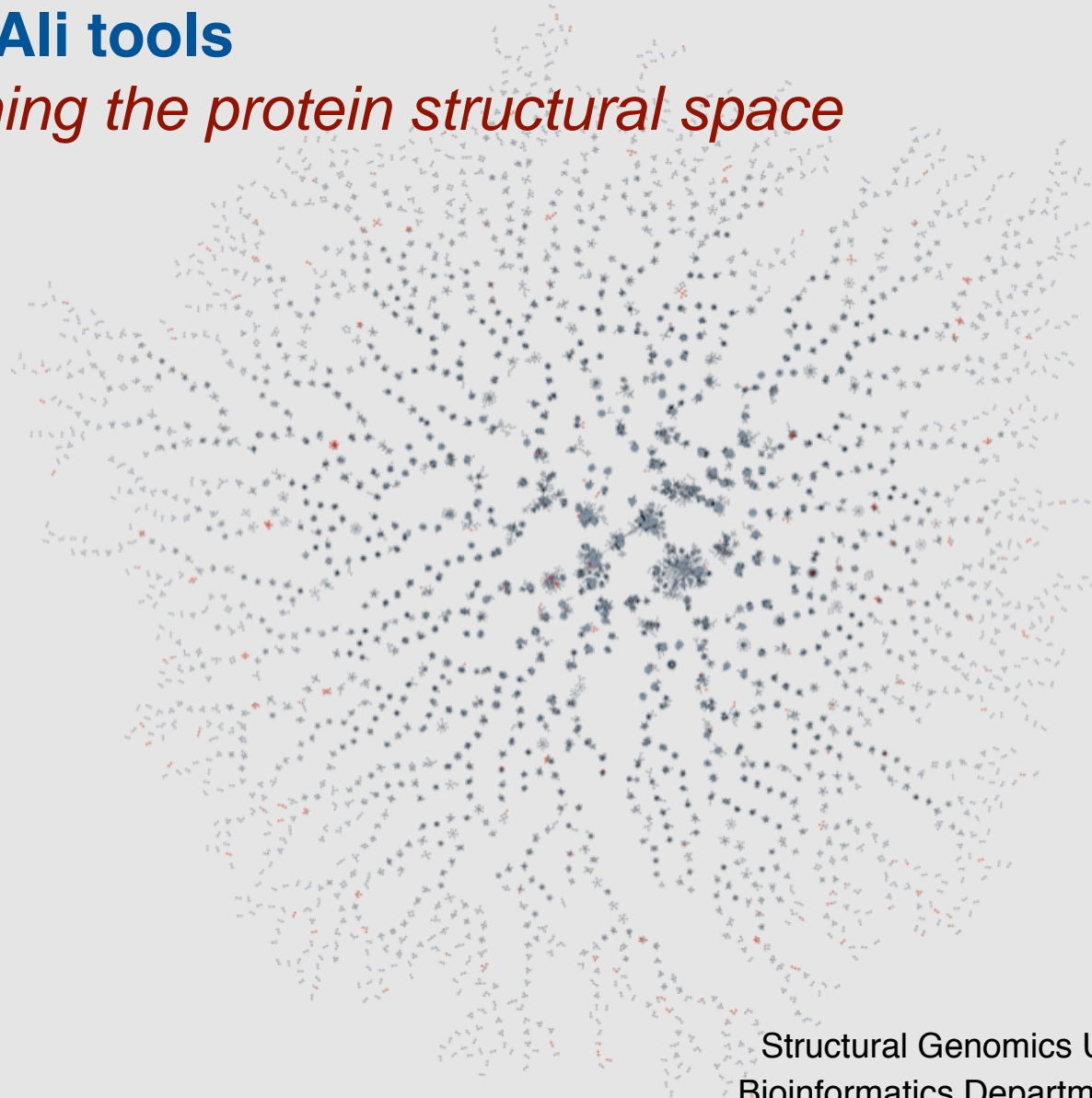


DBAli tools

Mining the protein structural space



Marc A. Marti-Renom

<http://bioinfo.cipf.es/squ/>

Structural Genomics Unit
Bioinformatics Department

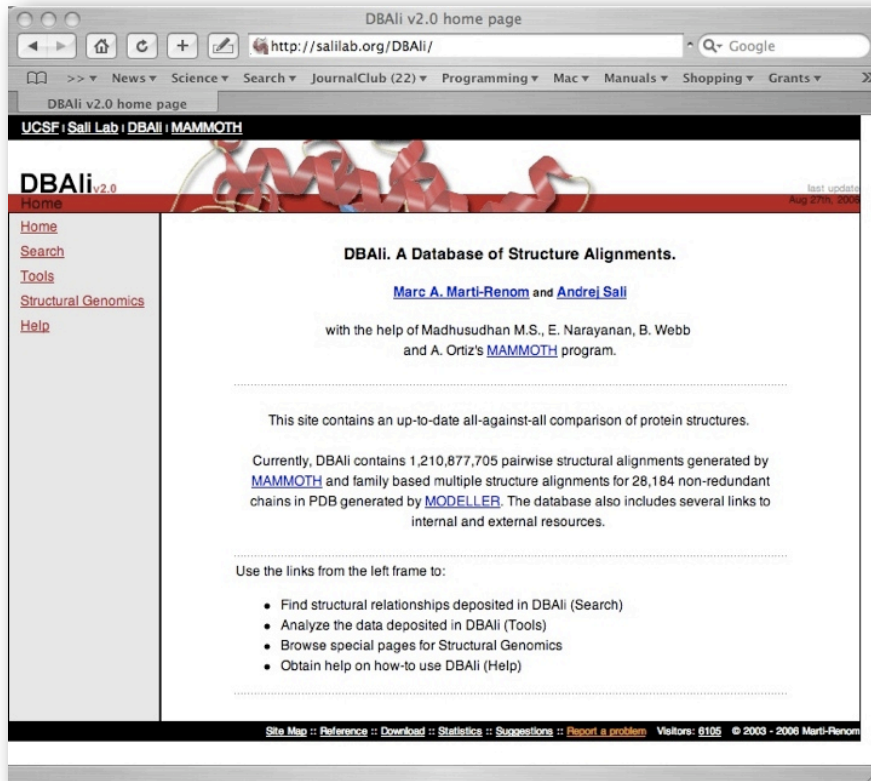
Prince Felipe Research Center (CIPF), Valencia, Spain



DBAli_{v2.0} database

<http://bioinfo.cipf.es/squ/services/DBAli/>

<http://www.salilab.org/DBAli/>



- ✓ Fully-automatic
- ✓ Data is kept up-to-date with PDB releases
- ✓ Tools for “on the fly” classification of families.
- ✓ Easy to navigate
- ✓ Provides tools for structure analysis

Does not provide a stable classification similar to that of CATH or SCOP

Pairwise structure alignments	
Last update:	March 22nd, 2007
Number of chains:	89,094
Number of structure-structure comparisons:*	1,460,445,131
Multiple structure alignments	
Last update:	January 23rd, 2007
Number of representative chains:	30,900
Number of families:	11,615

Uses MAMMOTH for similarity detection

- ✓ **VERY FAST!!!**
- ✓ **Good scoring system with significance**

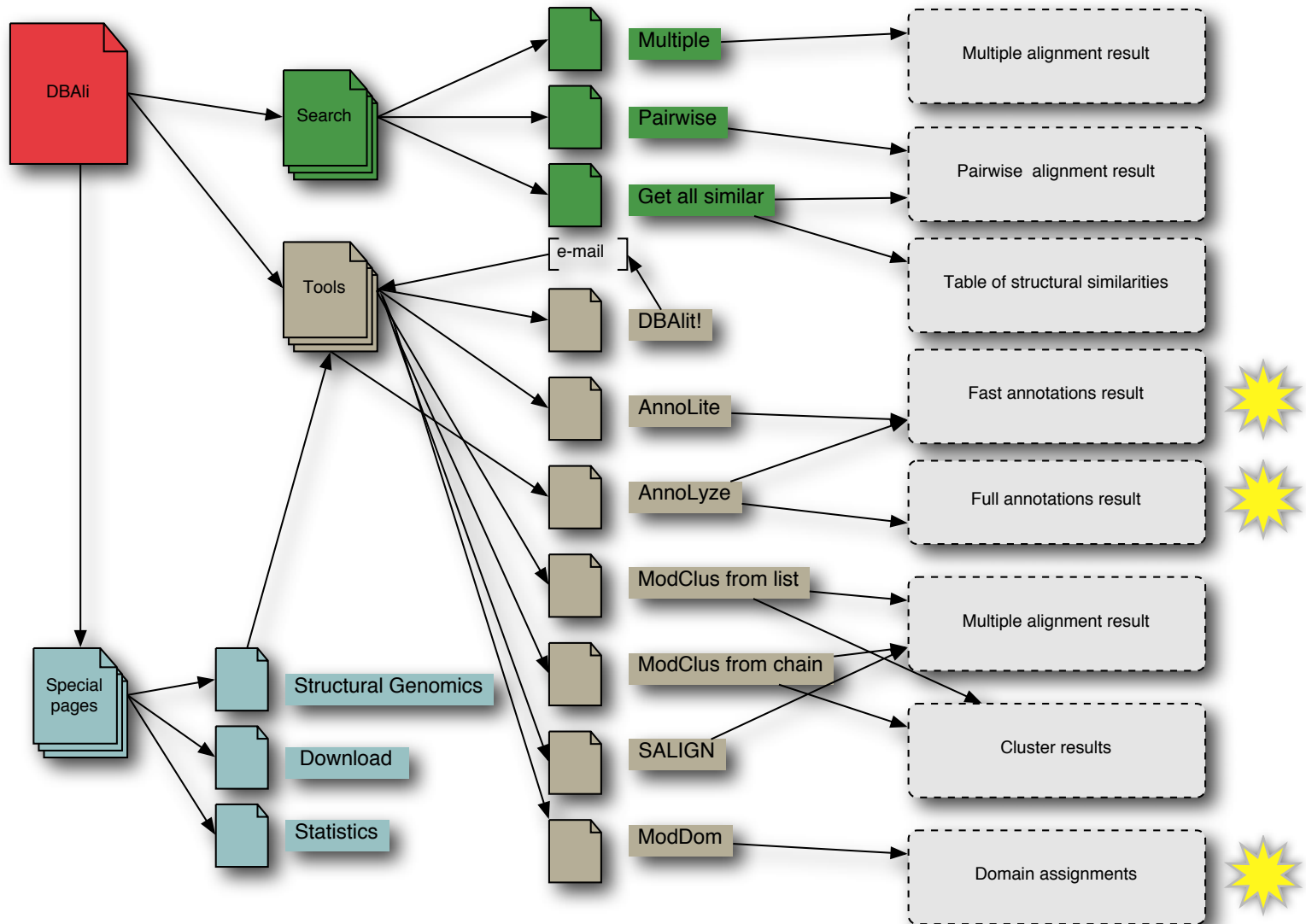
Ortiz AR, (2002) Protein Sci. 11 pp2606

Marti-Renom et al. 2001. Bioinformatics. 17, 746

DBAli_{v2.0} database

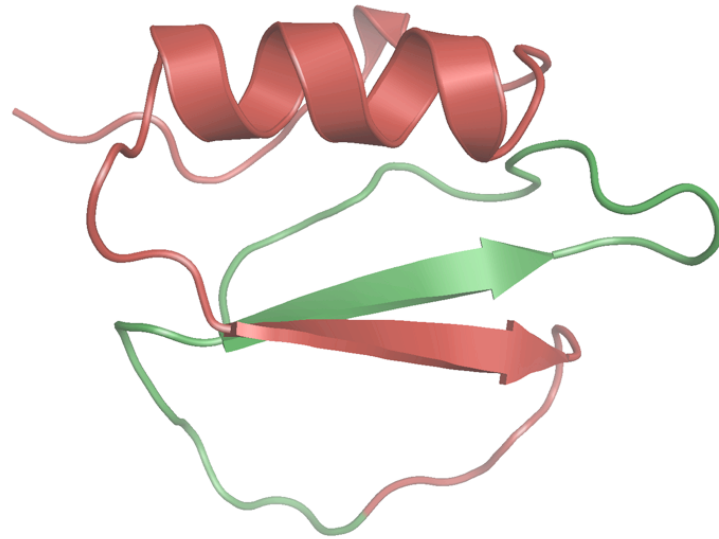
<http://bioinfo.cipf.es/squ/services/DBAli/>

