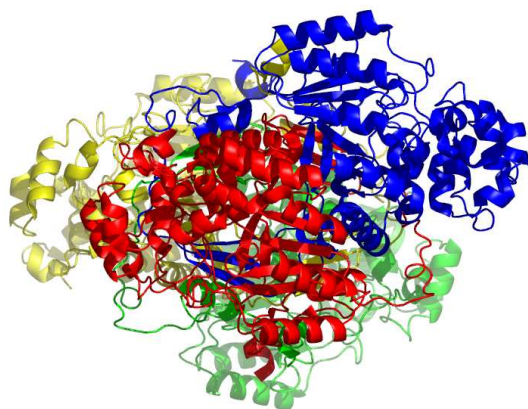


# 2a5h

Evolutionary trace report by **report\_maker**

April 22, 2010



## CONTENTS

### 1 Introduction

### 2 Chain 2a5hA

- 2.1 Q9XBQ8 overview
- 2.2 Multiple sequence alignment for 2a5hA
- 2.3 Residue ranking in 2a5hA
- 2.4 Top ranking residues in 2a5hA and their position on the structure
  - 2.4.1 Clustering of residues at 25% coverage.
  - 2.4.2 Overlap with known functional surfaces at 25% coverage.
  - 2.4.3 Possible novel functional surfaces at 25% coverage.

### 3 Notes on using trace results

- 3.1 Coverage
- 3.2 Known substitutions
- 3.3 Surface
- 3.4 Number of contacts
- 3.5 Annotation
- 3.6 Mutation suggestions

### 4 Appendix

- 4.1 File formats
- 4.2 Color schemes used
- 4.3 Credits

4.3.1	<b>Alistat</b>	12
4.3.2	<b>CE</b>	12
4.3.3	<b>DSSP</b>	12
4.3.4	<b>HSSP</b>	12
4.3.5	<b>LaTex</b>	12
4.3.6	<b>Muscle</b>	12
4.3.7	<b>Pymol</b>	12
4.4	Note about ET Viewer	12
4.5	Citing this work	12
4.6	About report_maker	12
4.7	Attachments	12

## 1 INTRODUCTION

From the original Protein Data Bank entry (PDB id 2a5h):

**Title:** 2.1 angstrom x-ray crystal structure of lysine-2,3- aminomutase from clostridium subterminale sb4, with michaelis analog (l-alpha-lysine external aldimine form of pyridoxal-5'-phosphate).

**Compound:** Mol id: 1; molecule: l-lysine 2,3-aminomutase; chain: a, b, c, d; ec: 5.4.3.2; engineered: yes

**Organism, scientific name:** Clostridium Subterminale;

2a5h contains a single unique chain 2a5hA (409 residues long) and its homologues 2a5hD, 2a5hC, and 2a5hB.

## 2 CHAIN 2A5HA

### 2.1 Q9XBQ8 overview

From SwissProt, id Q9XBQ8, 94% identical to 2a5hA:

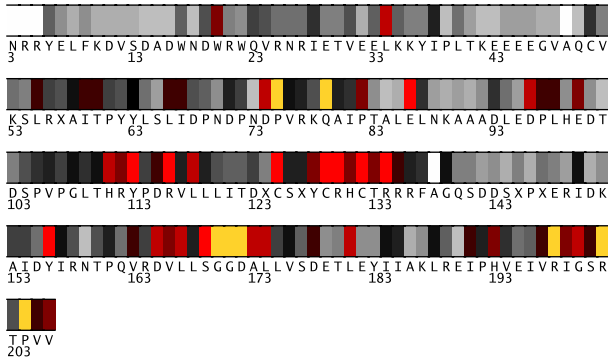
**Description:** L-lysine 2,3-aminomutase (EC 5.4.3.2).

**Organism, scientific name:** Clostridium subterminale.

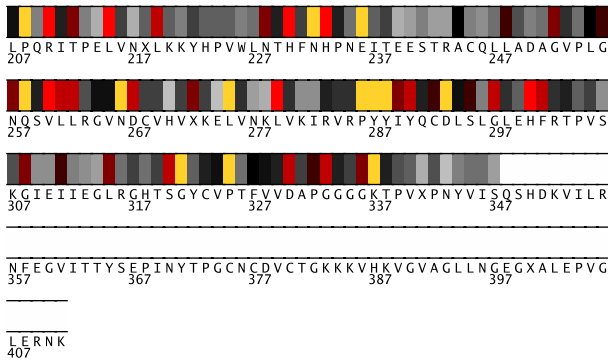
**Taxonomy:** Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.

### 2.2 Multiple sequence alignment for 2a5hA

For the chain 2a5hA, the alignment 2a5hA.msf (attached) with 43 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 2a5hA.msf. Its statistics, from the *alistat* program are the following:



**Fig. 1.** Residues 3-206 in 2a5hA colored by their relative importance. (See Appendix, Fig.13, for the coloring scheme.)



