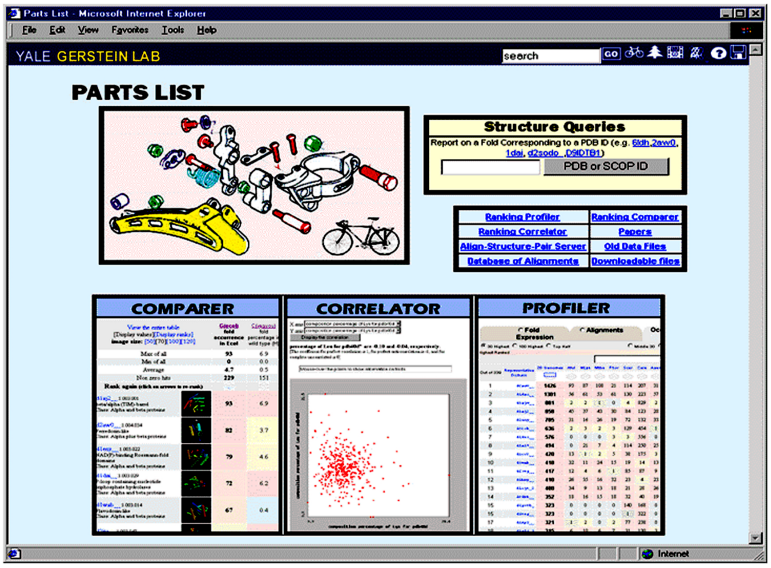
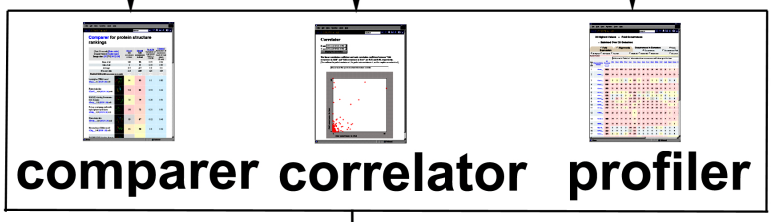
Category	Symbol	Definition of Symbol	Attributes in Category	ref.
Genome Occurrence	G(x)	Number of times a particular *PART* occurs in genome x. (These are based on PSI-blast comparisons between PDB and the genomes with an e-value cutoff in these comparisons of .0001.)	20	(19.31,32)
Expression	L(e)	average expression level over all genes that contain a this *PART.*	8	(33)
	C(e)	*PART* composition of the yeast transcriptome in expression level experiment e. This refers to the fraction of the mRNA population with this *PART* as opposed to all other parts. (This is only applicable to expression experiments, such as SAGE and GeneChips, that measure absolute mRNA levels in copies per cell.)	8	
	E(e)	Transcriptome enrichment compared to genome in experiment e. (Transcriptome enrichment is defined as percentage difference of *PART* composition in the transcriptome and the genome. In symbols: E(e) = [C(e)-G(Scer)] / G(Scer) .)	8	
	F(r)	Expression level fluctuation in experiment r. (This is the standard deviation in the expression ratio measurement R(i,t) over a timecourse, viz: $<(R(i,t)-<R(i,t)>)^2>$ where one averages over all times t and genes i that have a particular *PART*.	7	(34)
Alignments	V(f)	The number of aligned pairs in pair-set f.	2	(35)
	U(f)	RMS deviation in Cα atoms averaged over all alignments in pair-set f	2	
	R(f)	Similar to U(f) for pair-set f but only the best fitting half of the atoms are included in the calculation	2	
	S(f)	Average percentage identity between pairs of aligned proteins in pair-set f	2	
	P(f)	Average sequence P-value for pair-set f	2	
	Q(f)	Average structural P-value for pair-set f	2	
Compositions	N(p)	The number of structures associated with a particular *PART* in dataset p.	2	
	B(a,p)	Composition of amino acid a in a particular *PART* where one averages over all structures in dataset p associated with the *PART*	40	
Motion	M(s,d)	The maximum value of statistic s derived from surveying set of motions d in the Macromolecular Motions Database for a particular *PART*, where s is only calculated from the entries in the database that are associated with the *PART*.	7	(36,37)
	A(s,d)	Similar to M(s,d) but now we take the average instead of the maximum.	7	
Interaction	I(y,c)	For a given *PART*, the number of types of protein-protein interactions in interaction dataset y subject to the restriction c regarding whether or not the proteins are on the same chain. The number of interaction types is the number of distinctly different *PARTs* that interacts with a given *PART*.	24	(38,39)
	J(y,c)	For a given PART, the total number of types of interactions in interaction dataset y subject to the restriction c regarding whether or not the proteins are on the same chain. Here we show all interactions observed not just the number of distinct *PART-PART* interactions tabulated in I(y,c).	24	
Transposon	T(b)	The sensitivity of the cell to a transposon inserted into genes containing a particular *PART* under different growth condition b. The sensitivity was indicated by negative logarithm of a P-value, which measures the degree to which the observations for one particular gene could have resulted from wild-type cells that randomly change their phenotype.	20	(40)
Miscelleneous	X(q)	Various miscellaneous ranks	5	
Total			182	