<http://www.salilab.org/DBAli/>

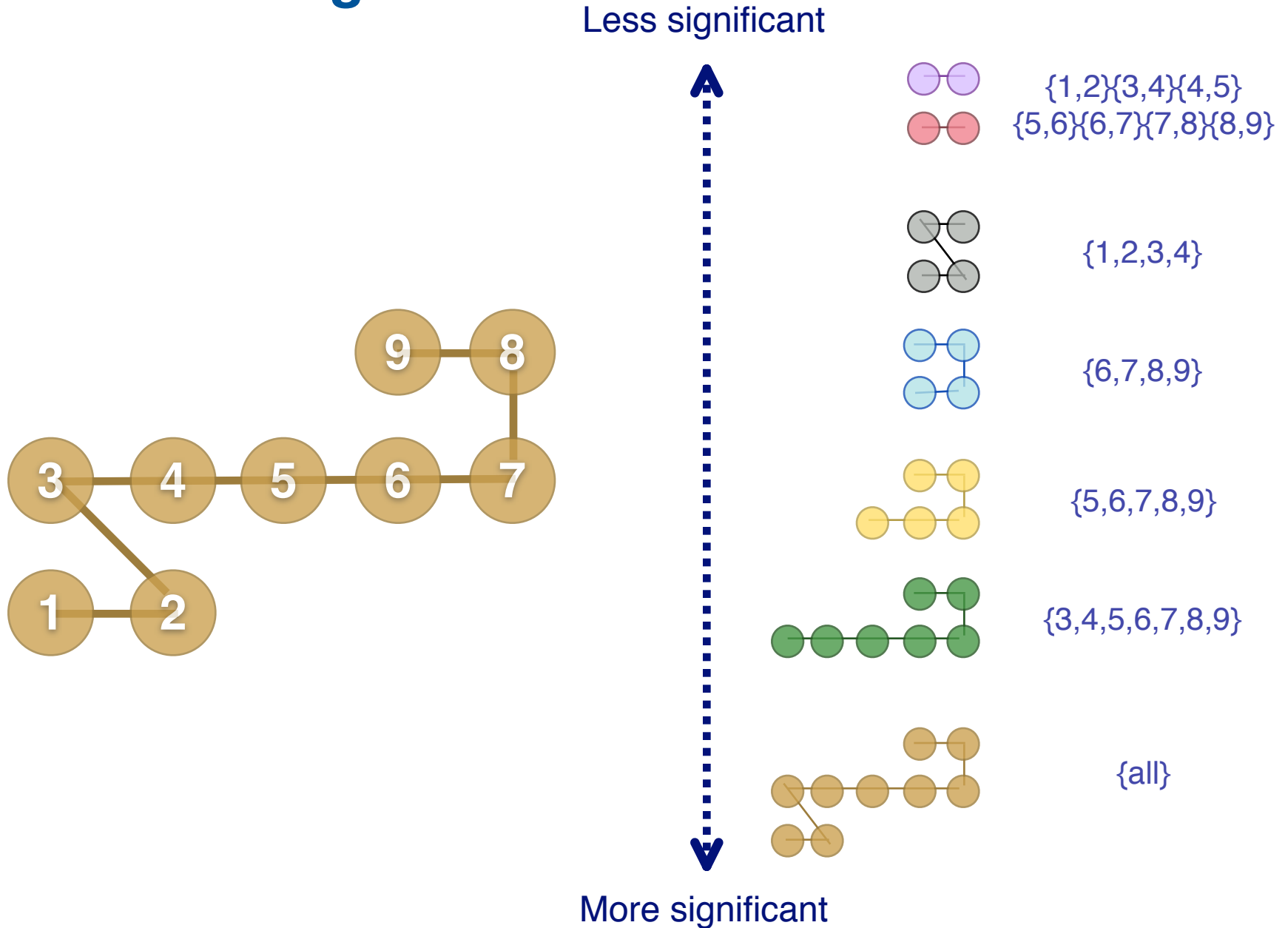


Protein domains/fragments from structure

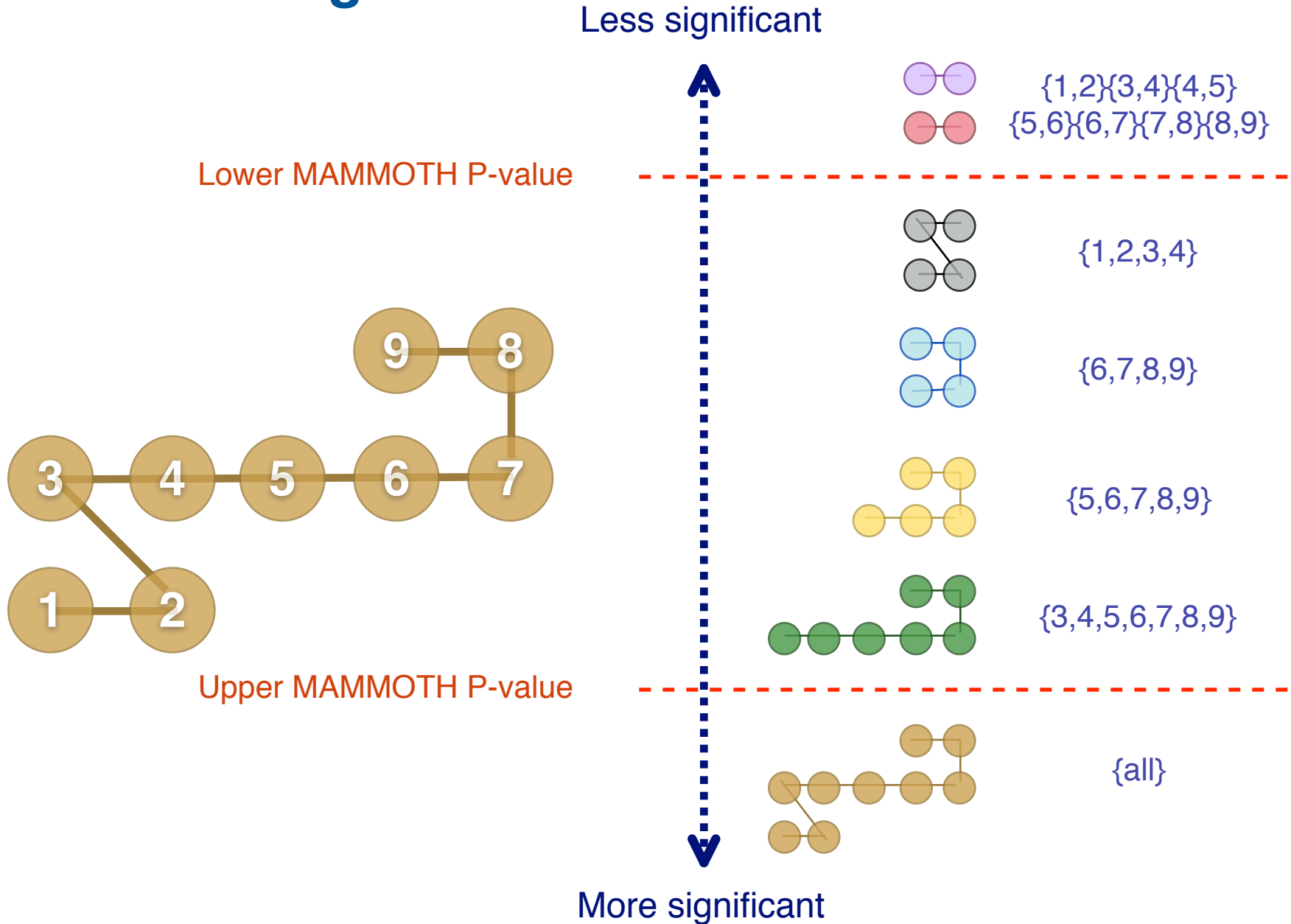
ModDom



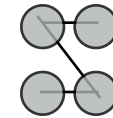
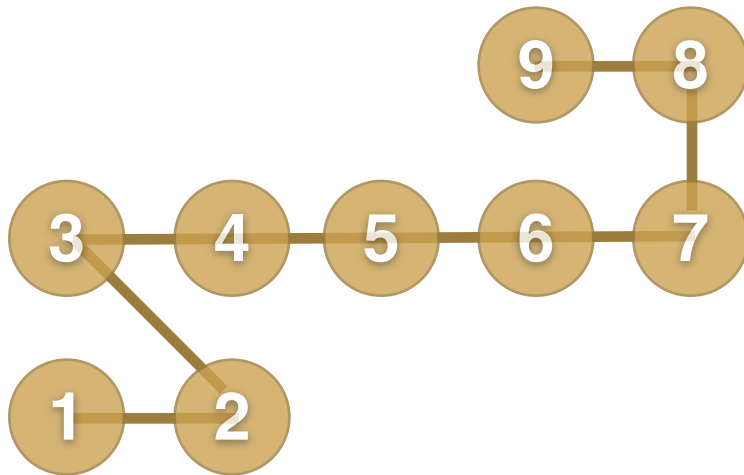
ModDom algorithm



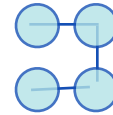
ModDom algorithm



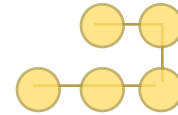
ModDom algorithm



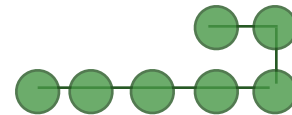
$\{1,2,3,4\}$



$\{6,7,8,9\}$

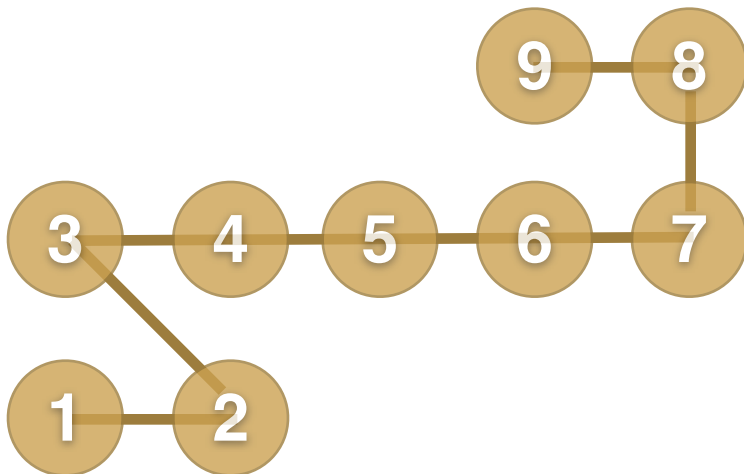
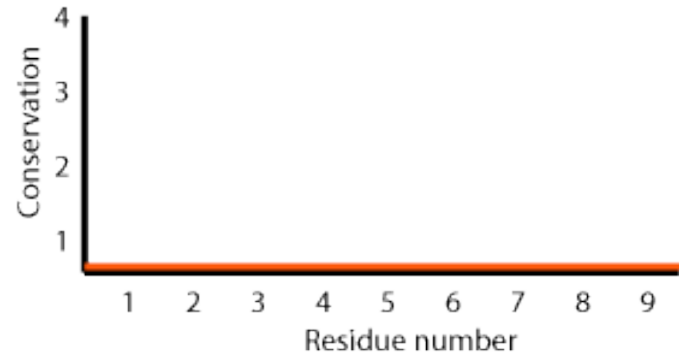
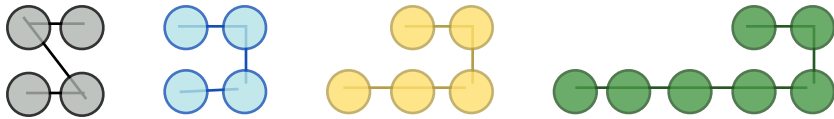


$\{5,6,7,8,9\}$



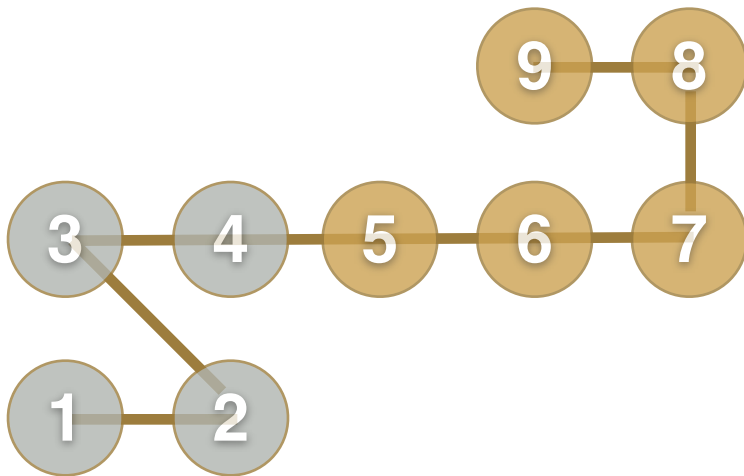
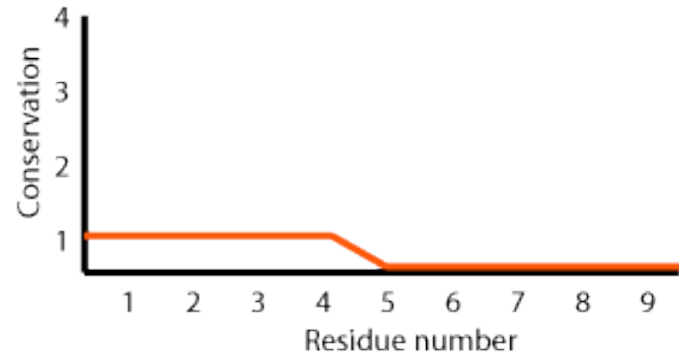
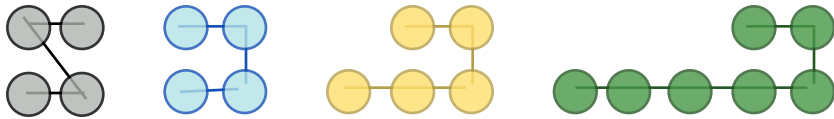
$\{3,4,5,6,7,8,9\}$

ModDom algorithm



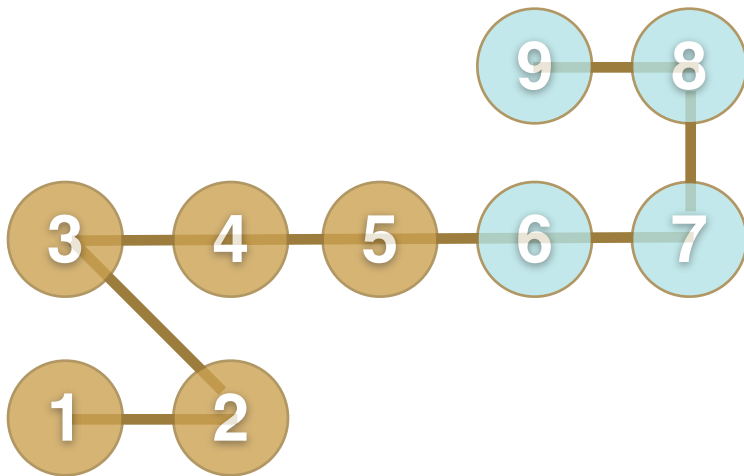
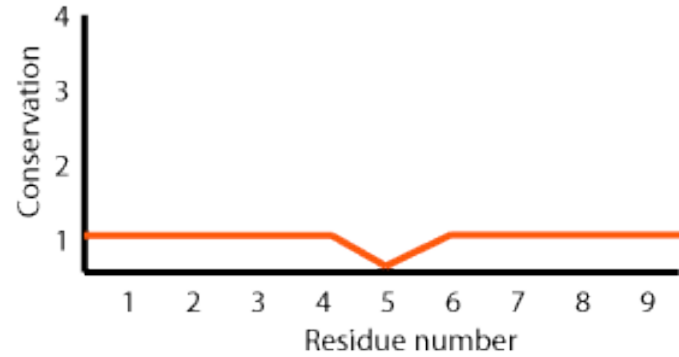
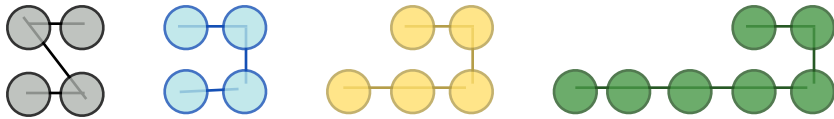
#	1	2	3	4	5	6	7	8	9
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0

