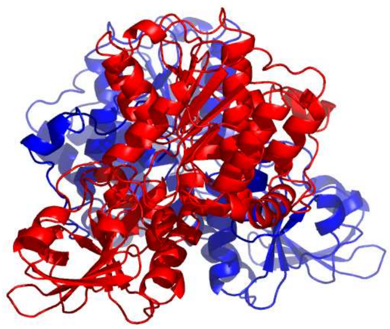
1a7a

Evolutionary trace report by **report_maker**

May 23, 2010



CONTENTS

1 INTRODUCTION

From the original Protein Data Bank entry (PDB id 1a7a):
Title: Structure of human placental s-adenosylhomocysteine hydrolase: determination of a 30 selenium atom substructure from data at a single wavelength
Compound: Mol id: 1; molecule: s-adenosylhomocysteine hydrolase; chain: a, b; ec: 3.3.1.1; engineered: yes; other details: selenomethionine derivatized protein
Organism, scientific name: Homo Sapiens;

1a7a contains a single unique chain 1a7aA (431 residues long) and its homologue 1a7aB.

2 CHAIN 1A7AA

2.1 P23526 overview

From SwissProt, id P23526, 90% identical to 1a7aA:
Description: Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine hydrolase) (AdoHcyase).
Organism, scientific name: Homo sapiens (Human).
Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
Function: Adenosylhomocysteine is a competitive inhibitor of S-adenosyl-L-methionine-dependent methyl transferase reactions; therefore adenosylhomocysteinase may play a key role in the control of methylations via regulation of the intracellular concentration of adenosylhomocysteine.
Catalytic activity: S-adenosyl-L-homocysteine + H(2)O = L- homocysteine + adenosine.
Cofactor: Binds 1 NAD per subunit.
Pathway: Activated methyl cycle.
Subunit: Homotetramer.
Subcellular location: Cytoplasmic.
Disease: Defects in AHCY are a cause of hypermethioninemia [MIM:180960]. It is a disease characterized by elevated levels of methionine in the sera.
Similarity: Belongs to the adenosylhomocysteinase family.

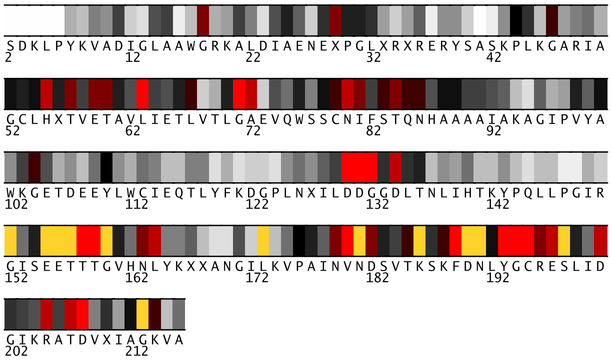


Fig. 1. Residues 2-216 in 1a7aA colored by their relative importance. (See Appendix, Fig.12, for the coloring scheme.)

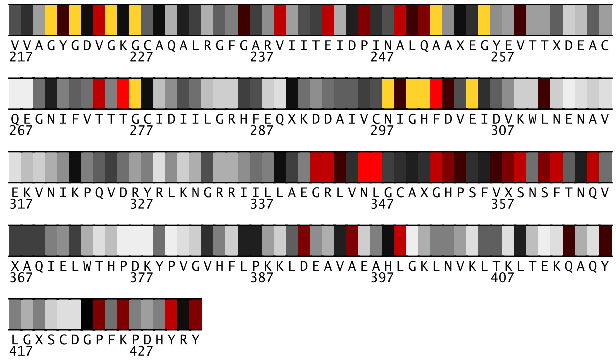


Fig. 2. Residues 217-432 in 1a7aA colored by their relative importance. (See Appendix, Fig.12, for the coloring scheme.)

2.2 Multiple sequence alignment for 1a7aA

For the chain 1a7aA, the alignment 1a7aA.msf (attached) with 631 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 1a7aA.msf. Its statistics, from the *alistat* program are the following:

```
Format:                 MSF
Number of sequences: 631
Total number of residues:     262895
Smallest:               166
Largest:                431
Average length:         416.6
Alignment length:       431
Average identity:       55%
Most related pair:      99%
Most unrelated pair:  0%
Most distant seq:       41%
```

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

The alignment consists of 10% eukaryotic (2% vertebrata, <1% arthropoda, 2% fungi, 2% plantae), 10% prokaryotic, and 2% 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 1a7aA.descr.