Attributes	Value	Description	ref.
Genome x =	aful	*Archaeoglobus fulgidus*	(19, 31, 32)
	mjan	*Methanococcus jannaschii*	
	mthe	*Methanobacterium thermoautotrophicum*	
	phor	*Pyrococcus horikoshii*	
	scer	*Saccharomyces cerevisiae*	
	cele	*Caenorhabditis elegans*	
	aaeo	*Aquifex aeolicus*	
	syne	*Synechocystis sp.*	
	ecol	*Escherichia coli*	
	bsub	*Bacillus subtilis*	
	mtub	*Mycobacterium tuberculosis*	
	hinf	*Haemophilus influenzae Rd*	
	hpyl	*Helicobacter pylor*	
	mgen	*Mycoplasma genitalium*	
	mpne	*Mycoplasma pneumoniae*	
	bbur	*Borrelia burgdorferi*	
	tpal	*Treponema pallidum*	
	ctra	*Chlamydia trachomatis*	
	cpne	*Chlamydia pneumoniae*	
	rpro	*Rickettsia prowazekii*	
Absolute Expression Expt. e =	vegsam	GeneChip mRNA expression analysis of 6200 yeast ORFs under vegetative growth conditions.	(41)
	vegyou	GeneChip mRNA expression analysis of 5455 yeast ORFs under vegetative growth conditions.	(42)
	sage	mRNA expression analysis of 3788 yeast ORFs determined by Serial Analysis of Gene Expression.	(43)
	matea	GeneChip mRNA expression analysis of yeast mating type a strain grown on glucose.	(44)
	mateal	GeneChip mRNA expression analysis of yeast mating type alpha strain grown on glucose	
	gal	GeneChip mRNA expression analysis of yeast mating type a strain grown on galactose	
	heat	GeneChip mRNA analysis of yeast mating type a strain grown on glucose at 30 degree before a 39 degree heat shock.	
	ref	Reference transcriptome. This is a scaling and merging of the above experiments.	(33)
Microarray Expt. r=	cdc28	cDNA microarray genome-wide characterization of mRNA transcript levels for CDC28 synchronized yeast cells during the cell cycle.	(45)
	cdc15	cDNA microarray genome-wide characterization of mRNA transcript levels for CDC15 synchronized yeast cells during the cell cycle.	
	alpha	Analysis using cDNA microarrays of yeast mRNA levels after synchronization of cell cycle via alpha arrest factor	
	diaux	Genome-wide cDNA microarray analysis of the temporal program of yeast mRNA expression accompanying the metabolic shift from fermentation to respiration	(46)
	spor	cDNA microarray genome-wide analysis to assay changes in gene expression during sporulation.	(47)
	heatec	cDNA microarray experiment and analysis on 4290 *E.coli* ORFs after exposure of the bacteria to heat shock.	(48)
	deve	Analysis of genome wide changes during successive larval stages using cDNA microarrays of ~12000 *C. elegan* ORFs.	(49)
Pair set f=	all	All pairs within a *PART* included in the calculations in Wilson et al. (For example, for fold rankings this would be the total number of pairs within a fold.)	(35)
	foldonly	A subset of the pair-set "all" that only includes pairs between structures that are in the same *PART* but different *sub-PART*. (If *PART* is fold, then *sub-PART* is superfamily; If *PART* is superfamily, then *sub-PART* is family.)	
Amino Acid a=		Ala, Cys, Asp, Glu, Phe, Gly, His, Ile, Lys, Leu, Met, Asn, Pro, Gln, Arg, Ser, Thr, Val, Trp, Tyr.	(50)
Data set p=	pdb100	All structures within the fold (as defined by SCOP pdb100d)	(50)
	pdb40	Similar to pdb100 but now using a version of the PDB clustered at 40% similarity (as defined by SCOP pdb40d)	