ModDom algorithm



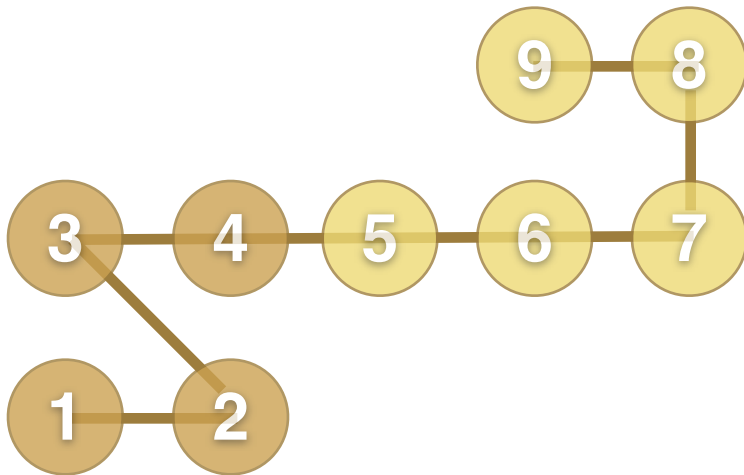
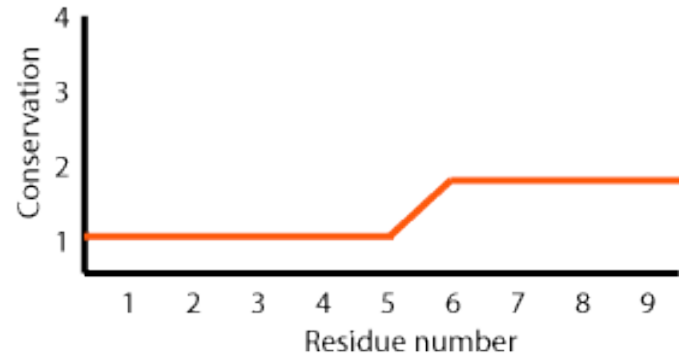
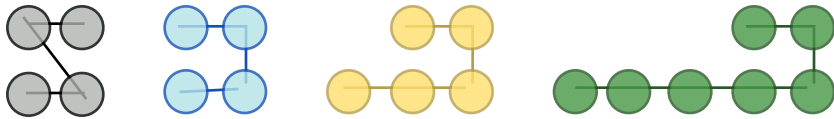
#	1	2	3	4	5	6	7	8	9
1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0
3	1	1	1	1	0	0	0	0	0
4	1	1	1	1	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0

ModDom algorithm



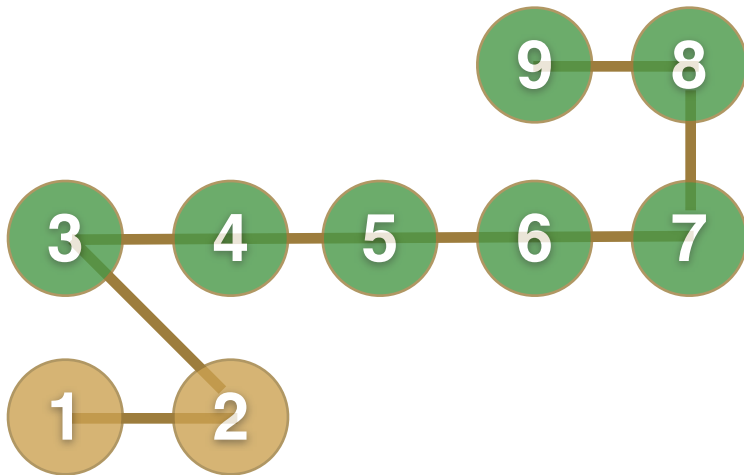
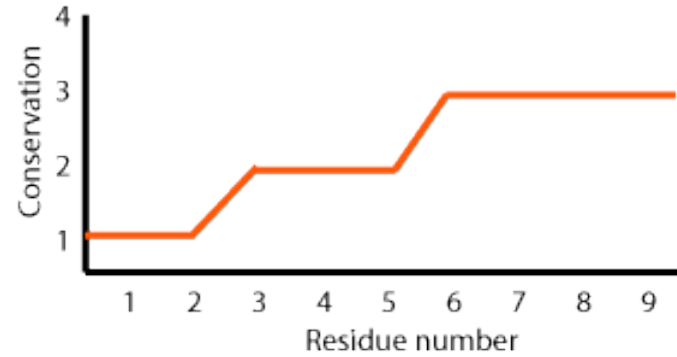
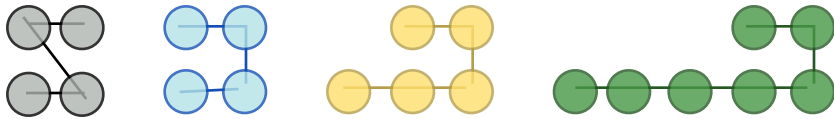
#	1	2	3	4	5	6	7	8	9
1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0
3	1	1	1	1	0	0	0	0	0
4	1	1	1	1	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	1	1	1
7	0	0	0	0	0	1	1	1	1
8	0	0	0	0	0	1	1	1	1
9	0	0	0	0	0	1	1	1	1

ModDom algorithm



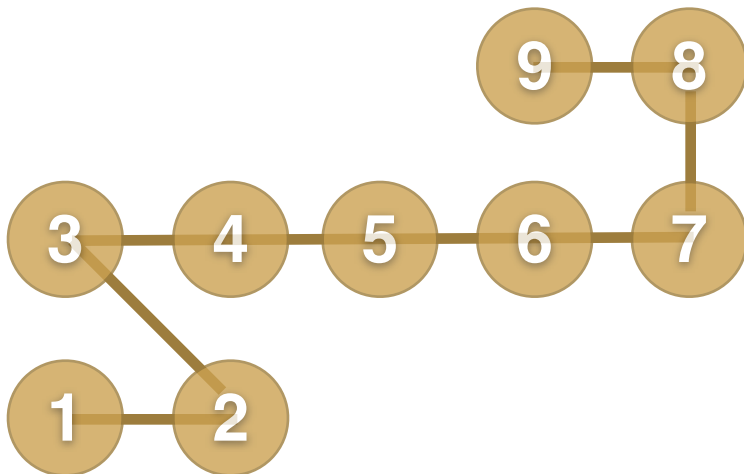
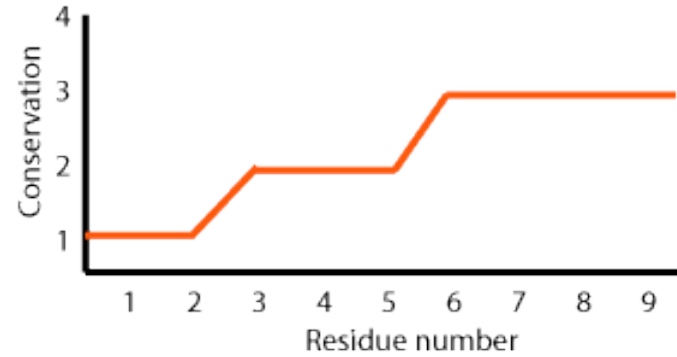
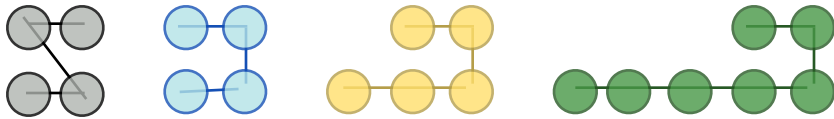
#	1	2	3	4	5	6	7	8	9
1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0
3	1	1	1	1	0	0	0	0	0
4	1	1	1	1	0	0	0	0	0
5	0	0	0	0	1	1	1	1	1
6	0	0	0	0	1	2	2	2	2
7	0	0	0	0	1	2	2	2	2
8	0	0	0	0	1	2	2	2	2
9	0	0	0	0	1	2	2	2	2

ModDom algorithm



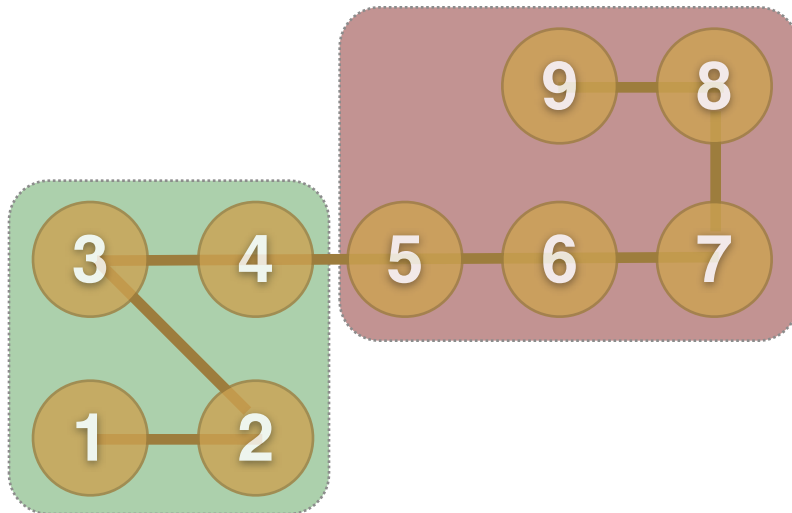
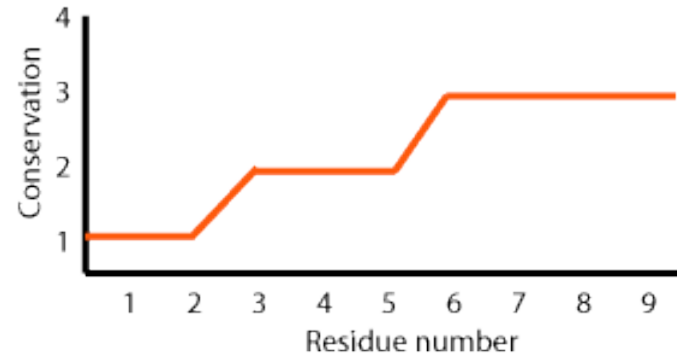
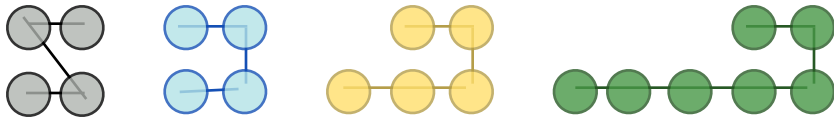
#	1	2	3	4	5	6	7	8	9
1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0
3	1	1	2	2	1	1	1	1	1
4	1	1	2	2	1	1	1	1	1
5	0	0	1	1	2	2	2	2	2
6	0	0	1	1	2	3	3	3	3
7	0	0	1	1	2	3	3	3	3
8	0	0	1	1	2	3	3	3	3
9	0	0	1	1	2	3	3	3	3

ModDom algorithm



#	1	2	3	4	5	6	7	8	9
1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0
3	1	1	2	2	1	1	1	1	1
4	1	1	2	2	1	1	1	1	1
5	0	0	1	1	2	2	2	2	2
6	0	0	1	1	2	3	3	3	3
7	0	0	1	1	2	3	3	3	3
8	0	0	1	1	2	3	3	3	3
9	0	0	1	1	2	3	3	3	3

ModDom algorithm



Threshold #3 MCL Cluster level (-I)

Stijn van Dongen (<http://micans.org/mcl/>)

#	1	2	3	4	5	6	7	8	9
1	1	1	1	1	0	0	0	0	0
2	1	1	1	1	0	0	0	0	0
3	1	1	2	2	1	1	1	1	1
4	1	1	2	2	1	1	1	1	1
5	0	0	1	1	2	2	2	2	2
6	0	0	1	1	2	3	3	3	3
7	0	0	1	1	2	3	3	3	3
8	0	0	1	1	2	3	3	3	3
9	0	0	1	1	2	3	3	3	3

Domains as recurrent fragments

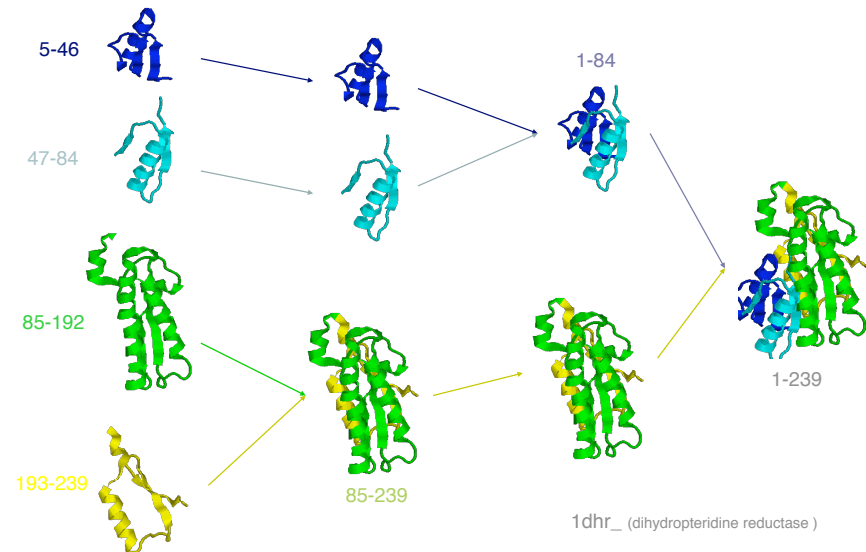
2163 chains from Islam et al. 1995 → 569 Non-redundant
<2Å && <30aa diff.

Divide randomly into two sets
Remove of incomplete or obsolete entries.

