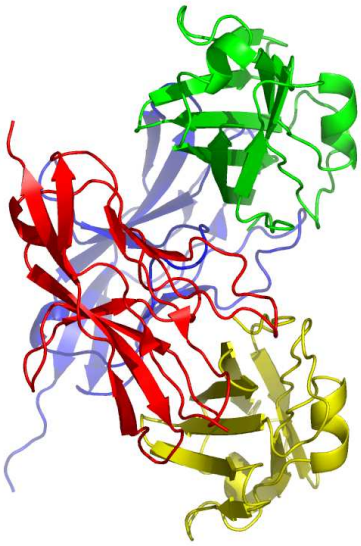
2jjs

Evolutionary trace report by **report_maker**

August 6, 2009



CONTENTS

1 INTRODUCTION

From the original Protein Data Bank entry (PDB id 2jjs):

Title: Structure of human cd47 in complex with human signal regulatory protein (sirp) alpha

Compound: Mol id: 1; molecule: tyrosine-protein phosphatase non-receptor type substrate 1; chain: a, b; fragment: n-terminal ectodomain, residues 31-148; synonym: sirp alpha, shp substrate 1, shps-1, inhibitory receptor shps-1, signal regulatory protein alpha-1, sirp-alpha-1, sirp-alpha-2, sirp-alpha- 3, myd-1 antigen, brain ig-like molecule with tyrosine- based activation motifs, bit, macrophage fusion receptor, p84, cd172a antigen; engineered: yes; mol id: 2; molecule: leukocyte surface antigen cd47; chain: c, d; fragment: immunoglobulin-superfamily ectodomain, residues 19-136; synonym: cd47, integrin-associated protein, iap, antigenic surface determinant protein oa3, protein mer6; engineered: yes; mutation: yes

Organism, scientific name: Homo Sapiens;

2jjs contains unique chains 2jjsB (119 residues) and 2jjsD (115 residues) 2jjsA is a homologue of chain 2jjsB. 2jjsC is a homologue of chain 2jjsD.

2 CHAIN 2JJSB

2.1 P78324 overview

From SwissProt, id P78324, 100% identical to 2jjsB:

Description: Tyrosine-protein phosphatase non-receptor type substrate 1 precursor (SHP substrate-1) (SHPS-1) (Inhibitory receptor SHPS-1) (Signal- regulatory protein alpha-1) (Sirp-alpha-1) (Sirp-alpha-2) (Sirp-alpha- 3) (MyD-1 antigen) (Brain Ig-like molecule

with tyrosine-based activation motifs) (Bit) (Macrophage fusion receptor) (p84).
Organism, scientific name: Homo sapiens (Human).
Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
Function: Immunoglobulin-like cell surface receptor for CD47. Acts as docking protein and induces translocation of PTPN6, PTPN11 and other binding partners from the cytosol to the plasma membrane. Supports adhesion of cerebellar neurons, neurite outgrowth and glial cell attachment. May play a key role in intracellular signaling during synaptogenesis and in synaptic function (By similarity). Involved in the negative regulation of receptor tyrosine kinase-coupled cellular responses induced by cell adhesion, growth factors or insulin. Mediates negative regulation of phagocytosis, mast cell activation and dendritic cell activation. CD47 binding prevents maturation of immature dendritic cells and inhibits cytokine production by mature dendritic cells.
Subunit: Binds PTPN11 when tyrosine-phosphorylated, except in macrophages, where it primarily binds PTPN6. Binds GRB2 in vitro. Binds FGR (By similarity). Binds JAK2 irrespective of its phosphorylation status and forms a stable complex. Binds SCAP1 and/or SCAP2. The resulting complex recruits FYB. Binds PTK2B.
Subcellular location: Type I membrane protein.
Alternative products:

Event=Alternative splicing; Named isoforms=3; Name=1; IsoId=P78324-1; Sequence=Displayed; Name=2; IsoId=P78324-2; Sequence=VSP 007030; Note=No experimental confirmation available; Name=3; IsoId=P78324-3; Sequence=VSP 007029; Note=No experimental confirmation available;

Tissue specificity: Ubiquitous. Highly expressed in brain. Detected on myeloid cells, but not T cells. Detected at lower levels in heart, placenta, lung, testis, ovary, colon, liver, small intestine, prostate, spleen, kidney, skeletal muscle and pancreas.
Ptm: N-glycosylated.
Ptm: Phosphorylated on tyrosine residues in response to stimulation with EGF, growth hormone, insulin and PDGF. Dephosphorylated by PTPN11.
Similarity: Contains 2 Ig-like C1-type (immunoglobulin-like) domains.
Similarity: Contains 1 Ig-like V-type (immunoglobulin-like) domain.
About: This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

2.2 Multiple sequence alignment for 2jjsB

For the chain 2jjsB, the alignment 2jjsB.msf (attached) with 112 sequences was used. The alignment was downloaded from the HSSP database, and fragments shorter than 75% of the query as well as duplicate sequences were removed. It can be found in the attachment to this report, under the name of 2jjsB.msf. Its statistics, from the *alistat* program are the following:

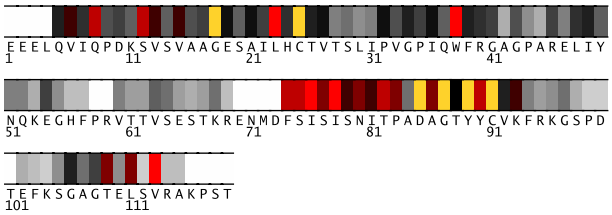


Fig. 1. Residues 1-119 in 2jjsB colored by their relative importance. (See Appendix, Fig.12, for the coloring scheme.)

```
Format:                MSF
Number of sequences: 112
Total number of residues:    11669
Smallest:              47
Largest:               119
Average length:        104.2
Alignment length:      119
Average identity:      36%
Most related pair:     98%
Most unrelated pair: 10%
Most distant seq:      30%
```

Furthermore, 1% of residues show as conserved in this alignment.

The alignment consists of 64% eukaryotic (61% vertebrata, 1% arthropoda) sequences. (Descriptions of some sequences were not readily available.) The file containing the sequence descriptions can be found in the attachment, under the name 2jjsB.descr.

2.3 Residue ranking in 2jjsB

The 2jjsB sequence is shown in Fig. 1, with each residue colored according to its estimated importance. The full listing of residues in 2jjsB can be found in the file called 2jjsB.ranks_sorted in the attachment.

2.4 Top ranking residues in 2jjsB and their position on the structure

In the following we consider residues ranking among top 25% of residues in the protein . Figure 2 shows residues in 2jjsB colored by their importance: bright red and yellow indicate more conserved/important residues (see Appendix for the coloring scheme). A Pymol script for producing this figure can be found in the attachment.

2.4.1 Clustering of residues at 25% coverage. Fig. 3 shows the top 25% of all residues, this time colored according to clusters they belong to. The clusters in Fig.3 are composed of the residues listed in Table 1.

Table 1.

cluster color	size	member residues
red	29	6,8,12,13,15,18,23,25,38,74 75,76,77,78,79,80,81,82,83 85,86,87,88,89,90,91,109,111 113

Table 1. Clusters of top ranking residues in 2jjsB.

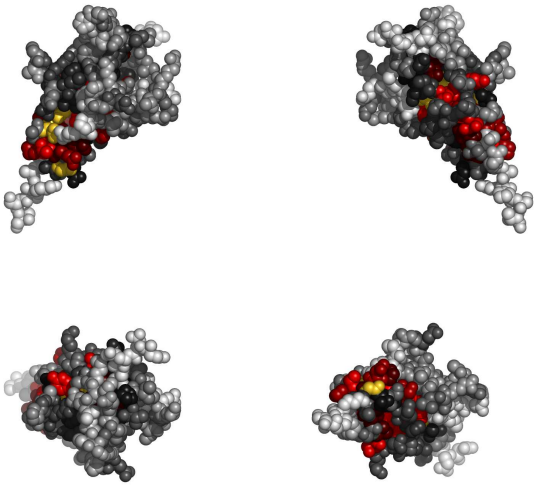


Fig. 2. Residues in 2jjsB, colored by their relative importance. Clockwise: front, back, top and bottom views.

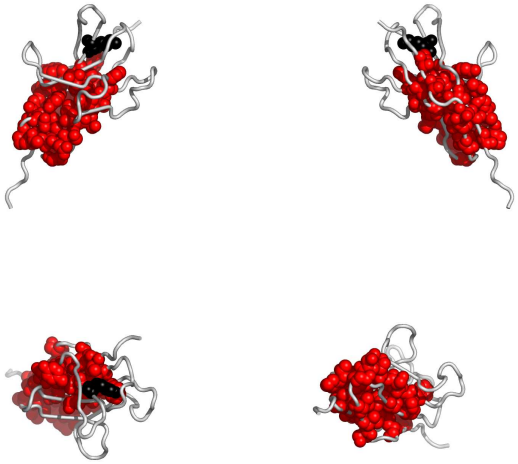


Fig. 3. Residues in 2jjsB, colored according to the cluster they belong to: red, followed by blue and yellow are the largest clusters (see Appendix for the coloring scheme). Clockwise: front, back, top and bottom views. The corresponding Pymol script is attached.

2.4.2 Overlap with known functional surfaces at 25% coverage. The name of the ligand is composed of the source PDB identifier and the heteroatom name used in that file.