Attributes	Value	Description	ref.
Interaction type y=	pdball	Interactions for a *PART* are computed with all other *PARTS* in the PDB databank based on the distances between atoms in the coordinate files. Five or more contacts between atoms separated by less than 5 A was considered a valid PART-PART contact.	(9,51, 52)
	pdba	A subset of "pdball". Interactions for a *PART* are computed just with all-alpha proteins (SCOP class 1) in the PDB.	
	pdbb	Similar to "pdba" but now just with all-beta proteins (SCOP class 2).	
	pdbab	Similar to "pdba" but now just with mixed helix-sheet proteins (SCOP class 3 and 4)	
	scerall	Interactions for a *PART* are computed with all other *PARTS* based on the yeast two-hybrid experimental data. In particular, interactions between structural domains in the yeast genome were obtained by assigning protein structures to the yeast proteins. Structural domains contained within the same ORF that were within 30 amino acids were assumed to interact in an intramolecular fashion. To derive intermolecular interactions, we combined three sets of protein-protein interactions: (i) the MIPS web pages on complexes and pairwise interactions (February 2000)(9), (ii) the global yeast-two-hybrid experiments by Uetz et al. (45) and (iii) large-scale yeast two-hybrid experiments by Ito et al. (46). Out of all these pairwise interactions known for yeast ORFs, there is a limited set in which both partners are completely covered by one structural domain (to within 100 residues).	
	scera	A subset of "scerall". Interactions for a *PART* are computed just with all-alpha proteins (SCOP class 1) in the yeast experiment.	
	scerb	Similar to "scera" but now just with all-beta proteins (SCOP class 2).	
	scerab	Similar to "scera" but now just with mixed helix-sheet proteins (SCOP class 3 and 4)	
Interaction restriction c=	inter	The interaction must occur between *PARTS* in different chains	
	intra	The interaction must occur between *PARTS* in the same chain.	
	none	The union of "inter" and "intra". Interactions can occur in *PARTS* on the same or different chains.	
Motion statistic s=	nresidue	Number of residues	(36,37)
	maxcadev	Maximal displacement of an Cα atom, in angstroms, of any residue during the motion (after fitting on the first core).	
	rmsoverall	Overall RMS of two structures after they are superimposed by a sieve-fit technique. Note that they are larger than traditionally used RMS (details see ref.).	
	nhinges	Number of hinges involved in the motion.	
	kappa	The rotation (in degrees) around the screw axis necessary to superimpose two domains of motion.	
	transe	Transition energy of the motion (maximum energy less minimum energy over the motion) (in kcal/mole).	
	deltae	Absolute value of energy difference between the "starting" and "ending" conformations of a motion (in kcal/mole).	
Motion dataset d=	goldstd	list of ~220 "gold-standard" manually curated motions	
	auto	list of ~4000 conformational different proteins based on analyzing the SCOP database for similar proteins with large conformational differences (as measured by RMS) but close sequence similarity	
Transposon conditions b=	caff	YPD + 8mM caffeine	(40)
	cyss	Cyclohexmide hypersensitivity: YPD + 0.08 μgml^{-1} cycloheximide at 30^{0}C	
	wr	White/red colour on YPD	
	ypg	YPGlycerol	
	calcs	Calcofluor hypersensitivity: YPD+12μgml^{-1} calcoluor at 30^{0}C	
	hyg	YPD + 46μgml^{-1} hygromycin at 30^{0}C	
	sds	YPD + 0.003%SDS	
	bens	Benomyl hypersensitivity: YPD + 10μgml^{-1} benomyl	
	bcip	YPD + 5-bromo-4-chloro-3-indolyl phosphate at 37^{0}C	
	mb	YPD + 0.001% methylene blue at 30^{0}C	
	benr	Benomyl resistance: YPD + 20μgml^{-1} benomyl	
	ypd37	YPD at 37^{0}C	
	egta	YPD + 2mM EGTA	
	mms	YPD + 0.008% MMS	
	hu	YPD + 75mM hydroxyurea	
	ypd11	YPD at 11^{0}C	
	calcr	Calcofluor resistance: YPD + 0.3μgml^{-1} calcofluor at 30^{0}C	
	cycr	Cyclohexmide resistance: YPD + 0.3μgml^{-1} cycloheximide	
	hhig	Hyperhaploid invasive growth mutants	
	nacl	YPD + 0.9M NaCl	
Misc. quantities q=	pseu	Number of pseudogenes in worm genome matching a particular *PART*	(53)
	func	Total number of functions associated with this *PART*. (In this survey all non-enzyme functions were lumped into a single category.)	(54)
	enz	Total number of enzymatic functions associated with this *PART*.	
	size	Average length of a *PART* in the pdb40d clustering of the PDB.	
	age	The year of the first structure that is part of the *PART* was determined.	