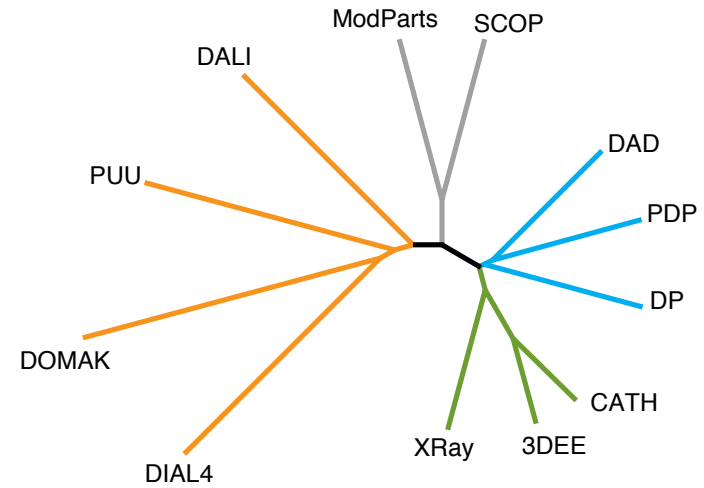
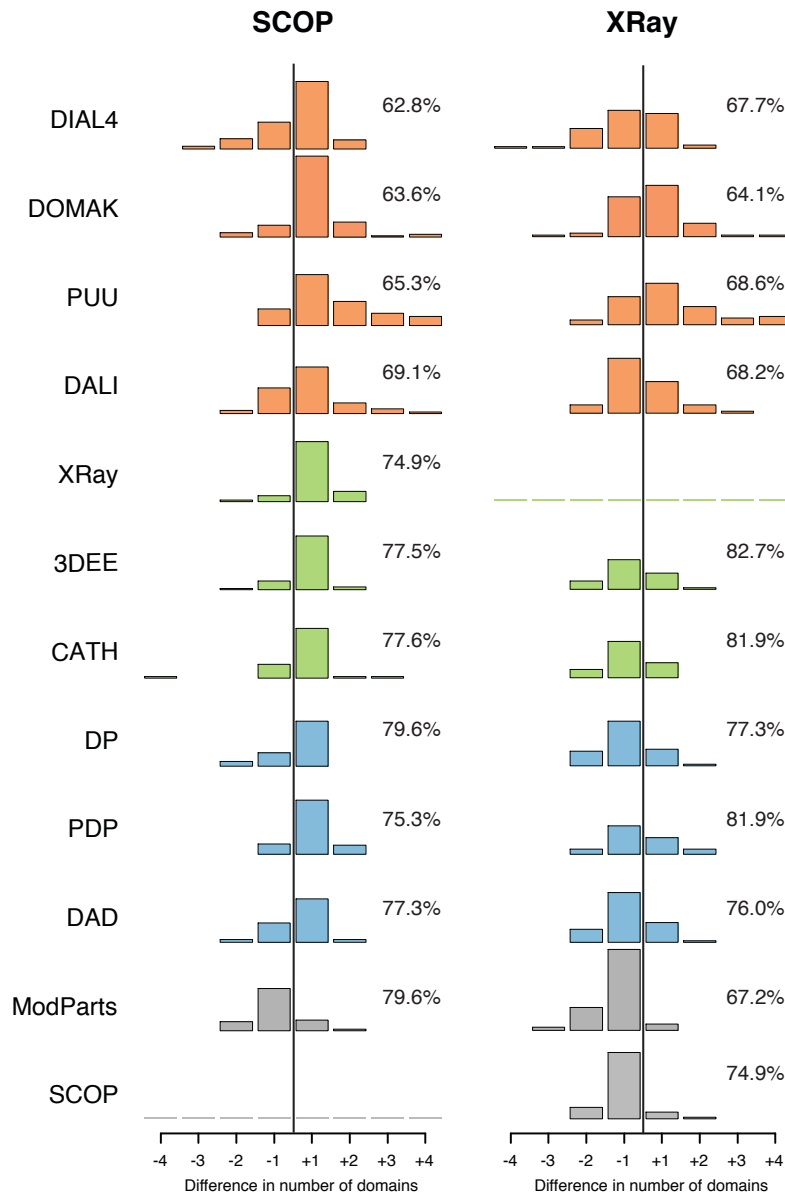
Training set → 242 chains

Testing set → 234 chains

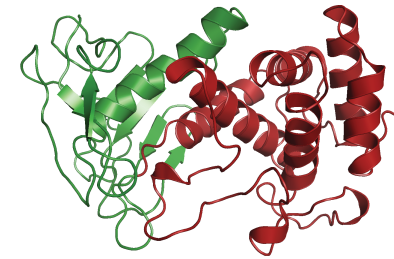
$$R = \text{Volume/ASA}$$



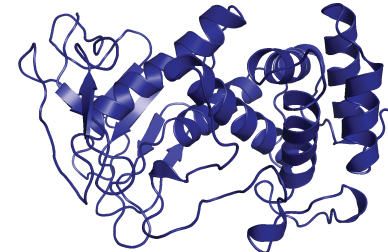
Domains as recurrent fragments



All other methods

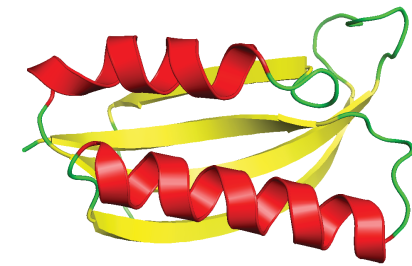
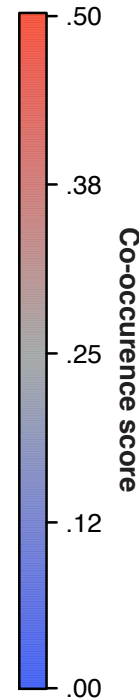
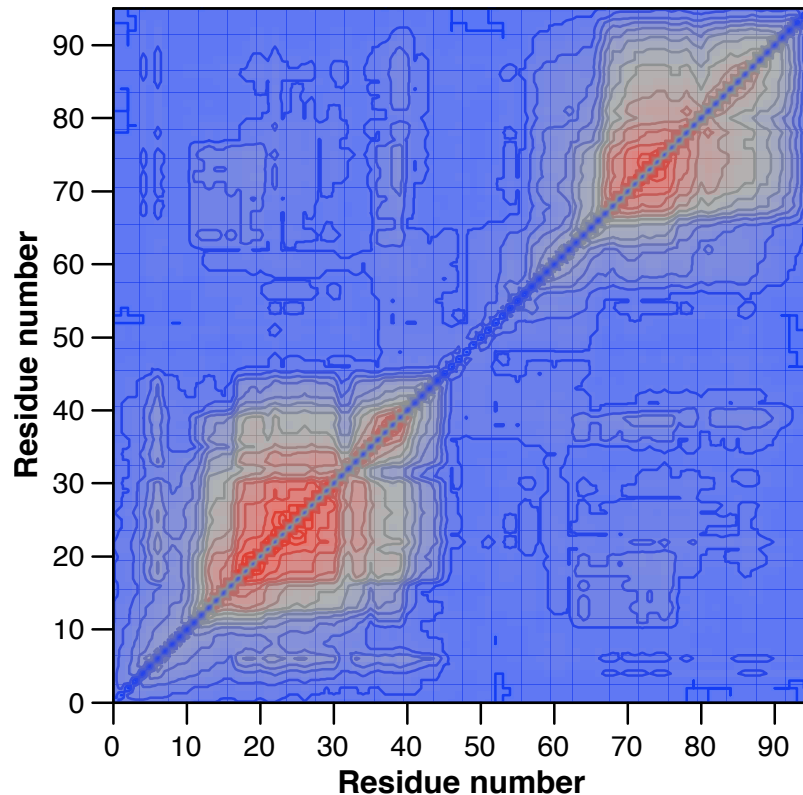


SCOP and ModParts

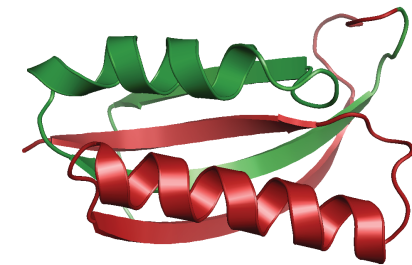


Repetitions as recurrent fragments

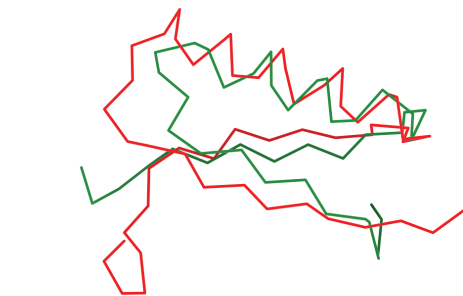
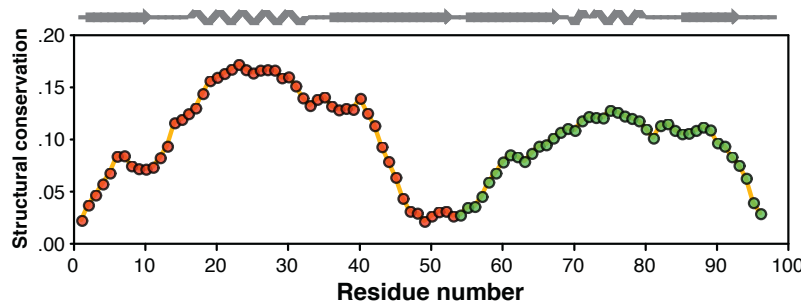
Ribosomal protein S6 (1ris) $\alpha+\beta \rightarrow$ *Ferredoxin Like domain*



1ris: _



unit 1 1-53
unit 2 54-97



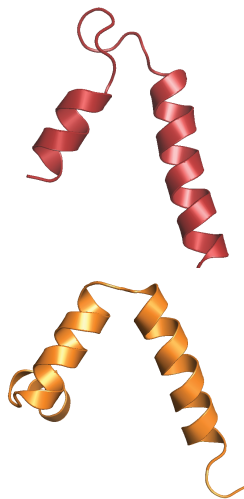
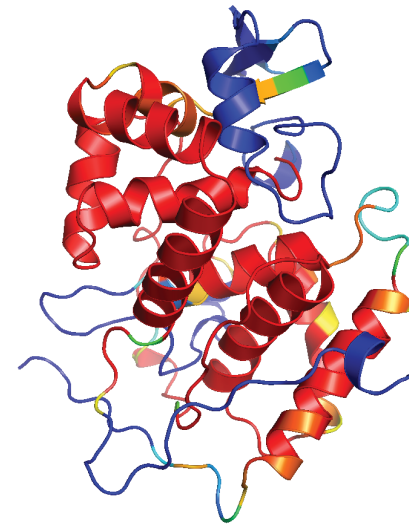
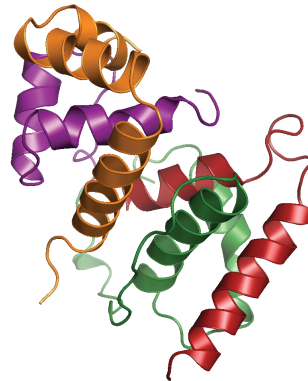
superimposed reversed unit 1 and unit 2
RMSD 3.1Å over 39 C α carbons

Repetitions as recurrent fragments

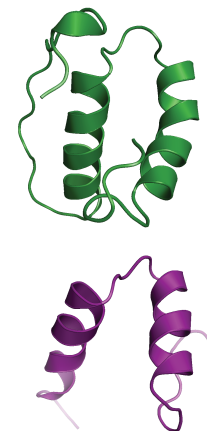
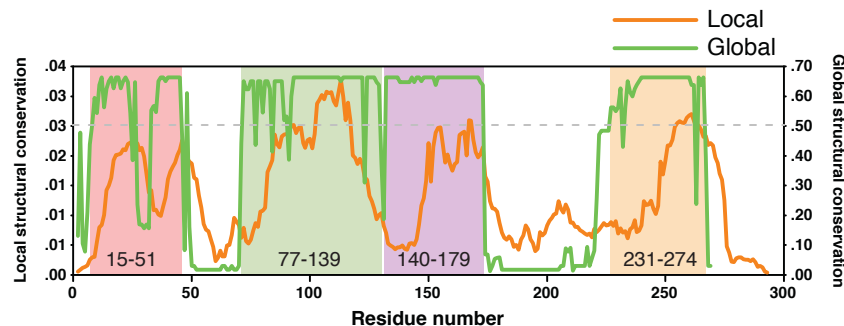
Cytochrome C Peroxidase (2cyp) *all- α* \rightarrow CCP-like domain



	limits	length
unit 1	2-62	61
unit 2	63-140	78
unit 3	141-197	57
unit 4	198-232	61
unit 5	233-293	35