**Fig. 2.** Residues 207-411 in 2a5hA colored by their relative importance. (See Appendix, Fig.13, for the coloring scheme.)

```

Format:                MSF
Number of sequences:   43
Total number of residues: 15401
Smallest:              324
Largest:               409
Average length:        358.2
Alignment length:      409
Average identity:       35%
Most related pair:     99%
Most unrelated pair:   17%
Most distant seq:      27%

```

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

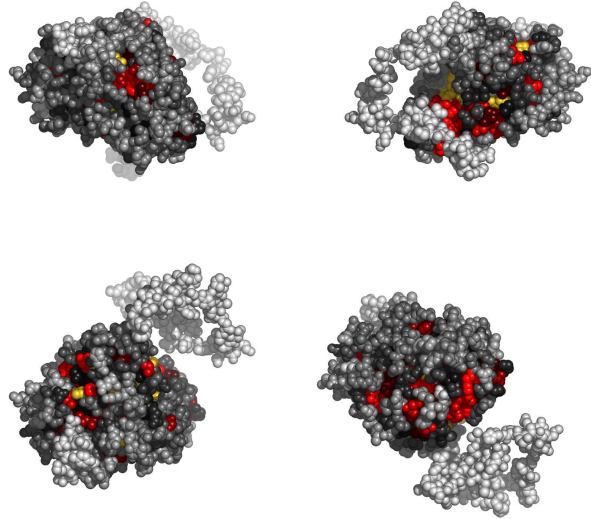
The alignment consists of 2% eukaryotic ( 2% fungi), 86% prokaryotic, and 9% archaean sequences. (Descriptions of some sequences were not readily available.) The file containing the sequence descriptions can be found in the attachment, under the name 2a5hA.descr.

### 2.3 Residue ranking in 2a5hA

The 2a5hA sequence is shown in Figs. 1–2, with each residue colored according to its estimated importance. The full listing of residues in 2a5hA can be found in the file called 2a5hA.ranks.sorted in the attachment.

### 2.4 Top ranking residues in 2a5hA and their position on the structure

In the following we consider residues ranking among top 25% of residues in the protein . Figure 3 shows residues in 2a5hA 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.

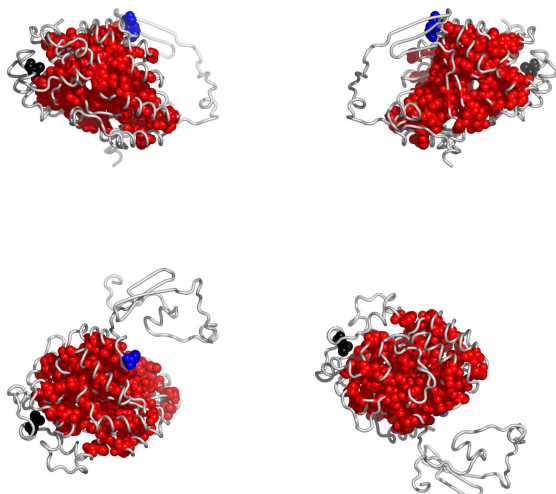


**Fig. 3.** Residues in 2a5hA, colored by their relative importance. Clockwise: front, back, top and bottom views.

2.4.1 Clustering of residues at 25% coverage. Fig. 4 shows the top 25% of all residues, this time colored according to clusters they belong to. The clusters in Fig.4 are composed of the residues listed in Table 1.

Table 1.		
cluster color	size	member residues
red	99	20, 55, 59, 60, 66, 67, 74, 75, 79, 82, 86, 96, 97, 98, 100, 111, 112, 113, 115, 116, 118, 125, 128, 129, 130, 131, 132, 133, 134, 135, 156, 163, 165, 166, 167, 169, 170, 171, 172, 173, 174, 178, 181, 191, 193, 197, 198, 199, 200, 201, 202, 204, 205, 206, 208, 210, 212, 215, 219, 228, 230, 232, 233, 236, 244, 248, 252, 256, 257, 258, 260, 261, 262, 266, 267, 272, 275, 279, 286, 287, 288, 289, 290, 292, 293, 295, 297

*continued in next column*



**Fig. 4.** Residues in 2a5hA, 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 1. continued		
cluster color	size	member residues
		300, 301, 308, 311, 315, 325, 327 330, 332, 333, 336, 337
blue	2	320, 321

**Table 1.** Clusters of top ranking residues in 2a5hA.

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.

**Interface with 2a5hD.** Table 2 lists the top 25% of residues at the interface with 2a5hD. The following table (Table 3) suggests possible disruptive replacements for these residues (see Section 3.6).

Table 2.					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
111	H	H(90) R(4) S(2) Q(2)	0.10	1/1	4.92
333	G	G(81) . (4) A(4) E(4) K(2)	0.11	7/7	3.30

continued in next column

Table 2. continued					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
112	R	Q(2) K(60) R(34) L(2)	0.15	14/8	3.61
332	P	Q(2) P(55) A(16) S(4) G(16) T(4) R(2)	0.24	10/5	3.73

**Table 2.** The top 25% of residues in 2a5hA at the interface with 2a5hD. (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 approach to the ligand.)

Table 3.		
res	type	disruptive mutations
111	H	(E)(TD)(M)(VQCLAPIG)
333	G	(FWHR)(E)(Y)(K)
112	R	(T)(Y)(D)(SCG)
332	P	(R)(Y)(H)(KE)

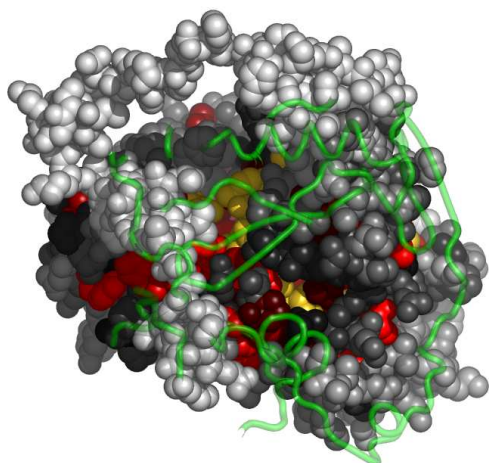
**Table 3.** List of disruptive mutations for the top 25% of residues in 2a5hA, that are at the interface with 2a5hD.