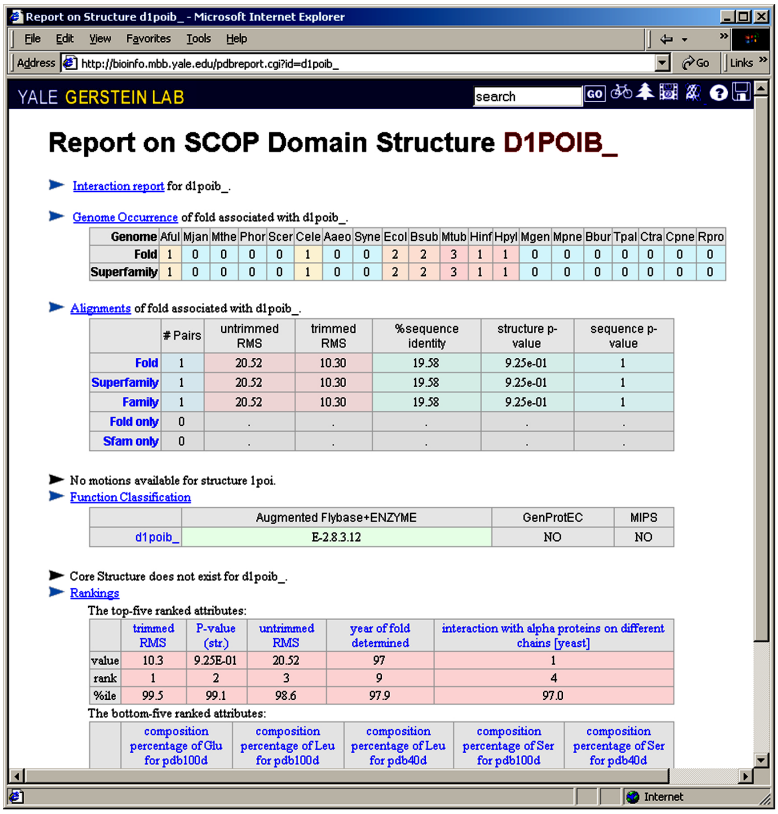
fold report

interaction

occurrence

motion

alignment



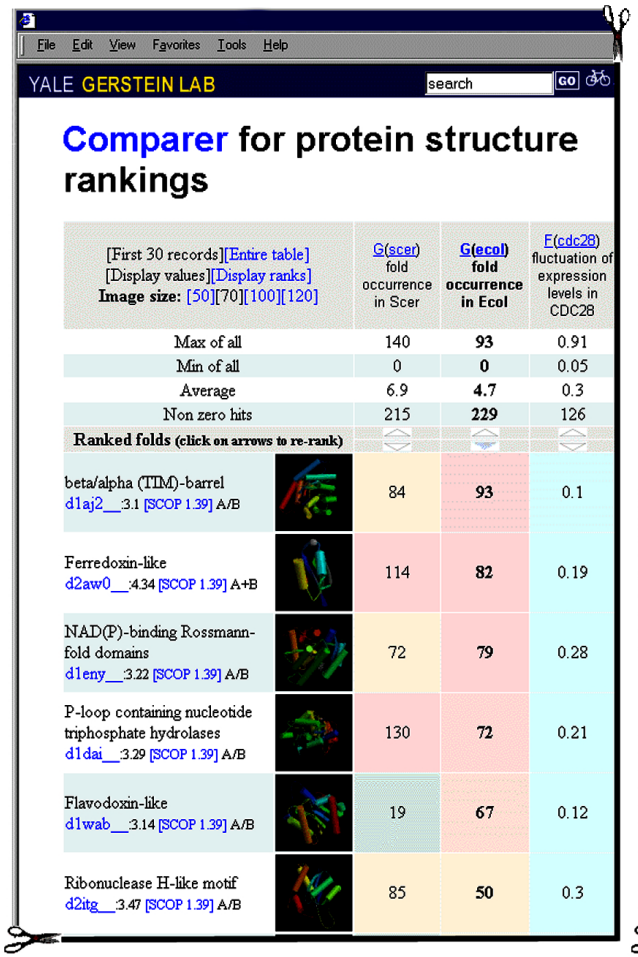
PDB report

core structure

function

rank report

A. Comparer



File Edit View Favorites Tools Help

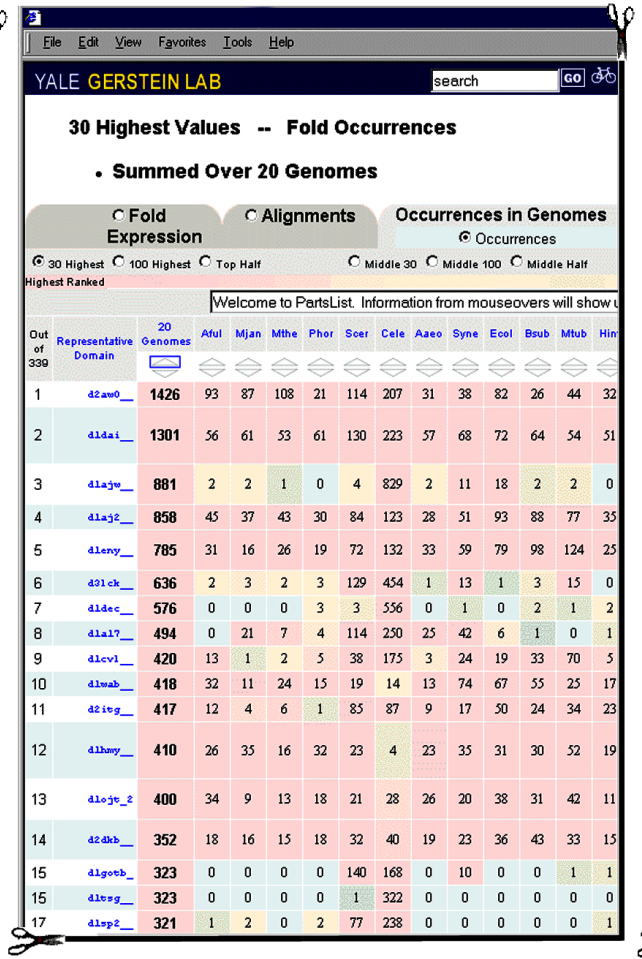
YALE GERSTEIN LAB search GO

Comparer for protein structure rankings

[First 30 records][Entire table]
[Display values][Display ranks]
Image size: [50][70][100][120]

	G(scer) fold occurrence in Scer	**G(ecol) fold occurrence in Ecol**	F(cdc28) fluctuation of expression levels in CDC28
Max of all	140	**93**	0.91
Min of all	0	**0**	0.05
Average	6.9	**4.7**	0.3
Non zero hits	215	**229**	126
Ranked folds (click on arrows to re-rank)			
beta/alpha (TIM)-barrel d1aj2__:3.1 [SCOP 1.39] A/B	84	**93**	0.1
Ferredoxin-like d2aw0__:4.34 [SCOP 1.39] A+B	114	**82**	0.19
NAD(P)-binding Rossmann-fold domains d1eny__:3.22 [SCOP 1.39] A/B	72	**79**	0.28
P-loop containing nucleotide triphosphate hydrolases d1dai__:3.29 [SCOP 1.39] A/B	130	**72**	0.21
Flavodoxin-like d1wab__:3.14 [SCOP 1.39] A/B	19	**67**	0.12
Ribonuclease H-like motif d2itg__:3.47 [SCOP 1.39] A/B	85	**50**	0.3