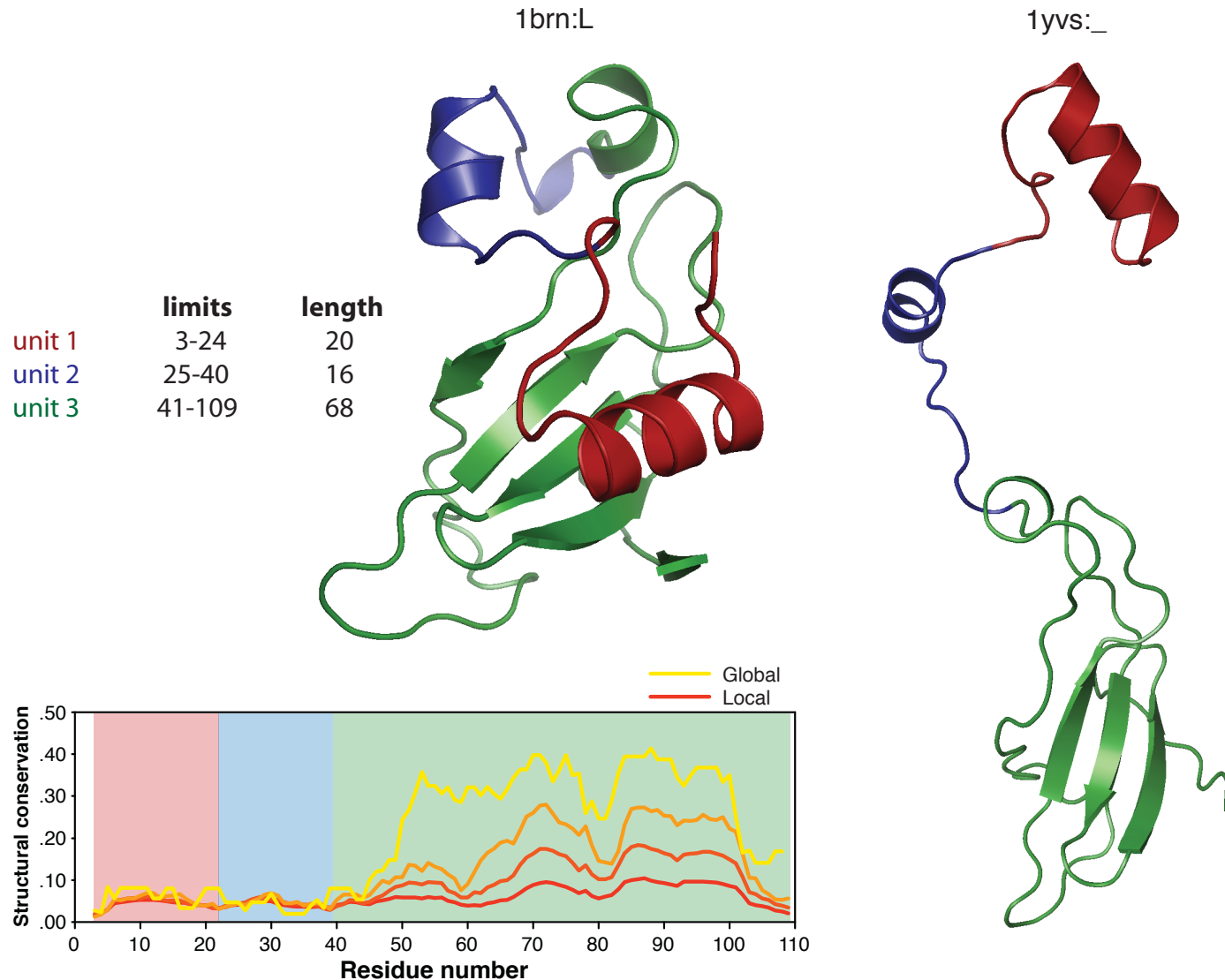
unit 1 and unit 5
3.5Å C α RMSD over 25 residues



unit 2 and unit 3
2.9Å C α RMSD over 30 residues

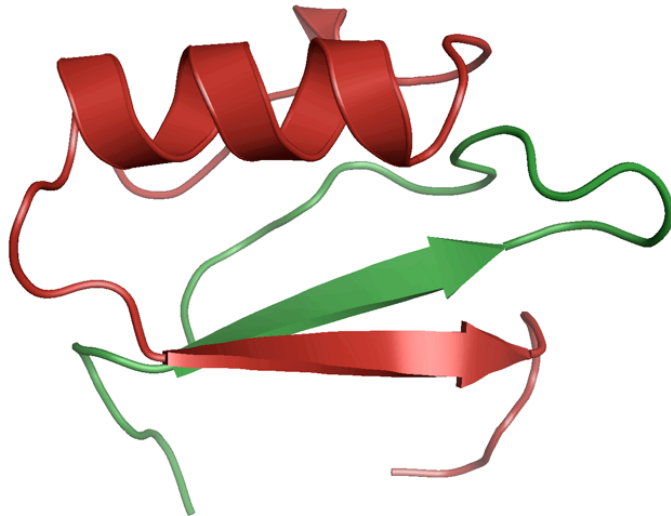
Swapping of recurrent fragments

Barnase Domain-Swapping

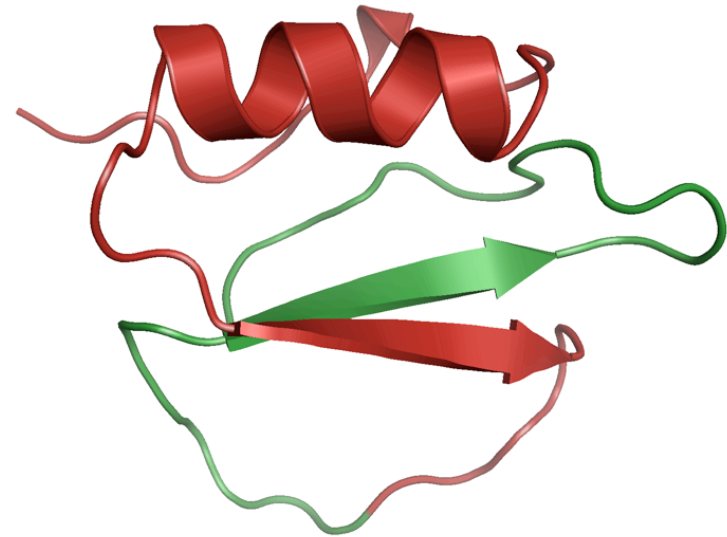


Co-folding of recurrent fragments

Chymotrypsin inhibitor 2



1-37 | 38-64

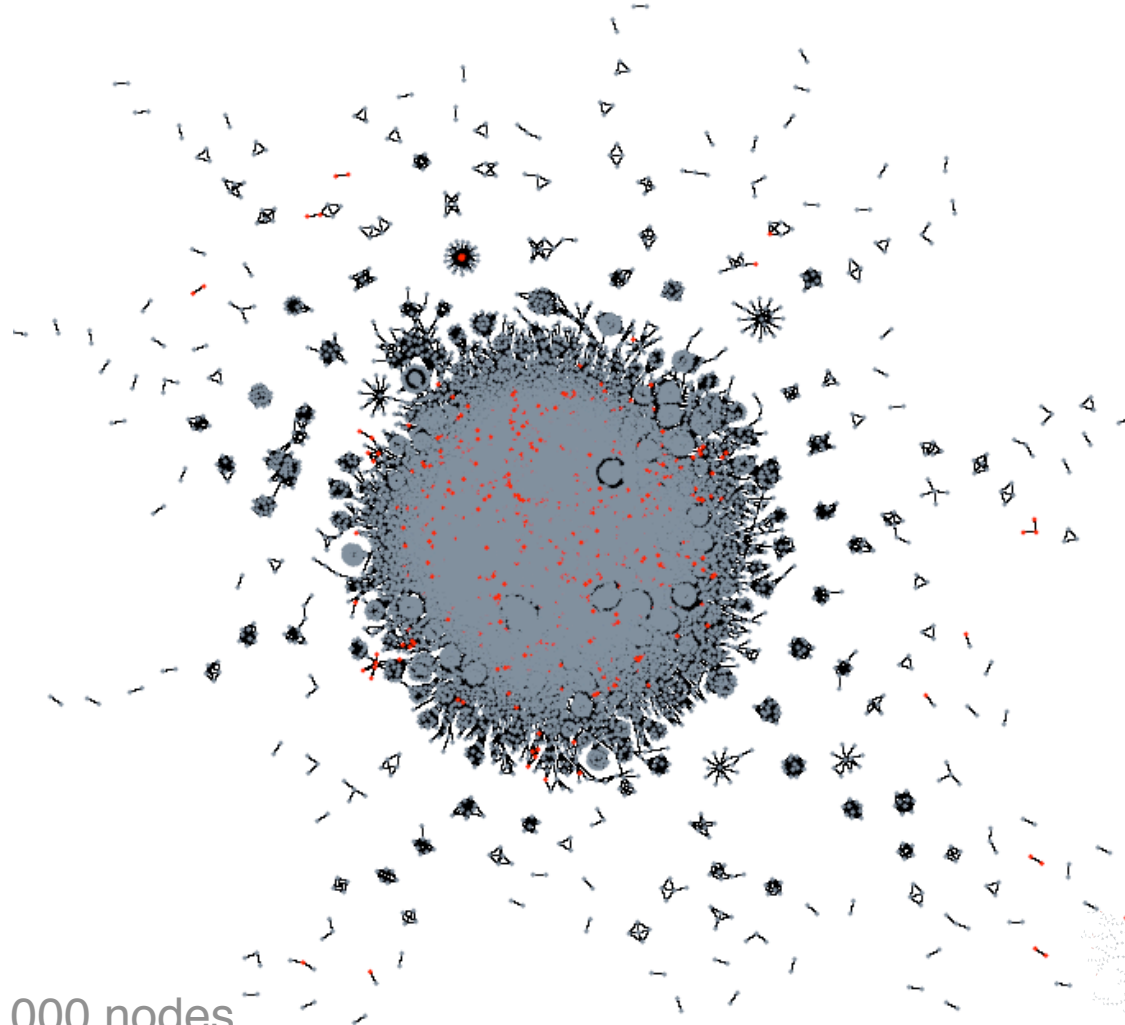


1-40 | 41-64

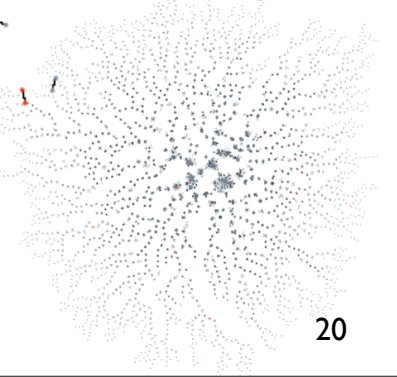
- Neira JL, Davis B, Ladurner AG, Buckle AM, Gay GP, Fersht AR. 1996. *Fold Des* 1:189-208.
- Ladurner AG, Itzhaki LS, de Prat GG, Fersht AR. 1997. *J Mol Biol* 273:317-329.

Sequence space .vs. Structure space

Structure map @ >20% sequence identity



Structure map @ >30%

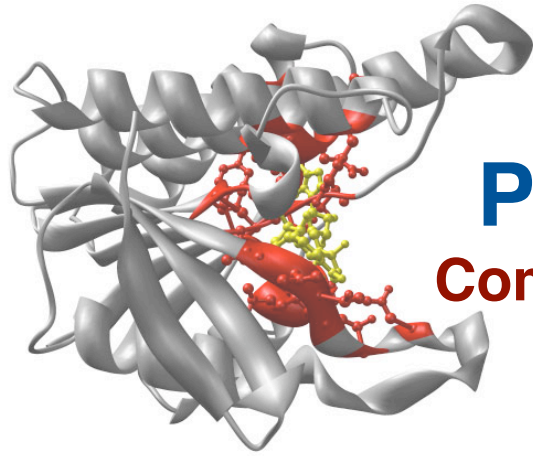


~22,000 nodes
186 clusters

Larger cluster contains 20,231 chains

Sequence space .vs. Structure space





Protein function from structure

Comparative annotation. AnnoLite and AnnoLyze.

For many protein structures function is *unknown*

	Structural Genomics*	Traditional methods
Annotated**	654	28,342
Not Annotated	506 (43.6%)	6,815 (19,4%)
Total deposited	1,160	35,157

* annotated as STRUCTURAL GENOMICS in the header of the PDB file

**annotated with either CATH, SCOP, Pfam or GO terms in the MSD database
36,317 protein structures, as of August 8th, 2006

For **20%** protein structures function is *unknown*

	Structural Genomics*	Traditional methods
Annotated**	654	28,342
Not Annotated	506 (43.6%)	6,815 (19,4%)
Total deposited	1,160	35,157

* annotated as STRUCTURAL GENOMICS in the header of the PDB file

**annotated with either CATH, SCOP, Pfam or GO terms in the MSD database
36,317 protein structures, as of August 8th, 2006

AnnoLite

	Conf. P-value	Link	Description
CATH:	7.5e-99	2.70.100.10	1,4-Beta-D-Glucan Cellobiohydrolase I, subunit A
SCOP:	0.00	b.29.1.10	Glycosyl hydrolase family 7 catalytic core
PFAM:	0.00	PF00840	Glycosyl hydrolase family 7
InterPro:	1.3e-99	IPR001722	Glycoside hydrolase, family 7
	6.0e-51	IPR008985	Concanavalin A-like lectin/glucanase
	1.0e-42	IPR000254	Cellulose-binding region, fungal
EC Number:	1.2e-44	3.2.1.91	Cellulose 1,4-beta-cellobiosidase.
	6.0e-41	3.2.1.4	Cellulase.
GO Molecular Function:	6.0e-36	0030248	cellulose binding ↕
	8.4e-36	0016162	cellulose 1,4-beta-cellobiosidase activity ↕
	1.0e-35	0004553	hydrolase activity, hydrolyzing O-glycosyl compounds ↕
	1.4e-30	0008810	cellulase activity ↕
	3.1e-20	0016798	hydrolase activity, acting on glycosyl bonds ↕
	1.0e+0	0016787	hydrolase activity ↕
GO Biological Process:	1.1e-63	0030245	cellulose catabolism ↕
	1.2e-54	0000272	polysaccharide catabolism ↕
	3.6e-20	0005975	carbohydrate metabolism ↕
GO Cellular Component:	1.2e-23	0005576	extracellular region ↕

● Information annotated in the MSD database.

● High, ● medium and ● low confidence annotations not annotated in the MSD database.

● High, ● medium and ● low confidence annotations already annotated in the MSD database.