Iodide ion binding site. Table 2 lists the top 25% of residues at the interface with 2jjsBIOD1124 (iodide ion). The following table (Table 3) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 2.

res	type	subst's (%)	cvg	noc/ bb	dist (Å)
87	G	G(91) A(8)	0.05	2/2	3.81
88	T	T(79) S(3) E(1) V(3) I(5) R(1)MWL DQ	0.25	6/4	3.39

Table 2. The top 25% of residues in 2jjsB at the interface with iodide ion.(Field names: res: residue number in the PDB entry; type: amino acid type; substs: substitutions seen in the alignment; with the percentage of each type in the bracket; noc/bb: number of contacts with the ligand, with the number of contacts realized through backbone atoms given in the bracket; dist: distance of closest apporach to the ligand.)

Table 3.

res	type	disruptive mutations
87	G	(KER)(QHD)(FYMW)(N)
88	T	(R)(K)(H)(FW)

Table 3. List of disruptive mutations for the top 25% of residues in 2jjsB, that are at the interface with iodide ion.

Figure 4 shows residues in 2jjsB colored by their importance, at the interface with 2jjsBIOD1124.

Interface with 2jjsC. Table 4 lists the top 25% of residues at the interface with 2jjsC. The following table (Table 5) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 4.

res	type	subst's (%)	cvg	noc/ bb	dist (Å)	antn
77	S	R(62) .(2) T(8) S(8) K(2) L(1) E(6) H(2)Q C(2)M	0.16	1/0	4.94	site

Table 4. The top 25% of residues in 2jjsB at the interface with 2jjsC. (Field names: res: residue number in the PDB entry; type: amino acid type; substs: substitutions seen in the alignment; with the percentage of each type in the bracket; noc/bb: number of contacts with the ligand, with the number of contacts realized through backbone atoms given in the bracket; dist: distance of closest apporach to the ligand.)

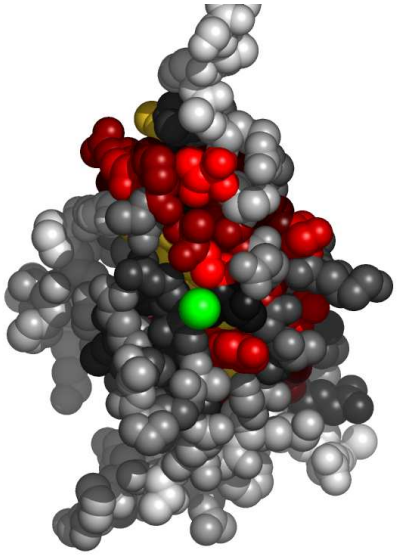


Fig. 4. Residues in 2jjsB, at the interface with iodide ion, colored by their relative importance. The ligand (iodide ion) is colored green. Atoms further than 30$\mathring{A}$ away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme for the protein chain 2jjsB.)

Table 5.		
res	**type**	**disruptive mutations**
77	S	(R)(FW)(KH)(Y)

Table 5. List of disruptive mutations for the top 25% of residues in 2jjsB, that are at the interface with 2jjsC.

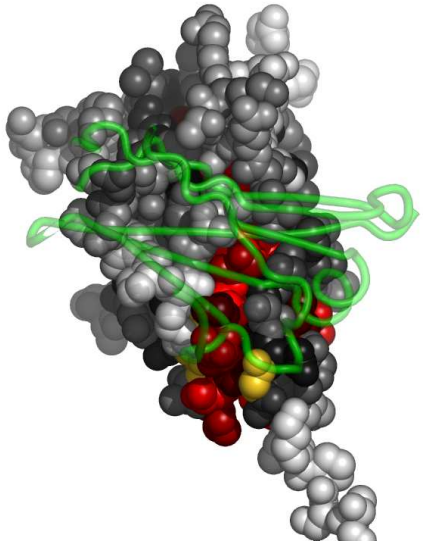


Fig. 5. Residues in 2jjsB, at the interface with 2jjsC, colored by their relative importance. 2jjsC is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 2jjsB.)

Figure 5 shows residues in 2jjsB colored by their importance, at the interface with 2jjsC.

Interface with 2jjsD. Table 6 lists the top 25% of residues at the interface with 2jjsD. The following table (Table 7) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 6.					
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist ($\mathring{A}$)**
74	F	F(73) .(4) K(2) G(4) Y(5)S L(5) A(1) V(1)	0.13	4/0	3.52
93	K	V(26) A(20) Q(1)	0.20	7/0	3.13

continued in next column

Table 6. *continued*					
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist (Å)**
		K(32) I(3) .(8) L(2)Y G(2)M			

Table 6. The top 25% of residues in 2jjsB at the interface with 2jjsD. (Field names: res: residue number in the PDB entry; type: amino acid type; substs: substitutions seen in the alignment; with the percentage of each type in the bracket; noc/bb: number of contacts with the ligand, with the number of contacts realized through backbone atoms given in the bracket; dist: distance of closest apporach to the ligand.)

Table 7.		
res	**type**	**disruptive mutations**
74	F	(K)(E)(Q)(D)
93	K	(Y)(T)(FW)(S)

Table 7. List of disruptive mutations for the top 25% of residues in 2jjsB, that are at the interface with 2jjsD.