2.3 Residue ranking in 1a7aA

The 1a7aA sequence is shown in Figs. 1–2, with each residue colored according to its estimated importance. The full listing of residues in 1a7aA can be found in the file called 1a7aA.ranks_sorted in the attachment.

2.4 Top ranking residues in 1a7aA 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 1a7aA 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.

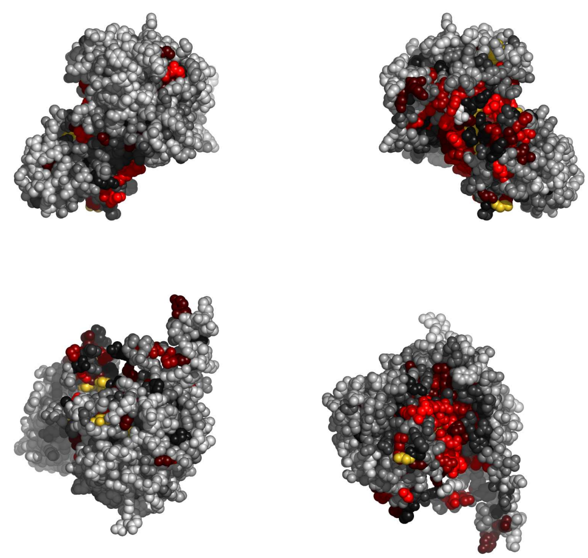


Fig. 3. Residues in 1a7aA, colored by their relative importance. Clockwise: front, back, top and bottom views.

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.

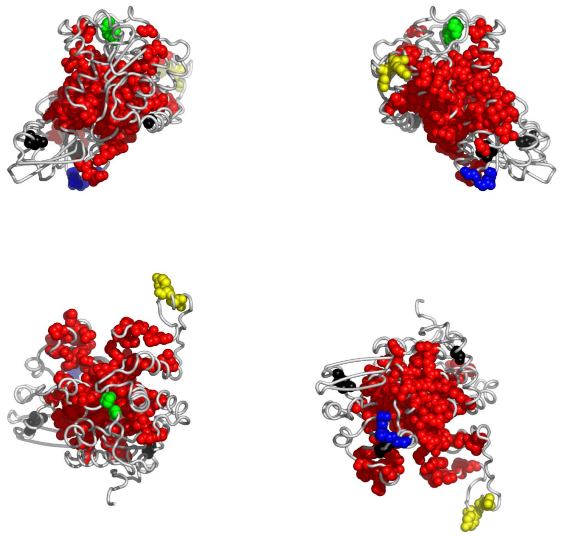


Fig. 4. Residues in 1a7aA, colored according to the cluster they belong to: red, followed by blue and yellow are the largest clusters (see Appendix for the coloring scheme). Clockwise: front, back, top and bottom views. The corresponding Pymol script is attached.

Table 1.		
cluster color	**size**	**member residues**
red	97	29,47,55,57,59,60,63,67,71 72,79,80,81,83,84,85,86,104 130,131,132,134,155,156,157 158,159,160,163,164,173,179 180,181,182,185,186,188,189 190,191,193,194,195,196,197 198,201,205,207,208,220,221 222,224,225,227,243,246,249 250,251,252,256,259,274,276 277,298,299,300,301,302,303 305,342,343,344,346,347,352 353,354,357,358,359,361,362 365,391,395,399,423,424,426 430,432
blue	3	213,214,236
yellow	2	413,416
green	2	152,176

Table 1. Clusters of top ranking residues in 1a7aA.

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 1a7aB1. Table 2 lists the top 25% of residues at the interface with 1a7aB1. 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)
213	G	G(97).T S(1)AE	0.04	18/18	3.15
256	G	G(98)S. TI	0.05	30/30	3.37
236	G	G(90)T. S(4) N(1)QRH WYIA	0.20	28/28	3.27

Table 2. The top 25% of residues in 1a7aA at the interface with 1a7aB1. (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
213	G	(R)(K)(H)(FW)
256	G	(R)(K)(E)(H)
236	G	(E)(K)(R)(D)

Table 3. List of disruptive mutations for the top 25% of residues in 1a7aA, that are at the interface with 1a7aB1.