Benchmark set

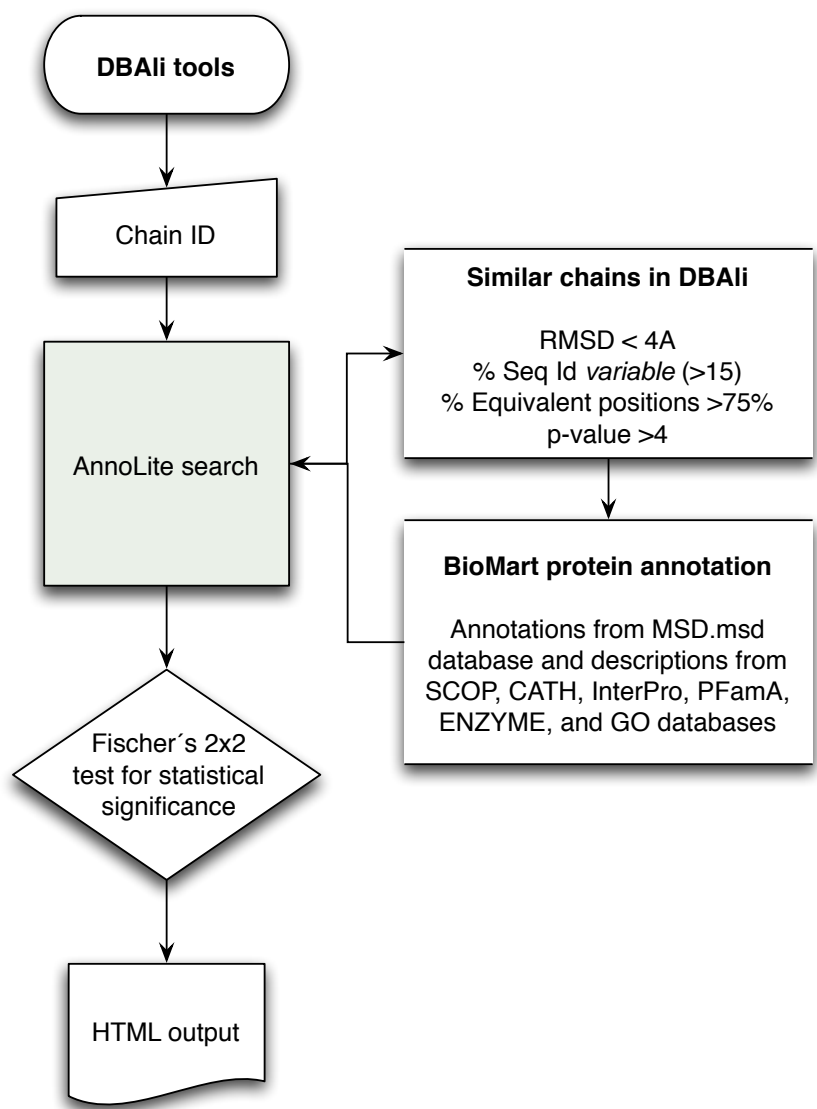
	Number of chains
Initial set*	50,223
FULL annotation**	10,997
Non-redundant set***	1,879

**data from BioMart MSD.3 (release February 2005)*

***annotated with CATH, SCOP, Pfam, EC, InterPro, and GO terms in the MSD database*

****not two chains can be structurally aligned within 2Å, superimposing more than 60% of their C atoms and have a length difference inferior to 30aa*

Method



AnnoLite results for chain [1gpi:A](#) based on [44](#) structural similar chains.

	Conf. P-value	Link	Description
CATH:	● 7.5e-99	2.70.100.10	1,4-Beta-D-Glucan Cellobiohydrolase I, subunit A
SCOP:	● 0.00	b.29.1.10	Glycosyl hydrolase family 7 catalytic core
PFAM:	● 0.00	PF00840	Glycosyl hydrolase family 7
InterPro:	● 1.3e-99	IPR001722	Glycoside hydrolase, family 7
	● 6.0e-51	IPR008985	Concanavalin A-like lectin/glucanase
	● 1.0e-42	IPR000254	Cellulose-binding region, fungal
EC Number:	● 1.2e-44	3.2.1.91	Cellulose 1,4-beta-cellobiosidase.
	● 6.0e-41	3.2.1.4	Cellulase.
GO Molecular Function:	● 6.0e-36	0030248	cellulose binding ↕
	● 8.4e-36	0016162	cellulose 1,4-beta-cellobiosidase activity ↕
	● 1.0e-35	0004553	hydrolase activity, hydrolyzing O-glycosyl compounds ↕
	● 1.4e-30	0008810	cellulase activity ↕
	● 3.1e-20	0016798	hydrolase activity, acting on glycosyl bonds ↕
	● 1.0e+0	0016787	hydrolase activity ↕
GO Biological Process:	● 1.1e-63	0030245	cellulose catabolism ↕
	● 1.2e-54	0000272	polysaccharide catabolism ↕
	● 3.6e-20	0005975	carbohydrate metabolism ↕
GO Cellular Component:	● 1.2e-23	0005576	extracellular region ↕

● Information annotated in the MSD database.

● High, ● medium and ● low confidence annotations not annotated in the MSD database.

● High, ● medium and ● low confidence annotations already annotated in the MSD database.

Scoring function

Fisher's 2x2 contingency test

	Non-similar	Similar	Total
Annotated	a	b	a+b
Not Annotated	c	d	c+d
Total	a+c	b+d	n

1b78A SCOP c.51.4.1	Similar	Not similar	Total
Annotated	4	2	6
Not Annotated	0	71,096	71,096
Total	4	71,098	71,102

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}}$$

$$= \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!a!b!c!d!}$$

$$p = 1.78e^{-19}$$

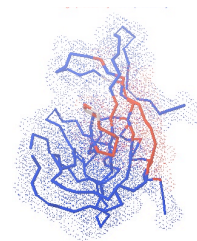
Sensitivity .vs. Precision

	Optimal cut-off	Sensitivity (%) Recall or TPR	Precision (%)
SCOP fold	1e-6	92.7	88.4
CATH fold	1e-3	95.7	90.1
InterPro	1e-3	88.4	78.2
PFam family	1e-4	90.5	82.8
EC number	1e-4	93.3	79.7
GO Molecular Function	1e-1	84.3	80.9
GO Biological Process	1e-3	85.5	74.8
GO Cellular Component	1e-2	77.6	58.6

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{Precision} = \frac{TP}{TP + FP}$$

AnnoLyze

Inherited ligands: 4			
Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
MO2	59.03	0.185	48 49 52 62 63 66 67 113 116
CRY	20.00	0.111	23 29 31 37 44 48 49 83 85 94 96 103 121
BOG	20.00	0.111	19 20 21 48 49 51 96 98 136
ACY	15.87	0.163	23 29 31 37 44 45 81 83 85 94 96 98 103 121 135
Inherited partners: 1			
Partner	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
d.113.1.1	23.68	0.948	19 20 50 51 52 53 54 55 56 57 58 77 78 79 80 81 82 83 84 85 93 95 97 99 134 135 138 142 145



Benchmark

	Number of chains
Initial set*	78,167
LigBase**	30,126
Non-redundant set***	4,948 (8,846 ligands)

**all PDB chains larger than 30 aminoacids in length (8th of August, 2006)*

***annotated with at least one ligand in the LigBase database*

****not two chains can be structurally aligned within 3Å, superimposing more than 75% of their C atoms, result in a sequence alignment with more than 30% identity, and have a length difference inferior to 50aa*

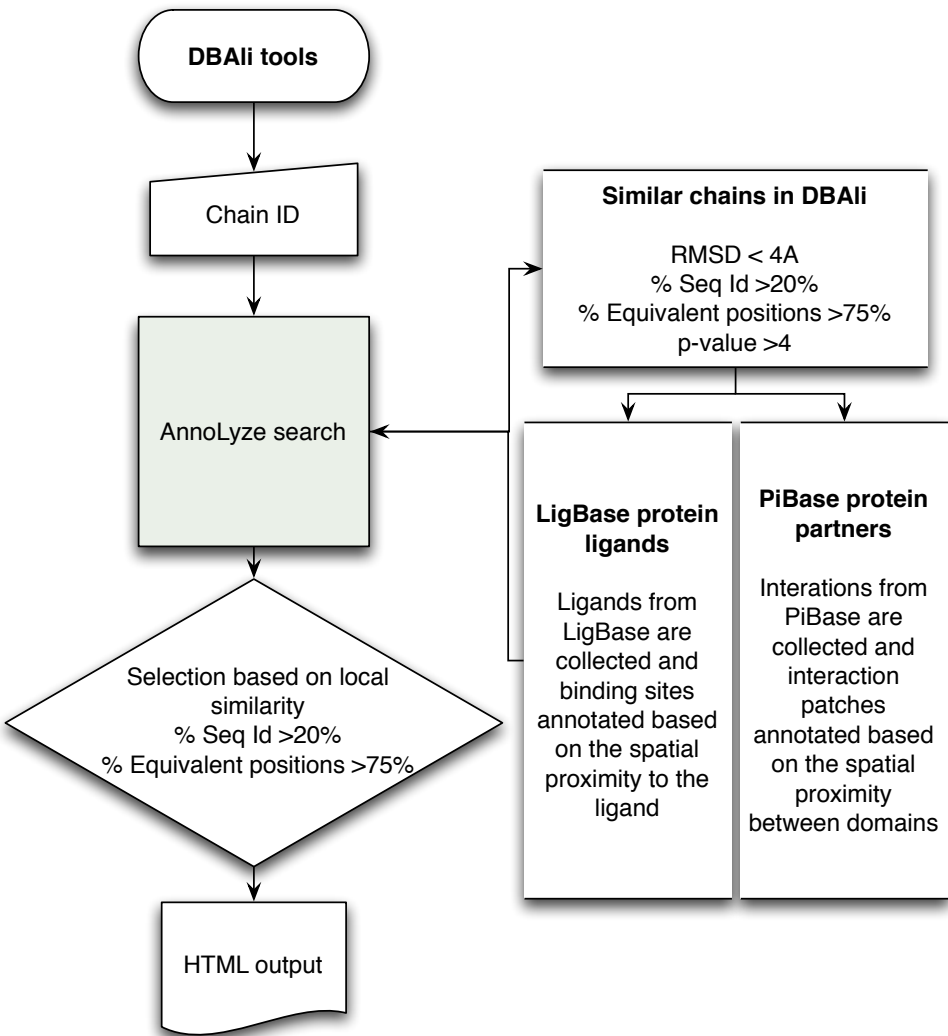
	Number of chains
Initial set*	78,167
πBase**	30,425
Non-redundant set***	4,613 (11,641 partnerships)

**all PDB chains larger than 30 aminoacids in length (8th of August, 2006)*

***annotated with at least one partner in the Base database*

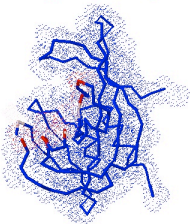
****not two chains can be structurally aligned within 3Å, superimposing more than 75% of their C atoms, result in a sequence alignment with more than 30% identity, and have a length difference inferior to 50aa*

Method



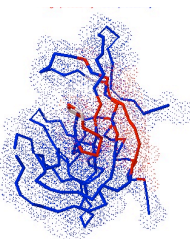
Inherited ligands: 4

Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
MO2	59.03	0.185	48 49 52 62 63 66 67 113 116
CRY	20.00	0.111	23 29 31 37 44 48 49 83 85 94 96 103 121
8OG	20.00	0.111	19 20 21 48 49 51 96 98 136
ACY	15.87	0.163	23 29 31 37 44 45 81 83 85 94 96 98 103 121 135



Inherited partners:1

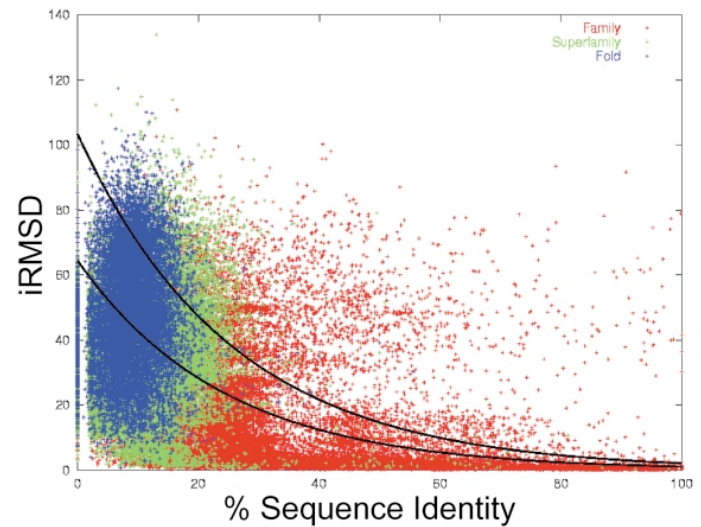
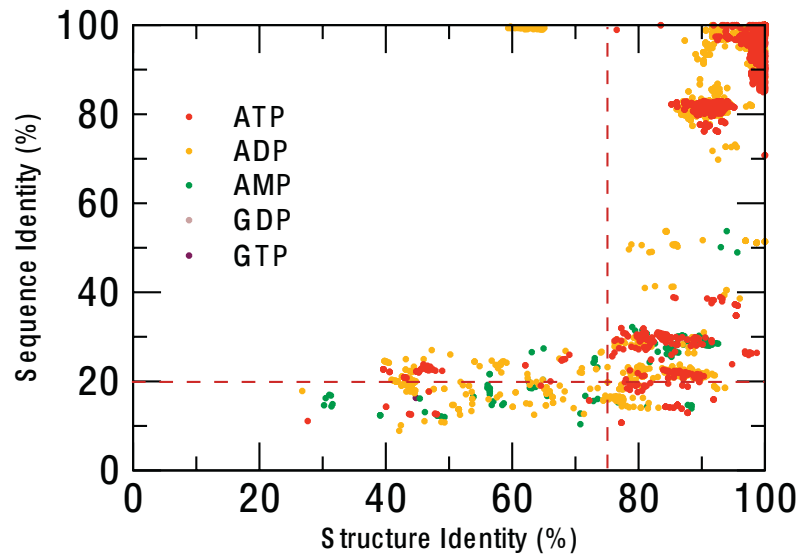
Partner	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
d.113.1.1	23.68	0.948	19 20 50 51 52 53 54 55 56 57 58 77 78 79 80 81 82 83 84 85 93 95 97 99 134 135 138 142 145



Scoring function

Ligands

Partners



Aloy *et al.* (2003) J.Mol.Biol. 332(5):989-98.