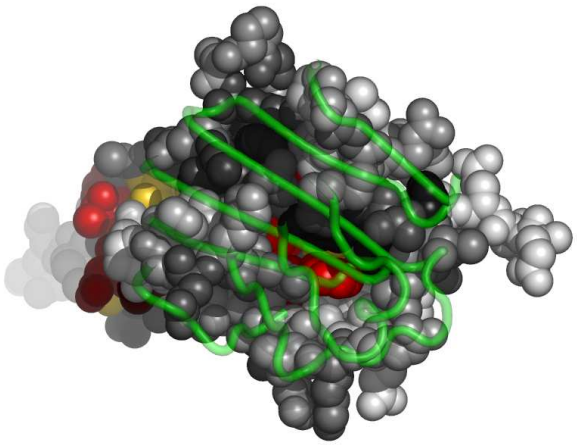


Fig. 6. Residues in 2jjsB, at the interface with 2jjsD, colored by their relative importance. 2jjsD is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 2jjsB.)

Figure 6 shows residues in 2jjsB colored by their importance, at the interface with 2jjsD.

NAG binding site. Table 8 lists the top 25% of residues at the interface with 2jjsBNAG1120 (nag). The following table (Table 9) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 8.						
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist (Å)**	**antn**
78	I	I(89) .(2) L(2)VMS K(2)	0.08	8/8	3.92	
77	S	R(62) .(2) T(8) S(8) K(2) L(1) E(6) H(2)Q C(2)M	0.16	21/11	2.56	site
79	S	N(36) .(2) S(37) L(6) H(1) K(3) Q(4)M E(1)G R(3)	0.23	27/9	3.50	site

Table 8. The top 25% of residues in 2jjsB at the interface with NAG.(Field names: res: residue number in the PDB entry; type: amino acid type; substs: substitutions seen in the alignment; with the percentage of each type in the bracket; noc/bb: number of contacts with the ligand, with the number of contacts realized through backbone atoms given in the bracket; dist: distance of closest apporach to the ligand.)

Table 9.		
res	**type**	**disruptive mutations**
78	I	(Y)(R)(H)(T)
77	S	(R)(FW)(KH)(Y)
79	S	(R)(FW)(H)(Y)

Table 9. List of disruptive mutations for the top 25% of residues in 2jjsB, that are at the interface with NAG.

Figure 7 shows residues in 2jjsB colored by their importance, at the interface with 2jjsBNAG1120.

2.4.3 Possible novel functional surfaces at 25% coverage. One group of residues is conserved on the 2jjsB surface, away from (or susbtantially larger than) other functional sites and interfaces recognizable in PDB entry 2jjs. It is shown in Fig. 8. The right panel shows (in blue) the rest of the larger cluster this surface belongs to. The residues belonging to this surface ”patch” are listed in Table 10, while Table 11 suggests possible disruptive replacements for these residues (see Section 4.6).

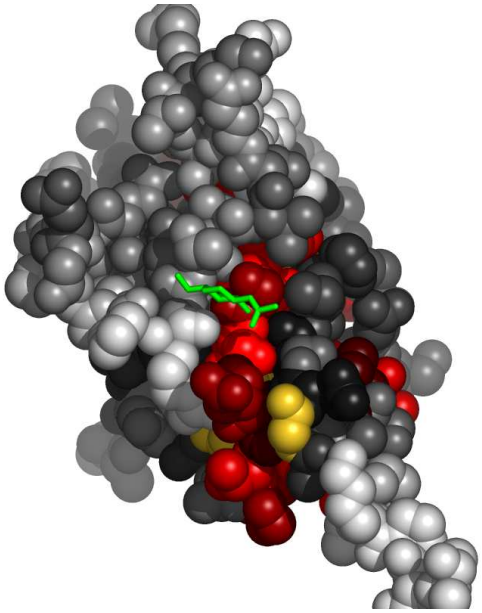


Fig. 7. Residues in 2jjsB, at the interface with NAG, colored by their relative importance. The ligand (NAG) is colored green. Atoms further than 30Å away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme for the protein chain 2jjsB.)

Fig. 8. A possible active surface on the chain 2jjsB. The larger cluster it belongs to is shown in blue.

Table 10.

res	type	substitutions(%)	cvg	antn
18	G	G(97).(2)	0.03	
85	D	D(96)K(2)H	0.03	
87	G	G(91)A(8)	0.05	
78	I	I(89).(2)L(2)VM SK(2)	0.08	
113	V	V(88).(6)I(1)M L(1)R	0.08	
38	W	W(83).(7)F(4) Q(1)Y(1)S	0.09	
90	Y	R(35)T(4)F(8)	0.11	