Figure 5 shows residues in 2a5hA colored by their importance, at the interface with 2a5hD.

**PLP binding site.** Table 4 lists the top 25% of residues at the interface with 2a5hAPLP419 (plp). The following table (Table 5) suggests possible disruptive replacements for these residues (see Section 3.6).

Table 4.					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
420				34/24	1.32
198	R	R(100)	0.02	19/0	2.64
288	Y	Y(100)	0.02	13/0	2.66
258	Q	Q(97) N(2)	0.03	1/0	4.74
287	Y	Y(97) G(2)	0.03	11/0	2.57
337	K	K(97) T(2)	0.03	26/0	2.91
116	R	R(90) I(2) T(4) K(2)	0.07	30/4	2.64

continued in next column



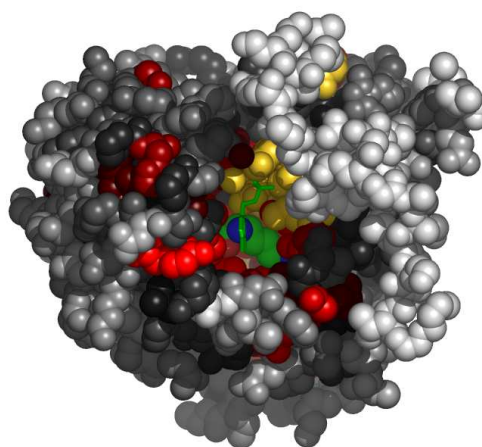
**Fig. 5.** Residues in 2a5hA, at the interface with 2a5hD, colored by their relative importance. 2a5hD is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 2a5hA.)

Table 4. continued					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
134	R	R (95) .(2) E (2)	0.07	1/0	4.42
113	Y	Y (95) H (2) K (2)	0.08	24/0	3.46
118	L	L (88) I (6) .(2) V (2)	0.11	28/7	3.69
167	L	L (32) I (62) V (4)	0.13	4/0	3.82
112	R	K (60) R (34) L (2) Q (2)	0.15	20/1	4.03
100	E	E (76) R (2) G (2) D (16) W (2)	0.19	1/0	4.67

**Table 4.** The top 25% of residues in 2a5hA at the interface with PLP. (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 approach to the ligand.)

Table 5.		
res	type	disruptive mutations
420		
198	R	(TD) (SYEVCLAPIG) (FMW) (N)
288	Y	(K) (QM) (NEVLAPIR) (D)
258	Q	(Y) (FTWH) (SVCAG) (D)
287	Y	(K) (QM) (ER) (NLPI)
337	K	(FYW) (TVAH) (SCDG) (ELPI)
116	R	(TYD) (E) (SFVCAWG) (LPI)
134	R	(T) (VCAG) (Y) (SLPDI)
113	Y	(KM) (VA) (EQ) (LPI)
118	L	(YR) (H) (T) (KE)
167	L	(YR) (H) (T) (KE)
112	R	(T) (Y) (D) (SCG)
100	E	(FWH) (R) (Y) (VCAG)

**Table 5.** List of disruptive mutations for the top 25% of residues in 2a5hA, that are at the interface with PLP.



**Fig. 6.** Residues in 2a5hA, at the interface with PLP, colored by their relative importance. The ligand (PLP) is colored green. Atoms further than 30 Å 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 2a5hA.)

Figure 6 shows residues in 2a5hA colored by their importance, at the interface with 2a5hAPLP419.

**S-adenosylmethionine binding site.** Table 6 lists the top 25% of residues at the interface with 2a5hASAM417 (s-adenosylmethionine). The following table (Table 7) suggests possible disruptive replacements for these residues (see Section 3.6).

Table 6.					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
420				67/11	3.16
170	G	G(100)	0.02	8/8	3.72
171	G	G(100)	0.02	9/9	3.18
172	D	D(100)	0.02	8/3	4.24
258	Q	Q(97)	0.03	25/9	3.04
		N(2)			
202	R	R(90)	0.05	17/9	4.01
		K(9)			
293	D	D(90)	0.05	31/10	2.88
		K(2)			
		R(4)			
		P(2)			
125	C	C(97)	0.06	1/0	4.58
		.(2)			
129	C	C(97)	0.06	1/0	4.91
		.(2)			
132	C	C(97)	0.06	12/12	3.91
		.(2)			
230	H	H(97)	0.06	48/4	2.96
		Q(2)			
134	R	R(95)	0.07	15/1	3.15
		.(2)			
		E(2)			
260	V	V(90)	0.07	56/2	3.49
		P(6)			
		T(2)			
169	S	S(83)	0.08	7/0	3.16
		T(16)			
200	G	G(30)	0.12	8/8	3.87
		H(60)			
		A(6)			
		C(2)			
290	Y	F(11)	0.14	32/4	3.45
		H(62)			
		Y(25)			
133	T	T(30)	0.15	41/12	3.08
		Q(2)			
		F(62)			
		.(2)			
		M(2)			
131	H	F(25)	0.17	20/10	2.81
		Y(51)			
		H(20)			
		.(2)			
228	N	N(23)	0.17	1/1	4.85
		S(2)			
		V(53)			
		M(4)			
		A(4)			

continued in next column

Table 6. continued					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
		H(4)			
		G(2)			
		I(4)			
292	C	A(2)	0.20	17/11	3.54
		P(6)			
		L(48)			
		E(4)			
		C(27)			
		G(4)			
		M(4)			
201	S	S(44)	0.23	14/14	4.11
		T(55)			