B. Profiler



File Edit View Favorites Tools Help

YALE GERSTEIN LAB search GO

30 Highest Values -- Fold Occurrences

- **Summed Over 20 Genomes**

Fold Expression | Alignments | Occurrences in Genomes (Occurrences)

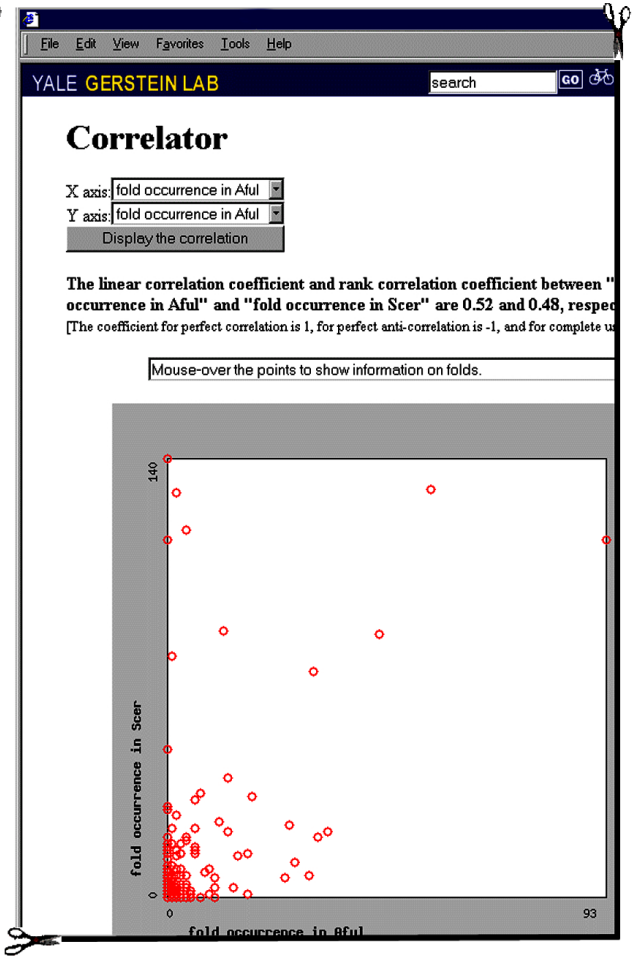
30 Highest | 100 Highest | Top Half | Middle 30 | Middle 100 | Middle Half

Highest Ranked

Welcome to PartsList. Information from mouseovers will show

Out of 339	Representative Domain	20 Genomes	Aful	Mjan	Mthe	Phor	Scer	Cele	Aaeo	Syne	Ecol	Bsub	Mtub	Hin
1	d2aw0__	**1426**	93	87	108	21	114	207	31	38	82	26	44	32
2	d1dai__	**1301**	56	61	53	61	130	223	57	68	72	64	54	51
3	d1ajw__	**881**	2	2	1	0	4	829	2	11	18	2	2	0
4	d1aj2__	**858**	45	37	43	30	84	123	28	51	93	88	77	35
5	d1eny__	**785**	31	16	26	19	72	132	33	59	79	98	124	25
6	d3lck__	**636**	2	3	2	3	129	454	1	13	1	3	15	0
7	d1dec__	**576**	0	0	0	3	3	556	0	1	0	2	1	2
8	d1a17__	**494**	0	21	7	4	114	250	25	42	6	1	0	1
9	d1cv1__	**420**	13	1	2	5	38	175	3	24	19	33	70	5
10	d1wab__	**418**	32	11	24	15	19	14	13	74	67	55	25	17
11	d2itg__	**417**	12	4	6	1	85	87	9	17	50	24	34	23
12	d1hmy__	**410**	26	35	16	32	23	4	23	35	31	30	52	19
13	d1ojt_2	**400**	34	9	13	18	21	28	26	20	38	31	42	11
14	d2dkb__	**352**	18	16	15	18	32	40	19	23	36	43	33	15
15	d1gotb_	**323**	0	0	0	0	140	168	0	10	0	0	1	1
15	d1tsg__	**323**	0	0	0	0	1	322	0	0	0	0	0	0
17	d1sp2__	**321**	1	2	0	2	77	238	0	0	0	0	0	1

C. Correlator



File Edit View Favorites Tools Help

YALE GERSTEIN LAB search GO

Correlator

X axis: fold occurrence in Aful
Y axis: fold occurrence in Aful
Display the correlation

The linear correlation coefficient and rank correlation coefficient between "occurrence in Aful" and "fold occurrence in Scer" are 0.52 and 0.48, respec

[The coefficient for perfect correlation is 1, for perfect anti-correlation is -1, and for complete u

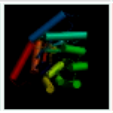
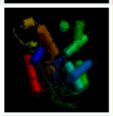
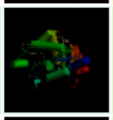
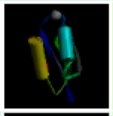
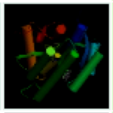
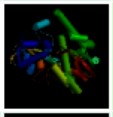
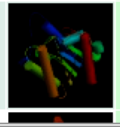
Mouse-over the points to show information on folds.