continued in next column

Table 10. *continued*

res	type	substitutions(%)	cvg	antn
		Y(40)E(1)K(1)Q I(3)V(2)G		
12	S	S(71)K(2)V(3) N(2)T(2)L(3) R(2)A.(2)P(4)Y D(1)	0.12	
74	F	F(73).(4)K(2) G(4)Y(5)SL(5) A(1)V(1)	0.13	
82	T	T(69).(2)S(8) K(3)RVQ(8)INEFL H	0.13	
8	Q	Q(75)R(2).(4)Y H(2)T(1)V(1) E(1)G(3)FK(1)AL	0.14	
86	A	G(33)Q(2)S(19) A(32)V(3)T(3) E(2)PY	0.15	
77	S	R(62).(2)T(8) S(8)K(2)L(1) E(6)H(2)QC(2)M	0.16	site
83	P	V(34).(6)I(1) E(2)P(36)R(1) S(1)TA(4)L(8)F	0.17	
80	N	D(40).(2)N(40) H(1)Q(1)S(2)G R(2)IK(1)E(3)Y	0.18	
93	K	V(26)A(20)Q(1) K(32)I(3).(8) L(2)YG(2)M	0.20	
81	I	L(39).(2)V(27) I(22)S(2)M(1) A(1)T(1)	0.21	
13	V	V(62)Q(6)F(5) K(4)E(3)L(5) I(6).(2)M(1) D(1)	0.22	
15	V	K(36)V(43)L(3) G(1)A(7)SI(1) .(2)M(1)	0.23	
79	S	N(36).(2)S(37) L(6)H(1)K(3) Q(4)ME(1)GR(3)	0.23	site
6	V	V(72)F(2).(11) Q(1)I(3)AL(3) M(1)EY	0.24	
88	T	T(79)S(3)E(1) V(3)I(5)R(1)MWL DQ	0.25	

Table 10. Residues forming surface ”patch” in 2jjsB.

Table 11.		
res	**type**	**disruptive mutations**
18	G	(KER)(FQMWHD)(NLPI)(Y)
85	D	(FWR)(VCAG)(Y)(TH)
87	G	(KER)(QHD)(FYMW)(N)
78	I	(Y)(R)(H)(T)
113	V	(Y)(ER)(H)(K)
38	W	(K)(E)(D)(Q)
90	Y	(K)(M)(Q)(ER)
12	S	(R)(K)(H)(FW)
74	F	(K)(E)(Q)(D)
82	T	(R)(K)(H)(FW)
8	Q	(Y)(H)(T)(FW)
86	A	(R)(K)(Y)(E)
77	S	(R)(FW)(KH)(Y)
83	P	(R)(Y)(H)(T)
80	N	(Y)(FW)(T)(H)
93	K	(Y)(T)(FW)(S)
81	I	(R)(Y)(H)(K)
13	V	(Y)(R)(H)(K)
15	V	(Y)(R)(E)(K)
79	S	(R)(FW)(H)(Y)
6	V	(R)(Y)(K)(E)
88	T	(R)(K)(H)(FW)

Table 11. Disruptive mutations for the surface patch in 2jjsB.

3 CHAIN 2JJSD

3.1 Q5REL0 overview

From SwissProt, id Q5REL0, 99% identical to 2jjsD:
Description: Hypothetical protein DKFZp459F197.
Organism, scientific name: Pongo pygmaeus (Orangutan).
Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pongo.

3.2 Multiple sequence alignment for 2jjsD

For the chain 2jjsD, the alignment 2jjsD.msf (attached) with 6 sequences was used. The alignment was assembled through combination of BLAST searching on the UniProt database and alignment using Muscle program. It can be found in the attachment to this report, under the name of 2jjsD.msf. Its statistics, from the *alistat* program are the following:

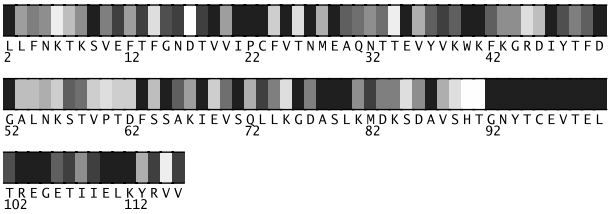


Fig. 9. Residues 2-116 in 2jjsD colored by their relative importance. (See Appendix, Fig.12, for the coloring scheme.)

```
Format:              MSF
Number of sequences: 6
Total number of residues:    679
Smallest:            112
Largest:             115
Average length:      113.2
Alignment length:    115
Average identity:    60%
Most related pair:   84%
Most unrelated pair: 50%
Most distant seq:    52%
```

Furthermore, 34% of residues show as conserved in this alignment.

The alignment consists of 83% eukaryotic (83% vertebrata) sequences. (Descriptions of some sequences were not readily available.) The file containing the sequence descriptions can be found in the attachment, under the name 2jjsD.descr.

3.3 Residue ranking in 2jjsD

The 2jjsD sequence is shown in Fig. 9, with each residue colored according to its estimated importance. The full listing of residues in 2jjsD can be found in the file called 2jjsD.ranks_sorted in the attachment.