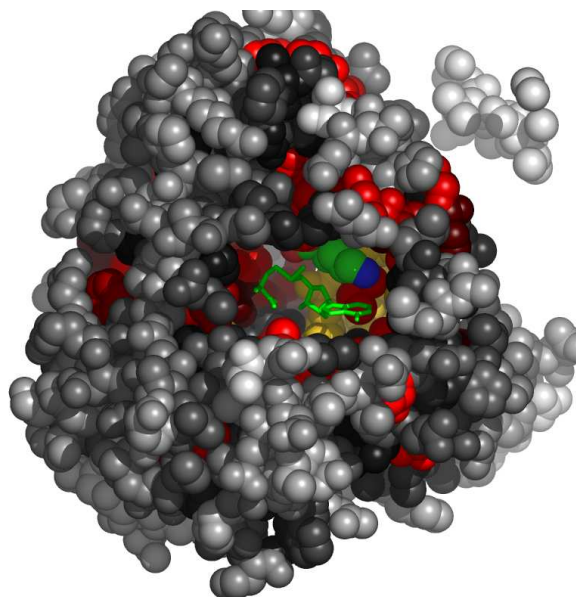
**Table 6.** The top 25% of residues in 2a5hA at the interface with s-adenosylmethionine.(Field names: res: residue number in the PDB entry; type: amino acid type; substs: substitutions seen in the alignment; with the percentage of each type in the bracket; noc/bb: number of contacts with the ligand, with the number of contacts realized through backbone atoms given in the bracket; dist: distance of closest approach to the ligand.)

Table 7.		
res	type	disruptive mutations
420		
170	G	(KER)(FQMWH)(NYLPI)(SVA)
171	G	(KER)(FQMWH)(NYLPI)(SVA)
172	D	(R)(FWH)(KYVCAG)(TQM)
258	Q	(Y)(FTWH)(SVCAG)(D)
202	R	(T)(YD)(SVCAG)(FELWPI)
293	D	(FYW)(HR)(CG)(TVA)
125	C	(KER)(FQMWH)(NLPI)(Y)
129	C	(KER)(FQMWH)(NLPI)(Y)
132	C	(KER)(FQMWH)(NLPI)(Y)
230	H	(TE)(D)(SVMCAG)(QLPI)
134	R	(T)(VCAG)(Y)(SLPDI)
260	V	(R)(K)(Y)(E)
169	S	(KR)(FQMWH)(NELPI)(Y)
200	G	(E)(K)(R)(QD)
290	Y	(K)(Q)(EM)(N)
133	T	(R)(K)(H)(FW)
131	H	(E)(Q)(MD)(TK)
228	N	(Y)(H)(R)(E)
292	C	(R)(K)(H)(E)
201	S	(KR)(FQMWH)(NELPI)(Y)

**Table 7.** List of disruptive mutations for the top 25% of residues in 2a5hA, that are at the interface with s-adenosylmethionine.

Figure 7 shows residues in 2a5hA colored by their importance, at the interface with 2a5hASAM417.

**Sulfate ion binding site.** Table 8 lists the top 25% of residues at the interface with 2a5hSO4592 (sulfate ion). The following table (Table 9) suggests possible disruptive replacements for these residues (see Section 3.6).



**Fig. 7.** Residues in 2a5hA, at the interface with s-adenosylmethionine, colored by their relative importance. The ligand (s-adenosylmethionine) 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 2a5hA.)

Table 8.					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
113	Y	Y (95) H (2) K (2)	0.08	4/4	3.38
111	H	H (90) R (4) S (2) Q (2)	0.10	30/8	2.87
112	R	K (60) R (34) L (2) Q (2)	0.15	9/9	3.79

**Table 8.** The top 25% of residues in 2a5hA at the interface with sulfate 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 approach to the ligand.)

Table 9.		
res	type	disruptive mutations
113	Y	(KM) (VA) (EQ) (LPI)
111	H	(E) (TD) (M) (VQCLAPIG)
112	R	(T) (Y) (D) (SCG)

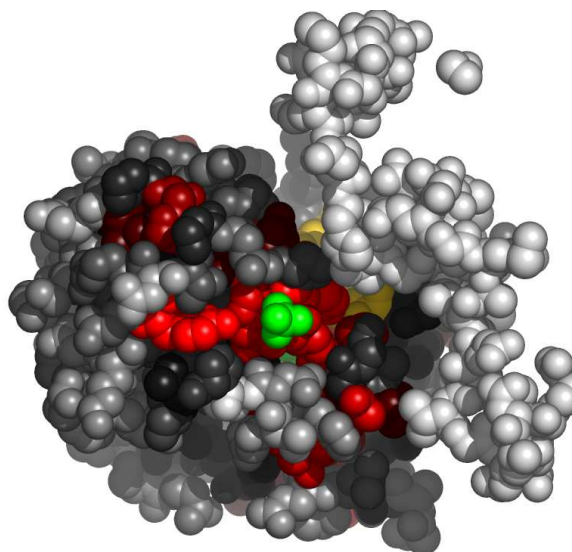
**Table 9.** List of disruptive mutations for the top 25% of residues in 2a5hA, that are at the interface with sulfate ion.