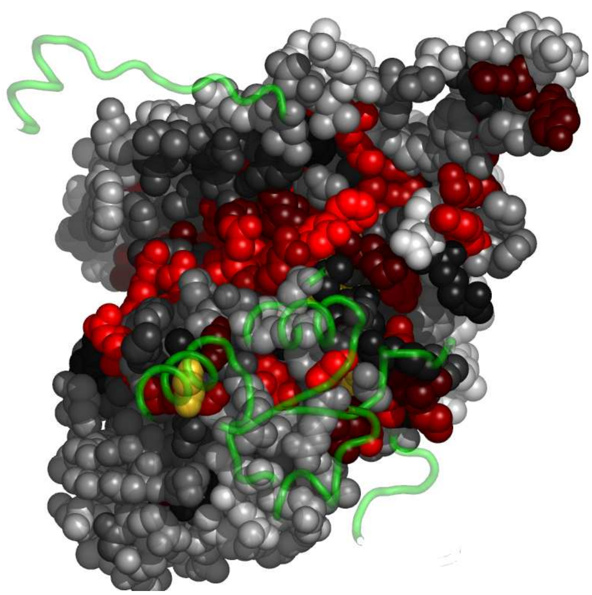


Fig. 5. Residues in 1a7aA, at the interface with 1a7aB1, colored by their relative importance. 1a7aB1 is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 1a7aA.)

Figure 5 shows residues in 1a7aA colored by their importance, at the interface with 1a7aB1.

Interface with 1a7aA1. Table 4 lists the top 25% of residues at the interface with 1a7aA1. 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)
213	G	G(97).T S(1)AE	0.04	14/14	3.12
193	Y	Y(96). H(2)IRF E	0.06	5/3	3.86
194	G	G(94) C(2).SL PTQ	0.06	1/1	4.76
208	D	N(20) D(77).K SLEGI	0.08	36/10	2.67
201	D	D(94).E S(1) A(1)TQH	0.13	13/0	3.52
205	R	R(94).K GHSIQAM PN(1)L	0.13	55/9	2.67
197	E	Q(18) H(22) E(57).S YDL	0.14	86/11	2.84
196	R	G(21) R(66). K(11)M	0.17	31/6	2.94
214	K	K(94) R(2). S(1)ANT QM	0.20	13/2	2.81
236	G	G(90)T. S(4) N(1)QRH WYIA	0.20	5/5	3.77
354	P	P(90)C .(1)G S(6)	0.22	35/9	3.21
357	V	V(85)L .(1) I(11)AT	0.22	11/5	3.48

Table 4. The top 25% of residues in 1a7aA at the interface with 1a7aA1. (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 5.		
res	type	disruptive mutations
213	G	(R)(K)(H)(FW)
193	Y	(K)(Q)(M)(EVA)
194	G	(R)(K)(E)(H)
208	D	(R)(H)(FW)(Y)
201	D	(R)(FWH)(KY)(VCAG)
205	R	(T)(Y)(D)(E)
197	E	(FWH)(R)(VCAG)(Y)
196	R	(TD)(Y)(S)(E)
214	K	(Y)(FW)(T)(H)
236	G	(E)(K)(R)(D)
354	P	(R)(Y)(H)(K)
357	V	(R)(K)(Y)(E)

Table 5. List of disruptive mutations for the top 25% of residues in 1a7aA, that are at the interface with 1a7aA1.

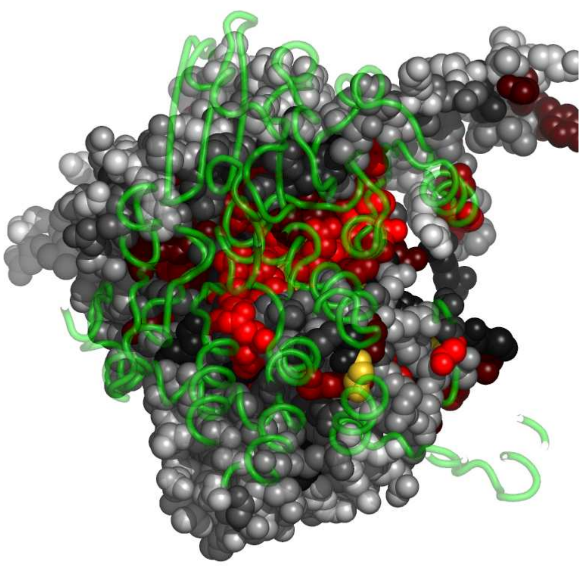


Fig. 6. Residues in 1a7aA, at the interface with 1a7aA1, colored by their relative importance. 1a7aA1 is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 1a7aA.)