Sensitivity .vs. Precision


	Optimal cut-off	Sensitivity (%) Recall or TPR	Precision (%)
Ligands	30%	71.9	13.7
Partners	40%	72.9	55.7

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{Precision} = \frac{TP}{TP + FP}$$

Example (2azwA)

Structural Genomics Unknown Function

Molecule: MutT/nudix family protein

PDB ID: 2azwA	
Header: STRUCTURAL GENOMICS, UNKNOWN FUNCTION	
Compound: MOL_ID: 1; MOLECULE: MUTT/NUDIX FAMILY PROTEIN; CHAIN: A; ENGINEERED: YES	
Source: MOL_ID: 1; ORGANISM: SCIENTIFIC: ENTEROCOCCUS FAECALIS V583; ORGANISM: COMMON: BACTERIA; EXPRESSION_SYSTEM: ESCHERICHIA COLI; EXPRESSION_SYSTEM_COMMON: BACTERIA; EXPRESSION_SYSTEM_STRAIN: BL21(DE3); EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID; EXPRESSION_SYSTEM_PLASMID: PET15B	Resolution: 1.90Å
Links: none	SCOP: none CATH: none
Sequence: Mds: 09b13d23ceae0dfcaddec636e2ddfa6KTPTAAS Length: 146	Ligands: none Interacting partners: none
	
KTPTFGKREE TLTYQTRYAA YIIIVSKPENN TMLVQAPNG AYFLPGGEIE GTETKEAHH REVLLEELGIS VEIGCYLGEA DEYFYSNHRQ TAYYNPGYFY VANTWRQLSE PLRNTLHWV APEEAVRLK RGSRWAVEK WLAAS	

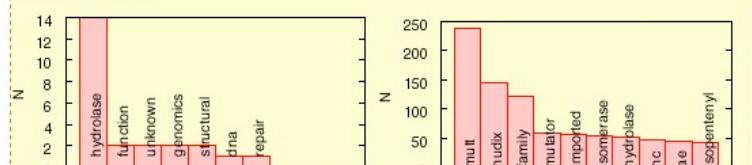
Similar structures: 20	P-value distribution:
Similar sequences: 890	P-value distribution for similar chains
Most similar structure in DBAli:	

Code	SeqId(%)	EqPos	RMSD	P-Value	See
1vc9:A	22.76	123	3.57	17.28	ali

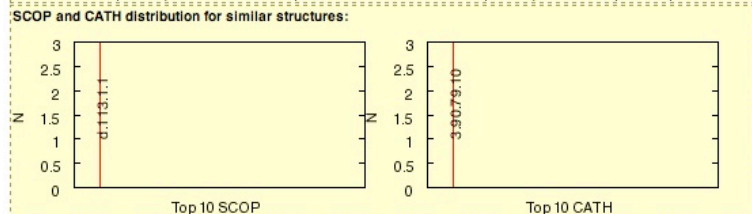
Code	SeqId(%)	EqPos	RMSD	P-Value	See
1vc9:B	24.59	122	3.47	17.00	ali

Code	SeqId(%)	EqPos	RMSD	P-Value	See
1vc9:B	24.59	122	3.47	17.00	ali

Keyword distribution:

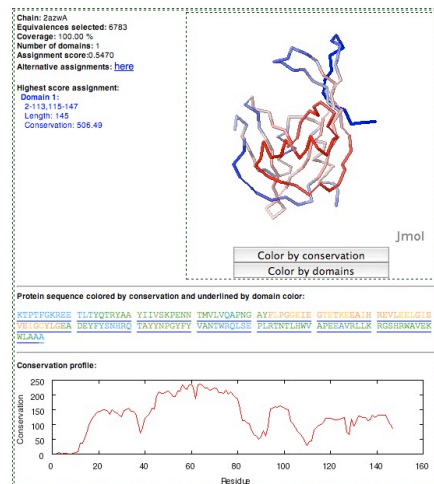


SCOP and CATH distribution for similar structures:



Inherited ligands: 4			
Ligand	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
MO2	59.03	0.185	48 49 52 62 63 66 67 113 116
CRY	20.00	0.111	23 29 31 37 44 48 49 83 85 94 96 103 121
BOG	20.00	0.111	19 20 21 48 49 51 96 98 136
ACY	15.87	0.163	23 29 31 37 44 45 81 83 85 94 96 98 103 121 135

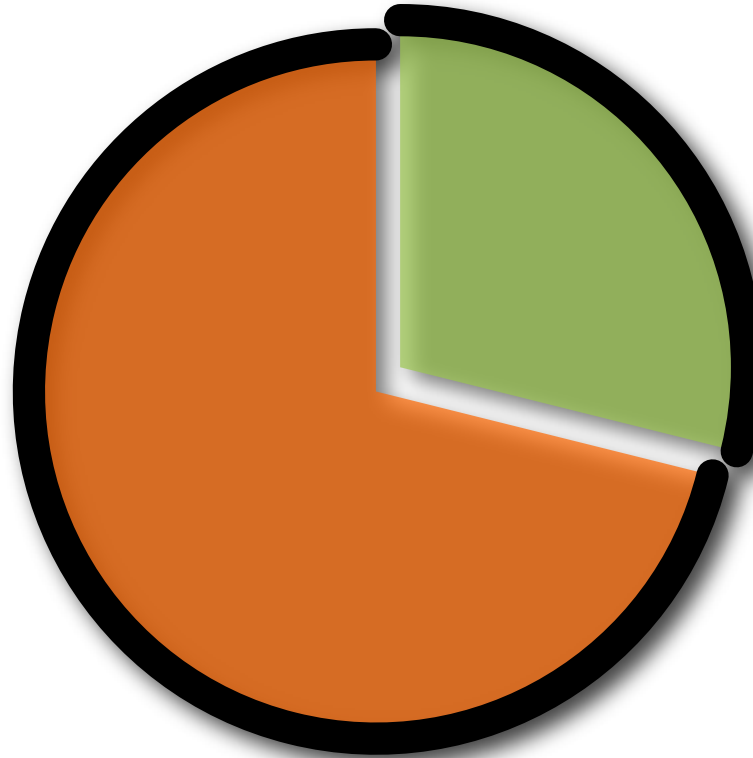
Inherited partners: 1			
Partner	Av. binding site seq. id.	Av. residue conservation	Residues in predicted binding site (size proportional to the local conservation)
d.113.1.1	23.68	0.948	19 20 50 51 52 53 54 55 56 57 58 77 78 79 80 81 82 83 84 85 93 95 97 99 134 135 138 142 145



	Conf. P-value	Link	Description
CATH:	1.1e-20	3.90.79.10	Nucleoside Triphosphate Pyrophosphohydrolase
SCOP:	4.2e-29	d.113.1.1	MutT-like
PFAM:	2.0e-74	PF00293	NUDIX domain
InterPro:	1.9e-65	IPR000086	NUDIX hydrolase
	2.7e-20	IPR003561	Mutator MutT
	2.9e-14	IPR002667	Isopentenyl-diphosphate delta-isomerase
EC Number:	1.7e-4	3.6.1.17	Bis(5'-nucleosyl)-tetraphosphatase (asymmetrical).
GO Molecular Function:	4.5e-19	0008413	8-oxo-7,8-dihydroguanine triphosphatase activity
	3.8e-13	0004452	isopentenyl-diphosphate delta-isomerase activity
	1.9e-6	0016787	hydrolase activity
	5.4e-3	0004081	bis(5'-nucleosyl)-tetraphosphatase (asymmetrical) activity
	1.9e-2	0000287	magnesium ion binding
GO Biological Process:	7.7e-11	0008299	isoprenoid biosynthesis
	1.5e-5	0006974	response to DNA damage stimulus
	1.7e-5	0006260	DNA replication
	2.4e-5	0006281	DNA repair

Tropical Disease Initiative (TDI)

Predicting binding sites in protein structure models.



<http://www.tropicaldisease.org>

“Unprofitable” Diseases and Global DALY (in 1000’s)

Malaria*	46,486
Tetanus	7,074
Lymphatic filariasis*	5,777
Syphilis	4,200
Trachoma	2,329
Leishmaniasis*	2,090
Ascariasis	1,817
Schistosomiasis*	1,702
Trypanosomiasis*	1,525

Trichuriasis	1,006
Japanese encephalitis	709
Chagas Disease*	667
Dengue*	616
Onchocerciasis*	484
Leprosy*	199
Diphtheria	185
Poliomyelitis	151
Hookworm disease	59

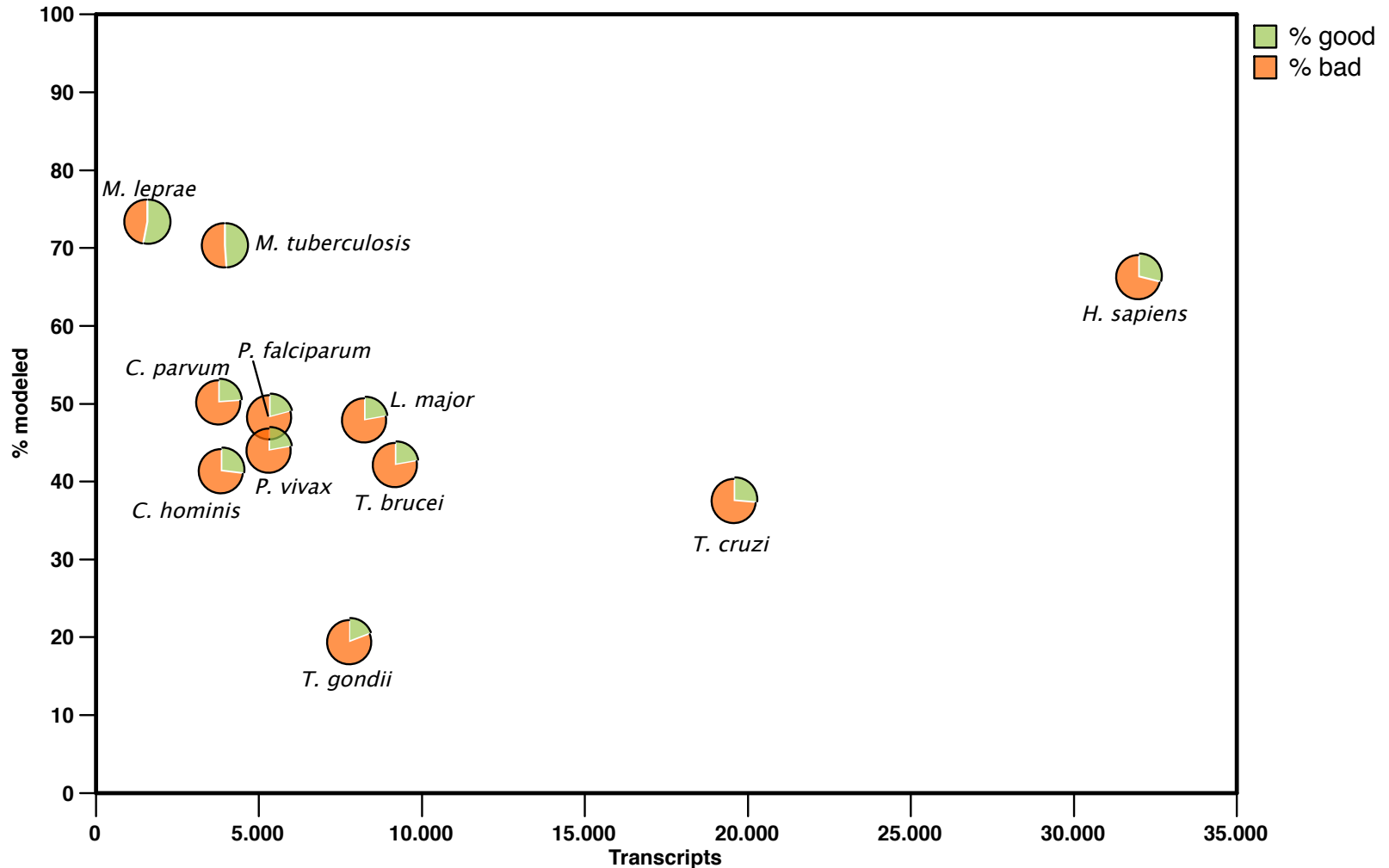
Disease data taken from WHO, *World Health Report 2004*

DALY - Disability adjusted life year in 1000’s.

* Officially listed in the WHO Tropical Disease Research [disease portfolio](#).

Modeling Genomes

data from models generated by ModPipe (Eswar, Pieper & Sali)

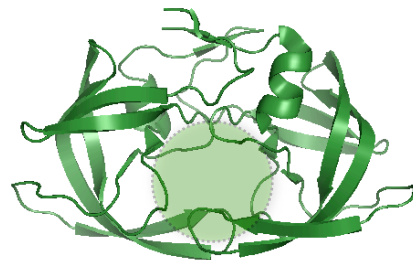
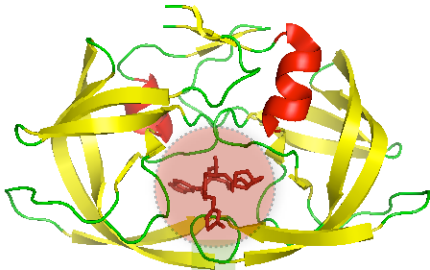