Figure 8 shows residues in 2a5hA colored by their importance, at the interface with 2a5hSO4592.

**PLP binding site.** Table 10 lists the top 25% of residues at the interface with 2a5hBPLP419 (plp). The following table (Table 11) suggests possible disruptive replacements for these residues (see Section 3.6).

Table 10.					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
321	G	G (95) S (2) N (2)	0.04	8/8	2.88
320	S	S (74) G (2) P (16) A (4) T (2)	0.14	9/7	3.36

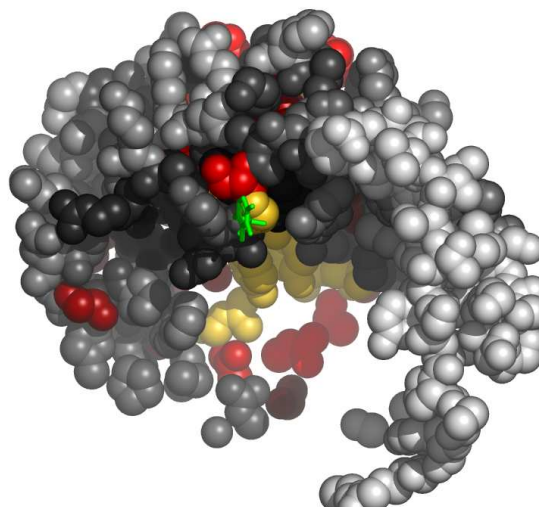
**Table 10.** The top 25% of residues in 2a5hA at the interface with PLP. (Field names: res: residue number in the PDB entry; type: amino acid type; substs: substitutions seen in the alignment; with the percentage of each type in the bracket; noc/ bb: number of contacts with the ligand, with the number of contacts realized through backbone atoms given in the bracket; dist: distance of closest approach to the ligand.)



**Fig. 8.** Residues in 2a5hA, at the interface with sulfate ion, colored by their relative importance. The ligand (sulfate ion) 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 2a5hA.)

Table 11.		
res	type	disruptive mutations
321	G	(R) (KE) (FWH) (M)
320	S	(R) (K) (H) (Q)

**Table 11.** List of disruptive mutations for the top 25% of residues in 2a5hA, that are at the interface with PLP.



**Fig. 9.** Residues in 2a5hA, at the interface with PLP, colored by their relative importance. The ligand (PLP) 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 2a5hA.)

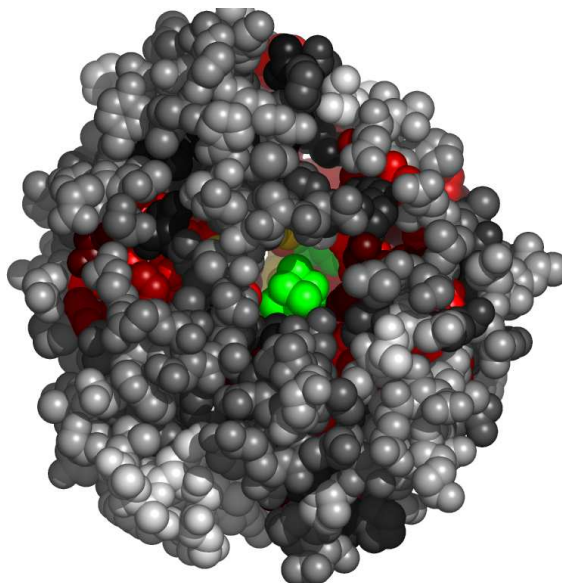
Figure 9 shows residues in 2a5hA colored by their importance, at the interface with 2a5hBPLP419.

**Iron/sulfur cluster binding site.** Table 12 lists the top 25% of residues at the interface with 2a5hASF4418 (iron/sulfur cluster). The following table (Table 13) suggests possible disruptive replacements for these residues (see Section 3.6).

Table 12.					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
170	G	G(100)	0.02	2/2	3.77
171	G	G(100)	0.02	9/9	3.62
172	D	D(100)	0.02	3/2	4.45
202	R	R(90) K(9)	0.05	6/2	3.58
125	C	C(97) . (2)	0.06	11/1	2.29

*continued in next column*

Table 12. continued					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
129	C	C (97) . (2)	0.06	17/6	2.30
132	C	C (97) . (2)	0.06	15/4	2.30
230	H	H (97) Q (2)	0.06	8/0	3.56
134	R	R (95) . (2) E (2)	0.07	1/0	4.88
128	Y	Y (48) I (2) N (32) H (13) . (2)	0.16	8/3	4.01
55	L	F (62) I (2) L (11) Y (9) M (4) H (2) R (2) Q (2) T (2)	0.23	1/0	4.91



**Fig. 10.** Residues in 2a5hA, at the interface with iron/sulfur cluster, colored by their relative importance. The ligand (iron/sulfur cluster) 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 2a5hA.)