3.4 Top ranking residues in 2jjsD and their position on the structure

In the following we consider residues ranking among top 35% of residues in the protein (the closest this analysis allows us to get to 25%). Figure 10 shows residues in 2jjsD colored by their importance: bright red and yellow indicate more conserved/important residues (see Appendix for the coloring scheme). A Pymol script for producing this figure can be found in the attachment.

3.4.1 Clustering of residues at 35% coverage. Fig. 11 shows the top 35% of all residues, this time colored according to clusters they belong to. The clusters in Fig.11 are composed of the residues listed in Table 12.

Table 12.		
cluster color	**size**	**member residues**
red	37	2,9,19,21,22,23,25,27,29,35 40,41,47,63,65,68,76,78,79 80,87,92,93,94,95,96,97,98 99,100,101,103,104,105,110
		continued in next column

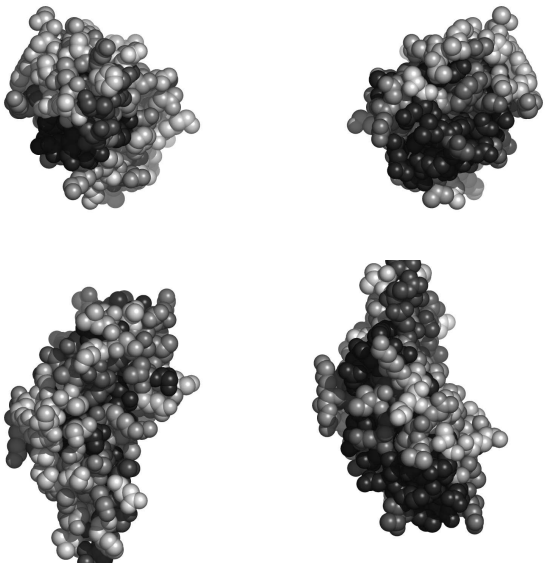


Fig. 10. Residues in 2jjsD, colored by their relative importance. Clockwise: front, back, top and bottom views.

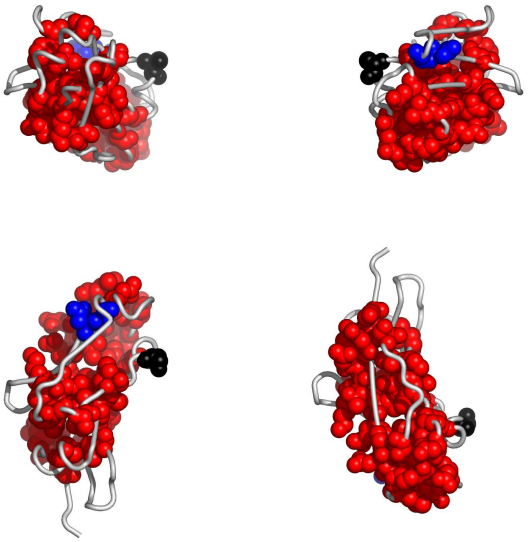


Fig. 11. Residues in 2jjsD, colored according to the cluster they belong to: red, followed by blue and yellow are the largest clusters (see Appendix for the coloring scheme). Clockwise: front, back, top and bottom views. The corresponding Pymol script is attached.

Table 12. *continued*

cluster color	size	member residues
blue	2	111,112 51,52

Table 12. Clusters of top ranking residues in 2jjsD.

4 NOTES ON USING TRACE RESULTS

4.1 Coverage

Trace results are commonly expressed in terms of coverage: the residue is important if its "coverage" is small - that is if it belongs to some small top percentage of residues [100% is all of the residues in a chain], according to trace. The ET results are presented in the form of a table, usually limited to top 25% percent of residues (or to some nearby percentage), sorted by the strength of the presumed evolutionary pressure. (I.e., the smaller the coverage, the stronger the pressure on the residue.) Starting from the top of that list, mutating a couple of residues should affect the protein somehow, with the exact effects to be determined experimentally.

4.2 Known substitutions

One of the table columns is "substitutions" - other amino acid types seen at the same position in the alignment. These amino acid types may be interchangeable at that position in the protein, so if one wants to affect the protein by a point mutation, they should be avoided. For example if the substitutions are "RVK" and the original protein has an R at that position, it is advisable to try anything, but RVK. Conversely, when looking for substitutions which will *n*ot affect the protein, one may try replacing, R with K, or (perhaps more surprisingly), with V. The percentage of times the substitution appears in the alignment is given in the immediately following bracket. No percentage is given in the cases when it is smaller than 1%. This is meant to be a rough guide - due to rounding errors these percentages often do not add up to 100%.

4.3 Surface

To detect candidates for novel functional interfaces, first we look for residues that are solvent accessible (according to DSSP program) by at least $10\mathring{A}^2$, which is roughly the area needed for one water molecule to come in the contact with the residue. Furthermore, we require that these residues form a "cluster" of residues which have neighbor within $5\mathring{A}$ from any of their heavy atoms.