A good model has MPQS of 1.1 or higher

Comparative docking

1. Expansion

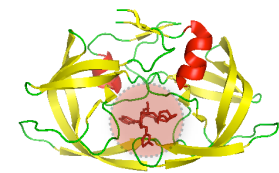
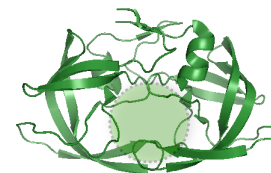
co-crystallized protein/ligand



crystalized protein

2. Inheritance

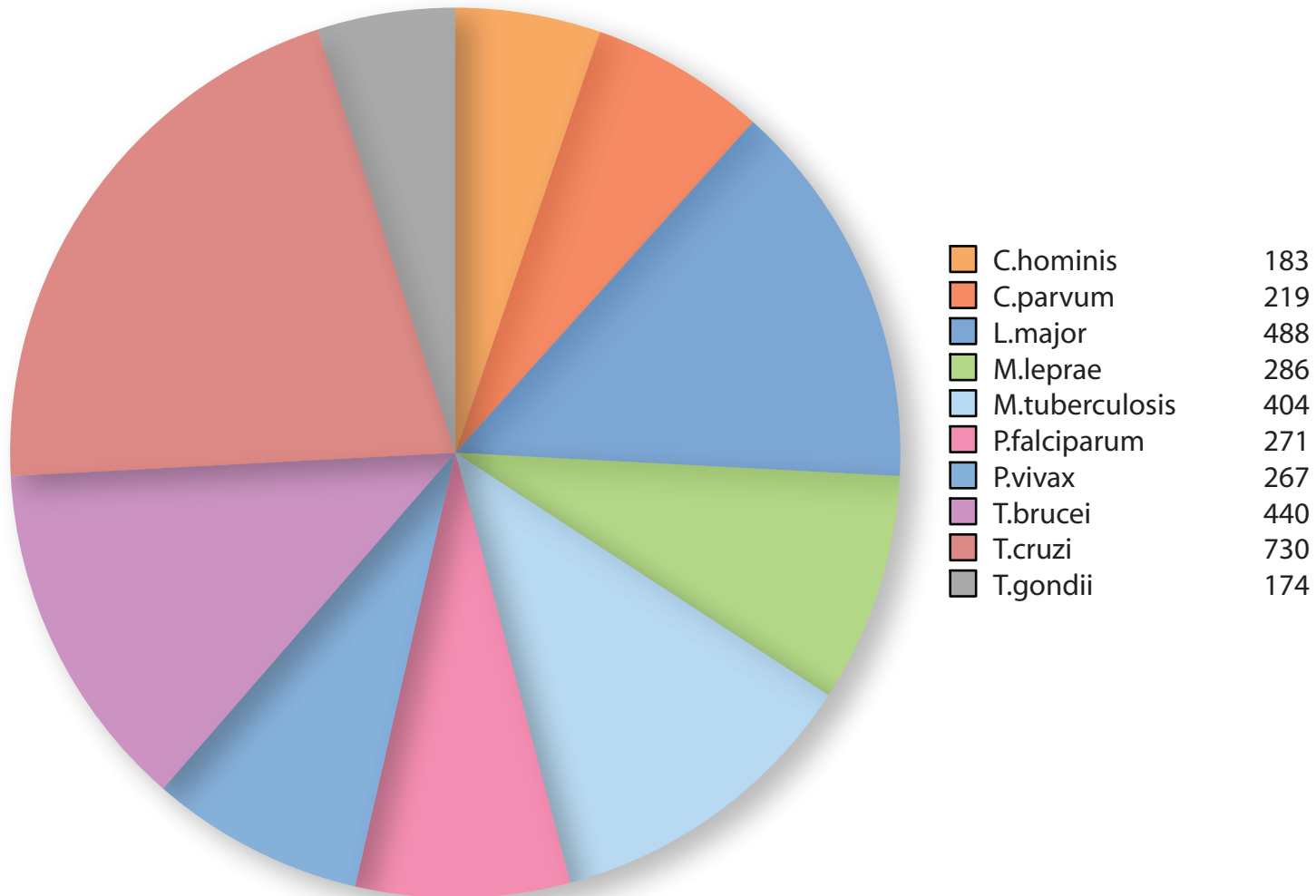
model



template

Distribution of models with inherited ligands

from 3,882 “good” models
using a 90% / 90% “inherited” cut-offs



Summary table

models with inherited ligands

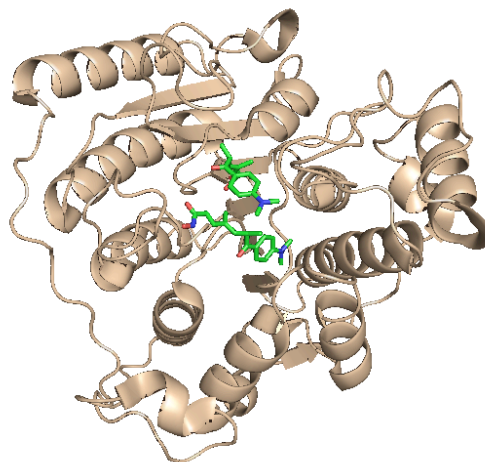
from 16,284 good models, 295 inherited a ligand/substance with at least one compound already approved by FDA and ready to be used from ZINC

	Transcripts	Good	Ligands	Lipinski	Lipinski+ZINC	FDA+ZINC
<i>C. hominis</i>	3,886	886	183	131	28	12 (10)
<i>C. parvum</i>	3,806	949	219	145	30	12 (10)
<i>L. major</i>	8,274	1,845	488	334	84	44 (34)
<i>M. leprae</i>	1,605	1,321	286	189	39	29 (25)
<i>M. tuberculosis</i>	3,991	2,887	404	285	71	44 (37)
<i>P. falciparum</i>	5,363	1,057	271	191	48	20 (16)
<i>P. vivax</i>	5,342	1,042	267	177	37	18 (15)
<i>T. brucei</i>	921	1,795	440	309	94	46 (36)
<i>T. cruzi</i>	19,607	3,915	730	493	127	62 (52)
<i>T. gondii</i>	7,793	587	174	124	28	8 (7)
TOTAL	60,588	16,284	3,462	2,378	586	295 (242)

Example of inheritance (expansion)

LmjF2 1.0680 from L. major “Histone deacetylase 2” (model 1)

Template 1t64A a human HDAC8 protein.



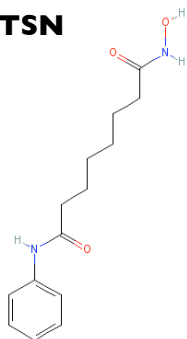
	Origen	Formula	Name	Cov.	Seq. Id. (%)
ZN	X-ray	Zn ²⁺	Zinc ion	--	--
NA	X-ray	Na ⁺	Sodium ion	--	--
CA	X-ray	Ca ²⁺	Calcium ion	--	--
TSN	X-ray	C ₁₇ H ₂₂ N ₂ O ₃	Trichostatin A	--	--
SHH	Expanded	C ₁₄ H ₂₀ N ₂ O ₃	Octadenioic acid hydroxyamide phenylamide	100.00	83.8

Example of inheritance (inheritance)

LmjF21.0680 from L. major "Histone deacetylase 2" (model 1)

	Formula	Name	Cov.	Seq. Id. (%)	Residues
TSN	C ₁₇ H ₂₂ N ₂ O ₃	Trichostatin A	100.00	90.9	90 131 132 140 141 167 169 256 263 293 295
SHH	C ₁₄ H ₂₀ N ₂ O ₃	Octadenioic acid hydroxyamide phenylamide	100.00	90.9	

TSN



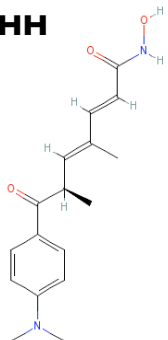
suberoylanilide hydroxamic acid

Pharmacological Action:

[Anti-Inflammatory Agents, Non-Steroidal](#)
[Antineoplastic Agents](#)
[Enzyme Inhibitors](#)
[Anticarcinogenic Agents](#)

Inhibits histone deacetylase 1 and 3

SHH



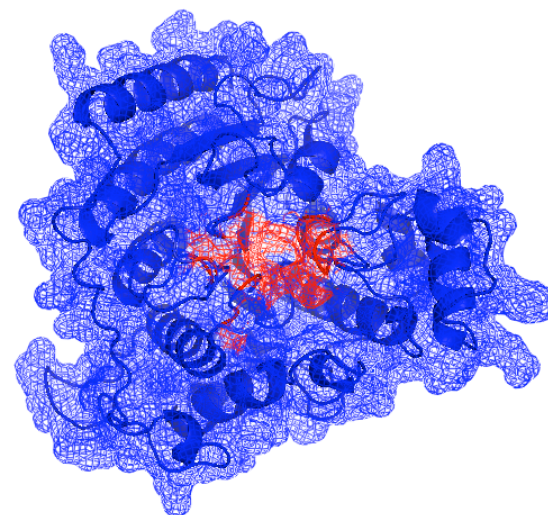
trichostatin A

Pharmacological Action:

[Antibiotics, Antifungal](#)
[Enzyme Inhibitors](#)
[Protein Synthesis Inhibitors](#)

chelates zinc ion in the active site of histone deacetylases, resulting in preventing histone unpacking so DNA is less available for transcription

	LmjF21.0680.1.pdb
Template	1t64A
Seq. Id (%)	38.00
MPQS	1.47



Example of inheritance (CDD-Roos-literature)

LmjF21.0680 from L. major “Histone deacetylase 2” (model 1)

Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 13143–13147, November 1996
Medical Sciences

Apicidin: A novel antiprotozoal agent that inhibits parasite histone deacetylase

(cyclic tetrapeptide/Apicomplexa/antiparasitic/malaria/coccidiosis)

SANDRA J. DARKIN-RATTRAY*[†], ANNE M. GURNETT*, ROBERT W. MYERS*, PAULA M. DULSKI*,
TAMI M. CRUMLEY*, JOHN J. ALLOCCO*, CHRISTINE CANNOVA*, PETER T. MEINKE[‡], STEVEN L. COLLETTI[‡],
MARIA A. BEDNAREK[‡], SHEO B. SINGH[§], MICHAEL A. GOETZ[§], ANNE W. DOMBROWSKI[§],
JON D. POLISHOOK[§], AND DENNIS M. SCHMATZ*

Departments of *Parasite Biochemistry and Cell Biology, [‡]Medicinal Chemistry, and [§]Natural Products Drug Discovery, Merck Research Laboratories, P.O. Box 2000, Rahway, NJ 07065

ANTIMICROBIAL AGENTS AND CHEMOTHERAPY, Apr. 2004, p. 1435–1436
0066-4804/04/\$08.00+0 DOI: 10.1128/AAC.48.4.1435–1436.2004
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

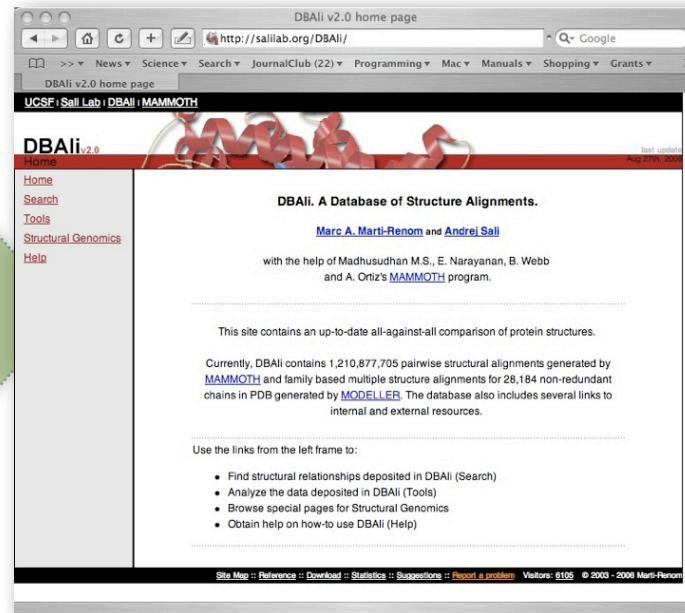
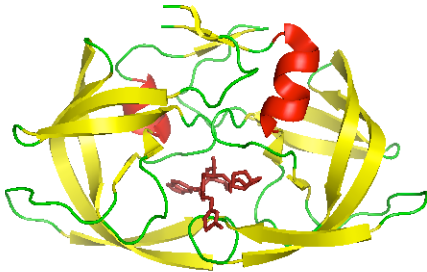
Vol. 48, No. 4

Antimalarial and Antileishmanial Activities of Aroyl-Pyrrolyl-Hydroxyamides, a New Class of Histone Deacetylase Inhibitors

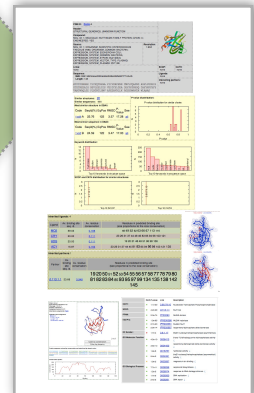
DBAli future work

<http://bioinfo.cipf.es/squ/services/DBAli/>

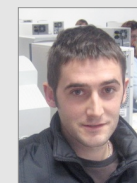
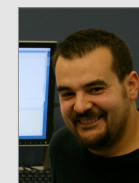
<http://www.salilab.org/DBAli/>



RSS



Acknowledgments



COMPARATIVE MODELING

Andrej Sali

M. S. Madhusudhan

Narayanan Eswar

Min-Yi Shen

Ursula Pieper

Ben Webb

Maya Topf

MODEL ASSESSMENT

David Eramian

Min-Yi Shen

Damien Devos

FUNCTIONAL ANNOTATION

Andrea Rossi

Fred Davis

FUNDING

Prince Felipe Research Center

Marie Curie Reintegration Grant

STREP EU Grant

MODEL ASSESSMENT

Francisco Melo (CU)

Alejandro Panjkovich (CU)

STRUCTURAL GENOMICS

Stephen Burley (SGX)

John Kuriyan (UCB)

NY-SGXRC

MAMMOTH

Angel R. Ortiz

FUNCTIONAL ANNOTATION

Fatima Al-Shahrour

Joaquin Dopazo

BIOLOGY

Jeff Friedman (RU)

James Hudsped (RU)

Partho Ghosh (UCSD)

Alvaro Monteiro (Cornell U)

Stephen Krilis (St. George H)

Tropical Disease Initiative

Stephen Maurer (UC Berkeley)

Arti Rai (Duke U)

Andrej Sali (UCSF)

Ginger Taylor (TSL)

CCPR Functional Proteomics

Patsy Babbitt (UCSF)

Fred Cohen (UCSF)

Ken Dill (UCSF)

Tom Ferrin (UCSF)

John Irwin (UCSF)

Matt Jacobson (UCSF)

Tack Kuntz (UCSF)

Andrej Sali (UCSF)

Brian Shoichet (UCSF)

Chris Voigt (UCSF)

EVA

Burkhard Rost (Columbia U)

Alfonso Valencia (CNB/UAM)

CAMP

Xavier Aviles (UAB)

Hans-Peter Nester (SANOFI)

Ernst Meinjohanns (ARPIDA)

Boris Turk (IJS)

Markus Gruetter (UE)

Matthias Wilmanns (EMBL)

Wolfram Bode (MPG)