**Table 12.** The top 25% of residues in 2a5hA at the interface with iron/sulfur cluster. (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 approach to the ligand.)

Table 13.		
res	type	disruptive mutations
170	G	(KER) (FQMWH) (NYLPI) (SVA)
171	G	(KER) (FQMWH) (NYLPI) (SVA)
172	D	(R) (FWH) (KYVCAG) (TQM)
202	R	(T) (YD) (SVCAG) (FELWPI)
125	C	(KER) (FQMWH) (NLPI) (Y)
129	C	(KER) (FQMWH) (NLPI) (Y)
132	C	(KER) (FQMWH) (NLPI) (Y)
230	H	(TE) (D) (SVMCAG) (QLPI)
134	R	(T) (VCAG) (Y) (SLPDI)
128	Y	(K) (M) (Q) (E)
55	L	(YR) (T) (E) (H)

**Table 13.** List of disruptive mutations for the top 25% of residues in 2a5hA, that are at the interface with iron/sulfur cluster.

Figure 10 shows residues in 2a5hA colored by their importance, at the interface with 2a5hASF4418.

**Interface with 2a5hB.** Table 14 lists the top 25% of residues at the interface with 2a5hB. The following table (Table 15) suggests possible disruptive replacements for these residues (see Section 3.6).



Table 14.					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
198	R	R(100)	0.02	22/0	3.40
288	Y	Y(100)	0.02	13/0	3.07
287	Y	Y(97)	0.03	42/0	3.10
337	K	G(2) K(97)	0.03	6/6	3.98
321	G	T(2) G(95)	0.04	21/21	3.25
293	D	S(2) N(2) D(90)	0.05	3/3	4.90
116	R	K(2) R(4) P(2) R(90)	0.07	63/0	2.89
113	Y	I(2) T(4) K(2) Y(95)	0.08	23/2	3.28
267	D	H(2) K(2) D(90)	0.10	42/6	2.81
165	D	R(2) N(4) S(2) E(67) D(32)	0.11	20/0	2.45
333	G	G(81) . (4) A(4) E(4) K(2) Q(2)	0.11	2/2	4.69
320	S	S(74) G(2) P(16) A(4) T(2)	0.14	48/22	2.59
112	R	K(60) R(34) L(2) Q(2)	0.15	1/1	4.67
252	G	G(79) N(9) . (2) A(6) K(2)	0.15	35/35	2.90
336	G	G(51) S(37) H(2)	0.16	1/1	4.62

continued in next column

Table 14. continued					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
315	L	F(2) N(6) L(72) A(6) M(6) V(6) I(6)	0.19	1/1	4.84
292	C	A(2) P(6) L(48) E(4) C(27) G(4) M(4)	0.20	11/10	2.69
115	D	D(39) T(4) S(4) E(4) N(27) G(18)	0.23	38/2	2.63
332	P	P(55) A(16) S(4) G(16) T(4) R(2)	0.24	26/10	2.67

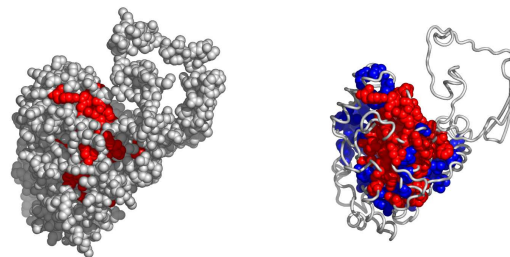
**Table 14.** The top 25% of residues in 2a5hA at the interface with 2a5hB. (Field names: res: residue number in the PDB entry; type: amino acid type; substs: substitutions seen in the alignment; with the percentage of each type in the bracket; noc/bb: number of contacts with the ligand, with the number of contacts realized through backbone atoms given in the bracket; dist: distance of closest approach to the ligand.)

Table 15.		
res	type	disruptive mutations
198	R	(TD) (SYEVCLAPIG) (FMW) (N)
288	Y	(K) (QM) (NEVLAPIR) (D)
287	Y	(K) (QM) (ER) (NLPI)
337	K	(FYW) (TVAH) (SCDG) (ELPI)
321	G	(R) (KE) (FWH) (M)
293	D	(FYW) (HR) (CG) (TVA)
116	R	(TYD) (E) (SFVCAWG) (LPI)
113	Y	(KM) (VA) (EQ) (LPI)
267	D	(FWR) (H) (Y) (VCAG)
165	D	(R) (FWH) (YVCAG) (K)
333	G	(FWHR) (E) (Y) (K)
320	S	(R) (K) (H) (Q)
112	R	(T) (Y) (D) (SCG)
252	G	(E) (R) (FWH) (YD)
336	G	(E) (K) (R) (QD)

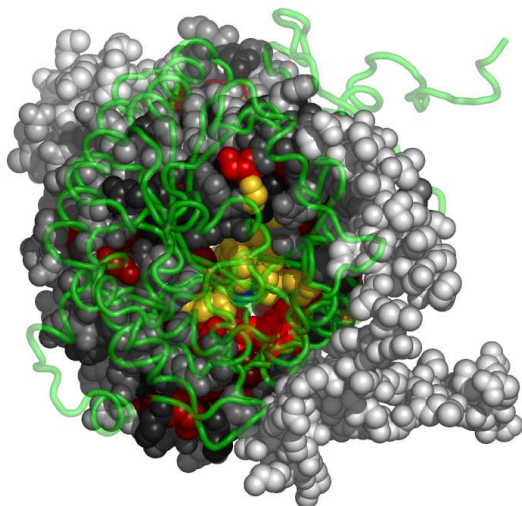
continued in next column

res	type	disruptive mutations
315	L	(Y) (R) (H) (T)
292	C	(R) (K) (H) (E)
115	D	(R) (FWH) (K) (Y)
332	P	(R) (Y) (H) (KE)