Figure 6 shows residues in 1a7aA colored by their importance, at the interface with 1a7aA1.

NAD binding site. Table 6 lists the top 25% of residues at the interface with 1a7aNAD433 (nad). 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)
190	D	D(98).G	0.00	3/0	3.70

continued in next column

Table 6. *continued*					
res	type	subst's (%)	cvg	noc/ bb	dist (Å)
		EHP			
186	K	K(98).T SQM	0.01	3/0	4.11
220	G	G(99).S	0.01	14/14	3.48
222	G	G(99).	0.01	25/25	2.73
300	G	G(98)P .(1)	0.01	16/16	3.47
277	G	G(98)F .(1)D	0.02	12/12	4.11
305	E	E(98)S .(1)L	0.02	1/0	4.86
157	T	T(94) S(4)A.	0.03	17/2	2.69
181	N	N(98).A SYRKML	0.03	2/0	4.87
191	N	N(97).D S(1)ETR	0.04	22/5	3.49
225	G	G(97)S. CAQL	0.05	1/1	4.52
301	H	H(96)K .(1)RGQ D	0.05	19/13	3.02
159	T	T(96) S(1)YDA .NI	0.06	18/8	2.74
158	T	T(94) V(3)SIW RGL.Y	0.07	43/11	2.62
276	T	T(98).V EYS	0.07	55/22	3.19
346	N	N(97)M .(1)AK	0.07	10/0	2.73
195	C	T(23) C(72).V I(1)GAQ	0.08	5/0	4.21
224	V	C(17) V(81).F PTI	0.10	37/9	2.72
243	E	E(98)L. DSH	0.10	44/11	2.83
353	H	H(92)AI .(1)K V(2)QN	0.15	11/0	3.11
344	L	L(96)K .(1)MVG AI	0.21	4/0	3.95
221	Y	Y(87) F(11).L	0.23	10/10	4.14
299	I	A(12) I(72) S(7) M(3)	0.23	21/12	2.99

continued in next column

Table 6. *continued*					
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist (Å)**
		.(1)C V(1)THG L			

Table 6. The top 25% of residues in 1a7aA at the interface with NAD.(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**
190	D	(R)(FWH)(KY)(VCAG)
186	K	(Y)(FW)(T)(H)
220	G	(KR)(E)(FMWH)(Q)
222	G	(KER)(FQMWHD)(NLPI)(Y)
300	G	(R)(KE)(H)(FWD)
277	G	(KR)(E)(Q)(MH)
305	E	(H)(FW)(R)(Y)
157	T	(KR)(QH)(FMW)(NE)
181	N	(Y)(TH)(FW)(E)
191	N	(Y)(FW)(H)(TR)
225	G	(R)(K)(E)(H)
301	H	(E)(T)(D)(M)
159	T	(R)(K)(H)(FQW)
158	T	(K)(R)(Q)(H)
276	T	(KR)(Q)(H)(FW)
346	N	(Y)(T)(H)(FW)
195	C	(R)(KE)(H)(FWD)
224	V	(R)(K)(E)(Y)
243	E	(FWH)(R)(YCG)(VA)
353	H	(E)(T)(D)(SQMCG)
344	L	(Y)(R)(H)(T)
221	Y	(K)(Q)(EMR)(N)
299	I	(R)(Y)(H)(K)

Table 7. List of disruptive mutations for the top 25% of residues in 1a7aA, that are at the interface with NAD.

Figure 7 shows residues in 1a7aA colored by their importance, at the interface with 1a7aNAD433.

NAD binding site. Table 8 lists the top 25% of residues at the interface with 1a7aNAD434 (nad). The following table (Table 9) suggests possible disruptive replacements for these residues (see Section 3.6).