Rankings2 - Microsoft Internet Explorer

File Edit View Favorites Tools Help

YALE GERSTEIN LAB search GO

Comparer for protein structure rankings

[First 30 records] [Entire table] [Display values] [Display ranks] **Image size:** [50] [70] [100] [120]		X(func) total number of functions associated with the fold	I(pdball,all) interaction types [pdb]
Max of all		**16**	35
Min of all		**1**	1
Average		**1.5**	2.7
Non zero hits		**211**	268
Ranked folds (click on arrows to re-rank)			
beta/alpha (TIM)-barrel d1aj2__:3.1 [SCOP 1.39] A/B		**16**	22
alpha/beta-Hydrolases d1cvl__:3.56 [SCOP 1.39] A/B		**9**	5
P-loop containing nucleotide triphosphate hydrolases d1dai__:3.29 [SCOP 1.39] A/B		**6**	23
Ferredoxin-like d2aw0__:4.34 [SCOP 1.39] A+B		**6**	17
NAD(P)-binding Rossmann-fold domains d1eny__:3.22 [SCOP 1.39] A/B		**6**	7
Cytochrome P450 d1cpt__:1.83 [SCOP 1.39] A		**5**	1
Flavodoxin-like d1wab__:3.14 [SCOP 1.39] A/B		**4**	.

Done Internet

Correlator - Microsoft Internet Explorer

File Edit View Favorites Tools Help

YALE GERSTEIN LAB search GO

Correlator

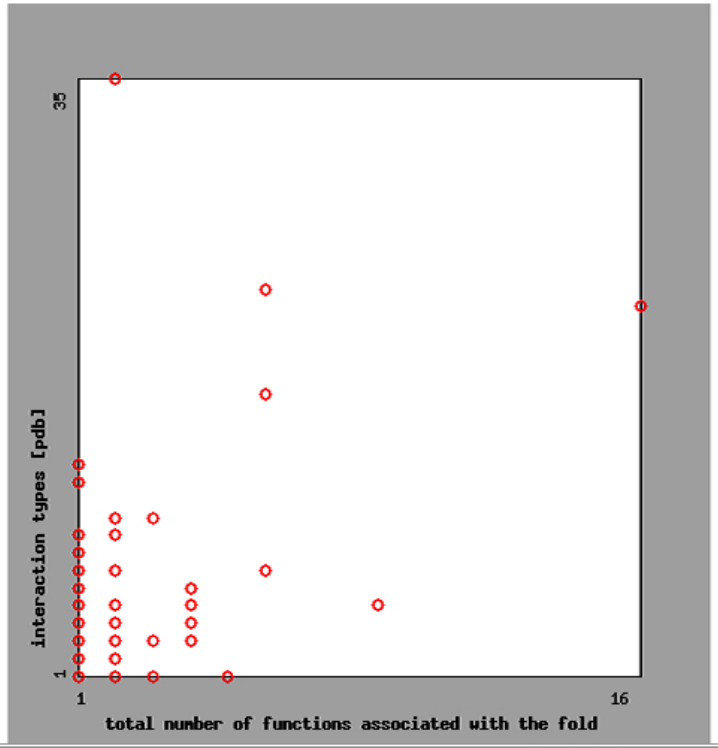
X axis: total number of functions associated with the fold