**Table 15.** List of disruptive mutations for the top 25% of residues in 2a5hA, that are at the interface with 2a5hB.



**Fig. 12.** A possible active surface on the chain 2a5hA. The larger cluster it belongs to is shown in blue.



**Fig. 11.** Residues in 2a5hA, at the interface with 2a5hB, colored by their relative importance. 2a5hB is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 2a5hA.)

Figure 11 shows residues in 2a5hA colored by their importance, at the interface with 2a5hB.

**2.4.3 Possible novel functional surfaces at 25% coverage.** One group of residues is conserved on the 2a5hA surface, away from (or substantially larger than) other functional sites and interfaces recognizable in PDB entry 2a5h. It is shown in Fig. 12. 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 16, while Table 17 suggests possible disruptive replacements for these residues (see Section 3.6).

res	type	substitutions(%)	cvg
170	G	G(100)	0.02
171	G	G(100)	0.02
172	D	D(100)	0.02
198	R	R(100)	0.02

*continued in next column*

res	type	substitutions(%)	cvg
286	P	P(100)	0.02
288	Y	Y(100)	0.02
258	Q	Q(97)N(2)	0.03
287	Y	Y(97)G(2)	0.03
337	K	K(97)T(2)	0.03
325	P	P(93)D(2)T(4)	0.04
202	R	R(90)K(9)	0.05
293	D	D(90)K(2)R(4) P(2)	0.05
125	C	C(97).(2)	0.06
129	C	C(97).(2)	0.06
132	C	C(97).(2)	0.06
230	H	H(97)Q(2)	0.06
116	R	R(90)I(2)T(4) K(2)	0.07
134	R	R(95).(2)E(2)	0.07
260	V	V(90)P(6)T(2)	0.07
113	Y	Y(95)H(2)K(2)	0.08
169	S	S(83)T(16)	0.08
86	E	E(88)K(2)M(4) I(2)D(2)	0.09
156	Y	Y(90)W(4)L(2) F(2)	0.09
300	H	H(86)D(2)Y(4) S(4)V(2)	0.09
111	H	H(90)R(4)S(2) Q(2)	0.10
118	L	L(88)I(6).(2) V(2)	0.11
165	D	E(67)D(32)	0.11
333	G	G(81).(4)A(4) E(4)K(2)Q(2)	0.11
96	D	D(95)E(4)	0.12
200	G	G(30)H(60)A(6) C(2)	0.12
330	D	D(41)N(2)E(46) S(4)T(2)P(2)	0.12

*continued in next column*

res	type	substitutions(%)	cvg
167	L	L(32)I(62)V(4)	0.13
290	Y	F(11)H(62)Y(25)	0.14
301	F	F(81)Y(9)M(2) R(2)L(4)	0.14
112	R	K(60)R(34)L(2) Q(2)	0.15
133	T	T(30)Q(2)F(62) . (2)M(2)	0.15
131	H	F(25)Y(51)H(20) . (2)	0.17
228	N	N(23)S(2)V(53) M(4)A(4)H(4) G(2)I(4)	0.17
257	N	C(4)N(83)A(2) S(9)	0.17
289	I	L(72)T(2)M(4) V(4)I(16)	0.17
100	E	E(76)R(2)G(2) D(16)W(2)	0.19
292	C	A(2)P(6)L(48) E(4)C(27)G(4) M(4)	0.20
256	G	G(23)Y(2)L(46) R(9)V(11)Q(2) E(2)F(2)	0.21
98	L	L(74)M(4)I(9) V(9)E(2)	0.22
135	R	K(25)N(2)R(60) A(4)S(2)Q(2) . (2)	0.22
55	L	F(62)I(2)L(11) Y(9)M(4)H(2) R(2)Q(2)T(2)	0.23
115	D	D(39)T(4)S(4) E(4)N(27)G(18)	0.23
201	S	S(44)T(55)	0.23
332	P	P(55)A(16)S(4) G(16)T(4)R(2)	0.24
255	L	L(60)I(13)M(11) V(13)	0.25
327	F	L(41)P(2)R(4) Y(25)F(25)	0.25

Table 16. Residues forming surface "patch" in 2a5hA.

res	type	disruptive mutations
170	G	(KER) (FQMWH) (NYLPI) (SVA)
171	G	(KER) (FQMWH) (NYLPI) (SVA)
172	D	(R) (FWH) (KYVCAG) (TQM)
198	R	(TD) (SYEVCLAPIG) (FMW) (N)
286	P	(YR) (TH) (SKECG) (FQWD)

