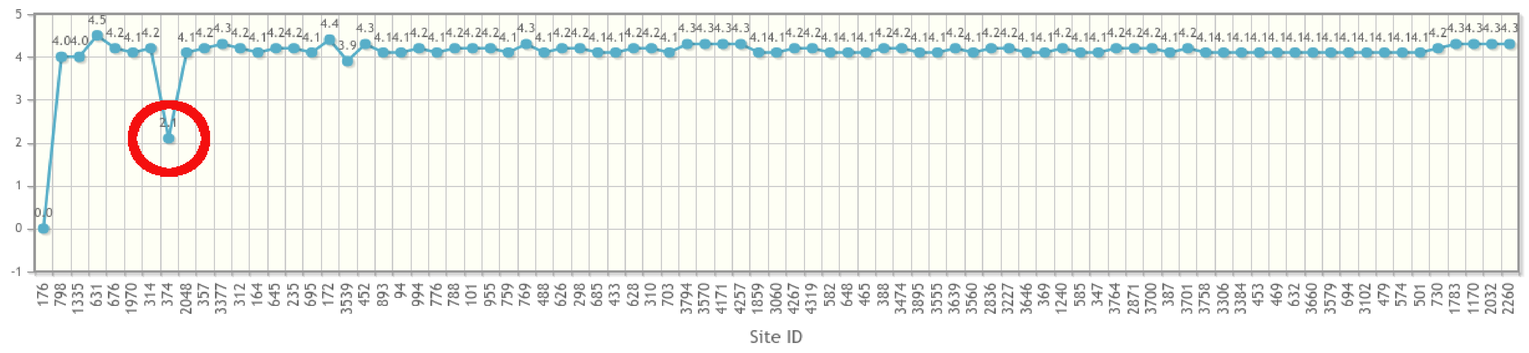
Show 10 entries Search:

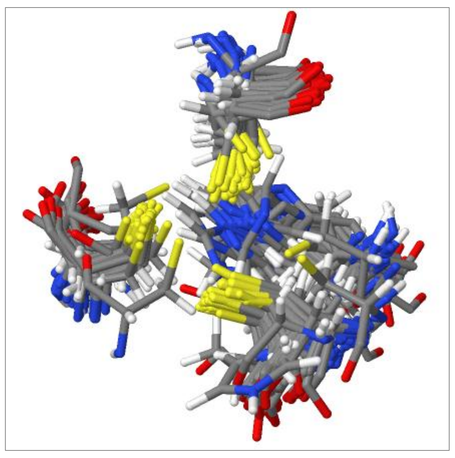
Site ID	Site	Chain	Protein	RMSD	Base
176	CYS15:CYS24:CYS30:HIS34	A	1m9o	0	yes
374	CYS68:CYS76:CYS82:HIS70	A	2d9n	2.1	no
3539	CYS60:CYS68:CYS73:HIS77	A	2rpp	3.9	no
798	CYS159:CYS168:CYS174:HIS178	A	1rgo	4	no
1335	CYS197:CYS206:CYS212:HIS216	A	1rgo	4	no
1970	CYS907:CYS915:CYS921:HIS925	A	2d9m	4.1	no
2048	CYS105:CYS110:CYS96:HIS114	A	2d9n	4.1	no
164	CYS105:CYS110:CYS96:HIS114	C	2rhk	4.1	no
695	CYS105:CYS110:CYS96:HIS114	D	2rhk	4.1	no
893	CYS53:CYS61:CYS66:HIS70	A	3d2n	4.1	no

Showing 1 to 10 of 83 entries



RMSD Stats for Cluster **3C1H-1**





ACKNOWLEDGEMENTS