Y axis: total number of functions associated with the fold

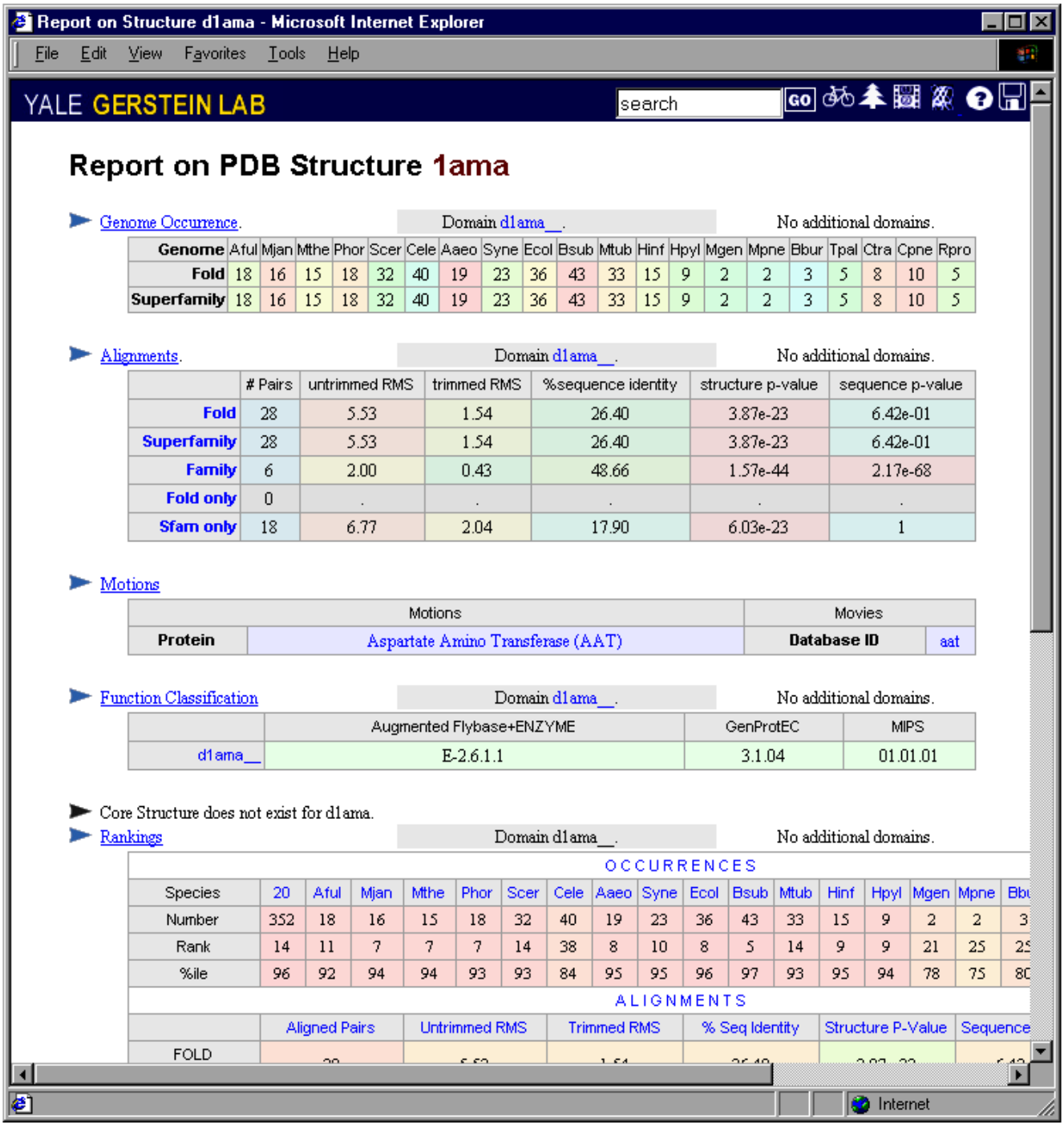
Display the correlation

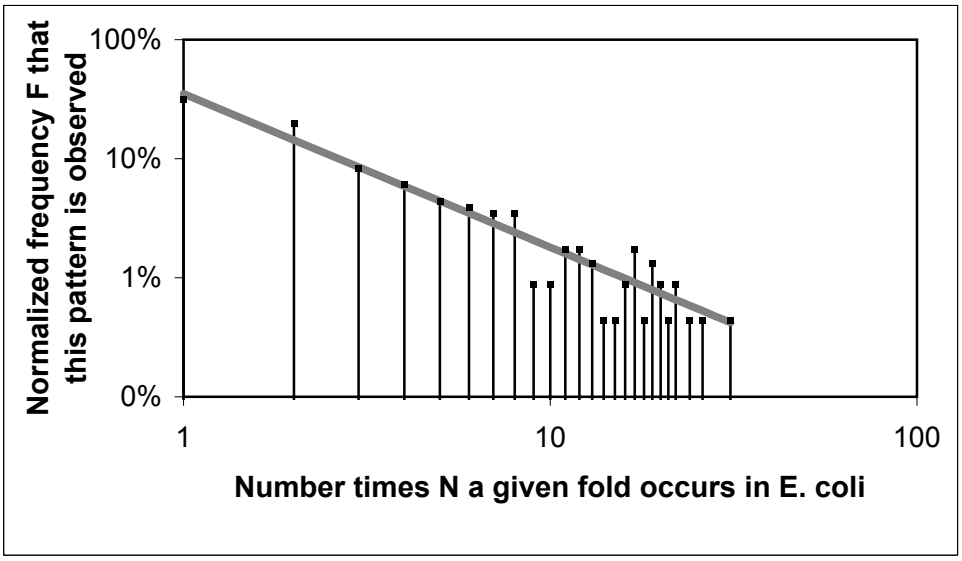
The linear correlation coefficient and rank correlation coefficient between "total number of functions associated with the fold" and "interaction types [pdb]" are 0.48 and 0.29, respectively.
[The coefficient for perfect correlation is 1, for perfect anti-correlation is -1, and for complete uncorrelated is 0]

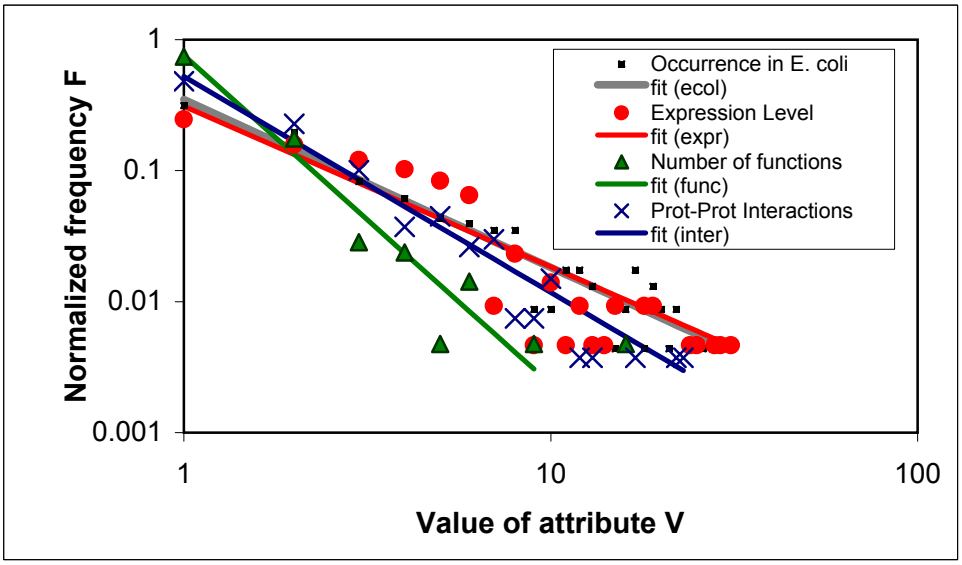
Mouse-over the points to show information on folds.