Note, however, that, if our picture of protein evolution is correct, the neighboring residues which *are not* surface accessible might be equally important in maintaining the interaction specificity - they should not be automatically dropped from consideration when choosing the set for mutagenesis. (Especially if they form a cluster with the surface residues.)

4.4 Number of contacts

Another column worth noting is denoted "noc/bb"; it tells the number of contacts heavy atoms of the residue in question make across the interface, as well as how many of them are realized through the backbone atoms (if all or most contacts are through the backbone, mutation presumably won't have strong impact). Two heavy atoms are considered to be "in contact" if their centers are closer than $5\mathring{A}$.

4.5 Annotation

If the residue annotation is available (either from the pdb file or from other sources), another column, with the header "annotation" appears. Annotations carried over from PDB are the following: site (indicating existence of related site record in PDB), S-S (disulfide

bond forming residue), hb (hydrogen bond forming residue, jb (james bond forming residue), and sb (for salt bridge forming residue).

4.6 Mutation suggestions

Mutation suggestions are completely heuristic and based on complementarity with the substitutions found in the alignment. Note that they are meant to be **disruptive** to the interaction of the protein with its ligand. The attempt is made to complement the following properties: small $[AVGSTC]$, medium $[LPNQDEMIK]$, large $[WFYHR]$, hydrophobic $[LPVAMWFI]$, polar $[GTCY]$; positively $[KHR]$, or negatively $[DE]$ charged, aromatic $[WFYH]$, long aliphatic chain $[EKRQM]$, OH-group possession $[SDETY]$, and NH2 group possession $[NQRK]$. The suggestions are listed according to how different they appear to be from the original amino acid, and they are grouped in round brackets if they appear equally disruptive. From left to right, each bracketed group of amino acid types resembles more strongly the original (i.e. is, presumably, less disruptive) These suggestions are tentative - they might prove disruptive to the fold rather than to the interaction. Many researcher will choose, however, the straightforward alanine mutations, especially in the beginning stages of their investigation.

5 APPENDIX

5.1 File formats

Files with extension "ranks_sorted" are the actual trace results. The fields in the table in this file:

- `alignment#` number of the position in the alignment
- `residue#` residue number in the PDB file
- `type` amino acid type
- `rank` rank of the position according to older version of ET
- `variability` has two subfields:
  1. number of different amino acids appearing in in this column of the alignment
  2. their type
- `rho` ET score - the smaller this value, the lesser variability of this position across the branches of the tree (and, presumably, the greater the importance for the protein)
- `cvg` coverage - percentage of the residues on the structure which have this rho or smaller
- `gaps` percentage of gaps in this column

5.2 Color schemes used

The following color scheme is used in figures with residues colored by cluster size: black is a single-residue cluster; clusters composed of more than one residue colored according to this hierarchy (ordered by descending size): red, blue, yellow, green, purple, azure, turquoise, brown, coral, magenta, LightSalmon, SkyBlue, violet, gold, bisque, LightSlateBlue, orchid, RosyBrown, MediumAquamarine, DarkOliveGreen, CornflowerBlue, grey55, burlywood, LimeGreen, tan, DarkOrange, DeepPink, maroon, BlanchedAlmond.

The colors used to distinguish the residues by the estimated evolutionary pressure they experience can be seen in Fig. 12.

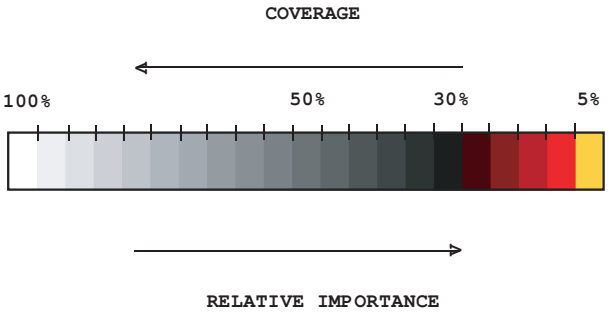


Fig. 12. Coloring scheme used to color residues by their relative importance.

5.3 Credits

5.3.1 **Alistat** *alistat* reads a multiple sequence alignment from the file and shows a number of simple statistics about it. These statistics include the format, the number of sequences, the total number of residues, the average and range of the sequence lengths, and the alignment length (e.g. including gap characters). Also shown are some percent identities. A percent pairwise alignment identity is defined as (idents / MIN(len1, len2)) where idents is the number of exact identities and len1, len2 are the unaligned lengths of the two sequences. The "average percent identity", "most related pair", and "most unrelated pair" of the alignment are the average, maximum, and minimum of all (N)(N-1)/2 pairs, respectively. The "most distant seq" is calculated by finding the maximum pairwise identity (best relative) for all N sequences, then finding the minimum of these N numbers (hence, the most outlying sequence). *alistat* is copyrighted by HHMI/Washington University School of Medicine, 1992-2001, and freely distributed under the GNU General Public License.