*continued in next column*

res	type	disruptive mutations
288	Y	(K) (QM) (NEVLAPIR) (D)
258	Q	(Y) (FTWH) (SVCAG) (D)
287	Y	(K) (QM) (ER) (NLPI)
337	K	(FYW) (TVAH) (SCDG) (ELPI)
325	P	(R) (YH) (K) (FTW)
202	R	(T) (YD) (SVCAG) (FELWPI)
293	D	(FYW) (HR) (CG) (TVA)
125	C	(KER) (FQMWH) (NLPI) (Y)
129	C	(KER) (FQMWH) (NLPI) (Y)
132	C	(KER) (FQMWH) (NLPI) (Y)
230	H	(TE) (D) (SVMCAG) (QLPI)
116	R	(TYD) (E) (SFVCAWG) (LPI)
134	R	(T) (VCAG) (Y) (SLPDI)
260	V	(R) (K) (Y) (E)
113	Y	(KM) (VA) (EQ) (LPI)
169	S	(KR) (FQMWH) (NELPI) (Y)
86	E	(H) (FYW) (R) (CG)
156	Y	(K) (Q) (E) (R)
300	H	(E) (Q) (KM) (D)
111	H	(E) (TD) (M) (VQCLAPIG)
118	L	(YR) (H) (T) (KE)
165	D	(R) (FWH) (YVCAG) (K)
333	G	(FWHR) (E) (Y) (K)
96	D	(R) (FWH) (YVCAG) (K)
200	G	(E) (K) (R) (QD)
330	D	(R) (H) (FW) (Y)
167	L	(YR) (H) (T) (KE)
290	Y	(K) (Q) (EM) (N)
301	F	(E) (K) (T) (D)
112	R	(T) (Y) (D) (SCG)
133	T	(R) (K) (H) (FW)
131	H	(E) (Q) (MD) (TK)
228	N	(Y) (H) (R) (E)
257	N	(Y) (H) (FWR) (E)
289	I	(R) (Y) (H) (TK)
100	E	(FWH) (R) (Y) (VCAG)
292	C	(R) (K) (H) (E)
256	G	(KER) (HD) (FQW) (YM)
98	L	(Y) (R) (H) (T)
135	R	(TY) (D) (ECG) (VA)
55	L	(YR) (T) (E) (H)
115	D	(R) (FWH) (K) (Y)
201	S	(KR) (FQMWH) (NELPI) (Y)
332	P	(R) (Y) (H) (KE)
255	L	(Y) (R) (H) (T)
327	F	(E) (K) (T) (D)

Table 17. Disruptive mutations for the surface patch in 2a5hA.

### 3 NOTES ON USING TRACE RESULTS

#### 3.1 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.

### 3.2 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 *not* 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%.

### 3.3 Surface

To detect candidates for novel functional interfaces, first we look for residues that are solvent accessible (according to DSSP program) by at least  $10\text{\AA}^2$ , 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  $5\text{\AA}$  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.)

### 3.4 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  $5\text{\AA}$ .

### 3.5 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).

### 3.6 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

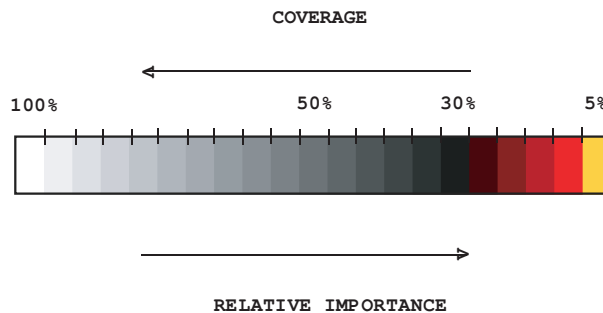


Fig. 13. Coloring scheme used to color residues by their relative importance.

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 NH<sub>2</sub> 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.

## 4 APPENDIX

### 4.1 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

## 4.2 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. 13.

## 4.3 Credits

**4.3.1 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  $(\text{idents} / \text{MIN}(\text{len1}, \text{len2}))$  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.

**4.3.2 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.

**4.3.3 DSSP** In this work a residue is considered solvent accessible if the DSSP program finds it exposed to water by at least  $10\text{\AA}^2$ , 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>.

**4.3.4 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/>

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

**4.3.6 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/>

**4.3.7 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.)

## 4.4 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.

## 4.5 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.

## 4.6 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.

## 4.7 Attachments

The following files should accompany this report:

- 2a5hA.complex.pdb - coordinates of 2a5hA with all of its interacting partners
- 2a5hA.etvx - ET viewer input file for 2a5hA
- 2a5hA.cluster\_report.summary - Cluster report summary for 2a5hA
- 2a5hA.ranks - Ranks file in sequence order for 2a5hA
- 2a5hA.clusters - Cluster descriptions for 2a5hA
- 2a5hA.msf - the multiple sequence alignment used for the chain 2a5hA
- 2a5hA.descr - description of sequences used in 2a5hA msf
- 2a5hA.ranks.sorted - full listing of residues and their ranking for 2a5hA
- 2a5hA.2a5hD.if.pml - Pymol script for Figure 5
- 2a5hA.cbvcg - used by other 2a5hA – related pymol scripts
- 2a5hA.2a5hAPLP419.if.pml - Pymol script for Figure 6

- 2a5hA.2a5hASAM417.if.pml - Pymol script for Figure 7
- 2a5hA.2a5hSO4592.if.pml - Pymol script for Figure 8
- 2a5hA.2a5hBPLP419.if.pml - Pymol script for Figure 9
- 2a5hA.2a5hASF4418.if.pml - Pymol script for Figure 10
- 2a5hA.2a5hB.if.pml - Pymol script for Figure 11