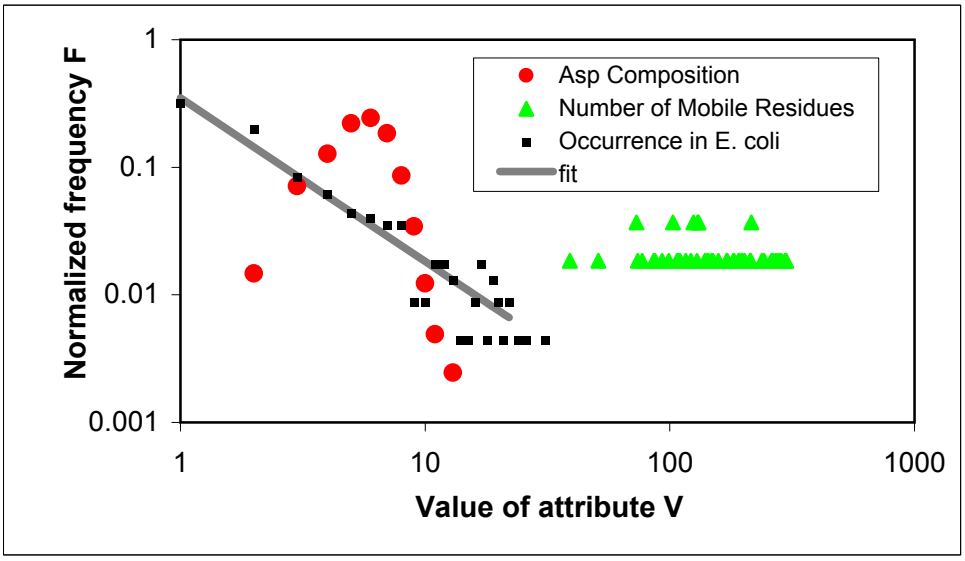


Done Internet


Report on Structure d1ama - Microsoft Internet Explorer
File Edit View Favorites Tools Help
YALE GERSTEIN LAB
search
GO
Report on PDB Structure 1ama
Genome Occurrence.
Domain d1ama__.
No additional domains.
Genome Aful Mjan Mthe Phor Scer Cele Aaeo Syne Ecol Bsub Mtub Hinf Hpyl Mgen Mpne Bbur Tpal Ctra Cpne Rpro
Fold 18 16 15 18 32 40 19 23 36 43 33 15 9 2 2 3 5 8 10 5
Superfamily 18 16 15 18 32 40 19 23 36 43 33 15 9 2 2 3 5 8 10 5
Alignments.
Domain d1ama__.
No additional domains.
Pairs untrimmed RMS trimmed RMS %sequence identity structure p-value sequence p-value
Fold 28 5.53 1.54 26.40 3.87e-23 6.42e-01
Superfamily 28 5.53 1.54 26.40 3.87e-23 6.42e-01
Family 6 2.00 0.43 48.66 1.57e-44 2.17e-68
Fold only 0 . . . . .
Sfam only 18 6.77 2.04 17.90 6.03e-23 1
Motions
Motions
Movies
Protein
Aspartate Amino Transferase (AAT)
Database ID
aat
Function Classification
Domain d1ama__.
No additional domains.
Augmented Flybase+ENZYME
GenProtEC
MIPS
d1ama__
E-2.6.1.1
3.1.04
01.01.01
Core Structure does not exist for d1ama.
Rankings
Domain d1ama__.
No additional domains.
OCCURRENCES
Species 20 Aful Mjan Mthe Phor Scer Cele Aaeo Syne Ecol Bsub Mtub Hinf Hpyl Mgen Mpne Bbu
Number 352 18 16 15 18 32 40 19 23 36 43 33 15 9 2 2 3
Rank 14 11 7 7 7 14 38 8 10 8 5 14 9 9 21 25 25
%ile 96 92 94 94 93 93 84 95 95 96 97 93 95 94 78 75 80
ALIGNMENTS
Aligned Pairs
Untrimmed RMS
Trimmed RMS
% Seq Identity
Structure P-Value
Sequence
FOLD
Internet


Normalized frequency F that this pattern is observed
100%
10%
1%
0%
1
10
100
Number times N a given fold occurs in E. coli


Normalized frequency F
1
0.1
0.01
0.001
1
10
100
Value of attribute V
Occurrence in E. coli
fit (ecol)
Expression Level
fit (expr)
Number of functions
fit (func)
Prot-Prot Interactions
fit (inter)


Normalized frequency F
1
0.1
0.01
0.001
1
10
100
1000
Value of attribute V
Asp Composition
Number of Mobile Residues
Occurrence in E. coli
fit