5.3.2 **CE** To map ligand binding sites from different source structures, report_maker uses the CE program: `http://cl.sdsc.edu/`. Shindyalov IN, Bourne PE (1998) *"Protein structure alignment by incremental combinatorial extension (CE) of the optimal path* . Protein Engineering 11(9) 739-747.

5.3.3 **DSSP** In this work a residue is considered solvent accessible if the DSSP program finds it exposed to water by at least 10Å^2, which is roughly the area needed for one water molecule to come in the contact with the residue. DSSP is copyrighted by W. Kabsch, C. Sander and MPI-MF, 1983, 1985, 1988, 1994 1995, CMBI version by Elmar.Krieger@`cmbi.kun.nl` November 18,2002,

`http://www.cmbi.kun.nl/gv/dssp/descrip.html.`

5.3.4 **HSSP** Whenever available, report_maker uses HSSP alignment as a starting point for the analysis (sequences shorter than 75% of the query are taken out, however); R. Schneider, A. de Daruvar, and C. Sander. *"The HSSP database of protein structure-sequence alignments."* Nucleic Acids Res., 25:226–230, 1997.

`http://swift.cmbi.kun.nl/swift/hssp/`

5.3.5 **LaTex** The text for this report was processed using LaTeX; Leslie Lamport, "LaTeX: A Document Preparation System Addison-Wesley," Reading, Mass. (1986).

5.3.6 **Muscle** When making alignments "from scratch", report maker uses Muscle alignment program: Edgar, Robert C. (2004), *"MUSCLE: multiple sequence alignment with high accuracy and high throughput."* Nucleic Acids Research 32(5), 1792-97.

`http://www.drive5.com/muscle/`

5.3.7 **Pymol** The figures in this report were produced using Pymol. The scripts can be found in the attachment. Pymol is an open-source application copyrighted by DeLano Scientific LLC (2005). For more information about Pymol see `http://pymol.sourceforge.net/`. (Note for Windows users: the attached package needs to be unzipped for Pymol to read the scripts and launch the viewer.)

5.4 Note about ET Viewer

Dan Morgan from the Lichtarge lab has developed a visualization tool specifically for viewing trace results. If you are interested, please visit:

`http://mammoth.bcm.tmc.edu/traceview/`

The viewer is self-unpacking and self-installing. Input files to be used with ETV (extension .etvx) can be found in the attachment to the main report.

5.5 Citing this work

The method used to rank residues and make predictions in this report can be found in Mihalek, I., I. Reš, O. Lichtarge. (2004). *"A Family of Evolution-Entropy Hybrid Methods for Ranking of Protein Residues by Importance"* J. Mol. Bio. **336**: 1265-82. For the original version of ET see O. Lichtarge, H.Bourne and F. Cohen (1996). *"An Evolutionary Trace Method Defines Binding Surfaces Common to Protein Families"* J. Mol. Bio. **257**: 342-358.

report_maker itself is described in Mihalek I., I. Res and O. Lichtarge (2006). *"Evolutionary Trace Report Maker: a new type of service for comparative analysis of proteins."* Bioinformatics **22**:1656-7.

5.6 About report_maker

report_maker was written in 2006 by Ivana Mihalek. The 1D ranking visualization program was written by Ivica Reš. **report_maker** is copyrighted by Lichtarge Lab, Baylor College of Medicine, Houston.

5.7 Attachments

The following files should accompany this report:

- 2jjsB.complex.pdb - coordinates of 2jjsB with all of its interacting partners
- 2jjsB.etvx - ET viewer input file for 2jjsB
- 2jjsB.cluster_report.summary - Cluster report summary for 2jjsB
- 2jjsB.ranks - Ranks file in sequence order for 2jjsB
- 2jjsB.clusters - Cluster descriptions for 2jjsB
- 2jjsB.msf - the multiple sequence alignment used for the chain 2jjsB
- 2jjsB.descr - description of sequences used in 2jjsB msf
- 2jjsB.ranks_sorted - full listing of residues and their ranking for 2jjsB
- 2jjsB.2jjsBIOD1124.if.pml - Pymol script for Figure 4
- 2jjsB.cbcvg - used by other 2jjsB – related pymol scripts
- 2jjsB.2jjsC.if.pml - Pymol script for Figure 5
- 2jjsB.2jjsD.if.pml - Pymol script for Figure 6
- 2jjsB.2jjsBNAG1120.if.pml - Pymol script for Figure 7
- 2jjsD.complex.pdb - coordinates of 2jjsD with all of its interacting partners
- 2jjsD.etvx - ET viewer input file for 2jjsD
- 2jjsD.cluster_report.summary - Cluster report summary for 2jjsD
- 2jjsD.ranks - Ranks file in sequence order for 2jjsD
- 2jjsD.clusters - Cluster descriptions for 2jjsD
- 2jjsD.msf - the multiple sequence alignment used for the chain 2jjsD
- 2jjsD.descr - description of sequences used in 2jjsD msf
- 2jjsD.ranks_sorted - full listing of residues and their ranking for 2jjsD