Table 8.					
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist (Å)**
430	Y	.(24) Y(75)V	0.14	15/0	2.84
426	K	.(24)	0.16	16/0	2.62
					continued in next column

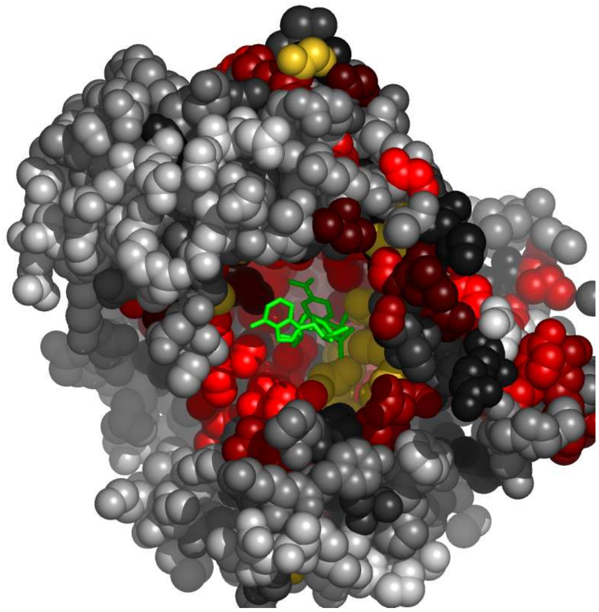


Fig. 7. Residues in 1a7aA, at the interface with NAD, colored by their relative importance. The ligand (NAD) is colored green. Atoms further than 30Å away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme for the protein chain 1a7aA.)

Table 8. *continued*					
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist (Å)**
413	Q	K(75)D Q(95)K .(3)ERD	0.24	5/0	3.64

Table 8. The top 25% of residues in 1a7aA at the interface with NAD.(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**
430	Y	(K)(Q)(M)(E)
426	K	(Y)(FW)(T)(VCAG)
413	Q	(Y)(FW)(T)(H)

Table 9. List of disruptive mutations for the top 25% of residues in 1a7aA, that are at the interface with NAD.

Figure 8 shows residues in 1a7aA colored by their importance, at the interface with 1a7aNAD434.

ADC binding site. Table 10 lists the top 25% of residues at the interface with 1a7aADC435 (adc). The following table (Table

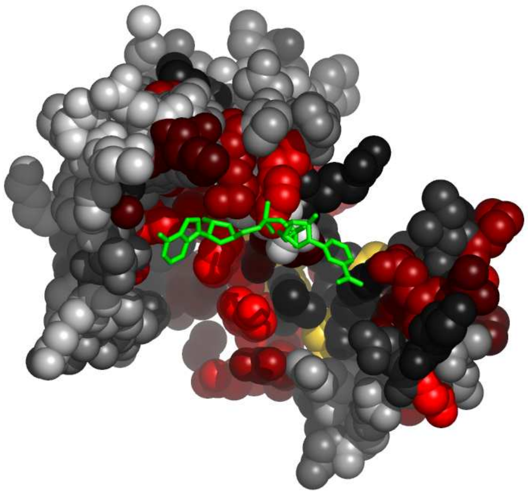


Fig. 8. Residues in 1a7aA, at the interface with NAD, colored by their relative importance. The ligand (NAD) is colored green. Atoms further than 30$\AA$ away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme for the protein chain 1a7aA.)

11) suggests possible disruptive replacements for these residues (see Section 3.6).

Table 10.

res	type	subst's (%)	cvg	noc/ bb	dist ($\AA$)
190	D	D(98).G EHP	0.00	15/0	2.74
186	K	K(98).T SQM	0.01	7/0	3.17
156	E	E(98)KP F.QG	0.03	13/2	2.64
157	T	T(94) S(4)A.	0.03	11/4	2.99
181	N	N(98).A SYRKML	0.03	1/0	4.87
191	N	N(97).D S(1)ETR	0.04	1/0	4.65
346	N	N(97)M .(1)AK	0.07	2/0	4.37
131	D	D(97) .(1)NSL EI	0.08	16/0	3.81
347	L	L(96) .(1)IY	0.09	22/0	3.67
352	G	G(93)R T(3) .(1)ASN	0.11	12/12	3.73
362	F	F(93)L A(2) .(1)MTS Y	0.12	7/0	3.75
55	H	H(96) .(2)LPS IV	0.14	27/5	3.36
57	T	E(11) T(84) .(1)Y N(1)	0.15	17/2	2.53
60	T	T(96) .(1)ACS V	0.15	8/2	3.29
353	H	H(92)AI .(1)K V(2)QN	0.15	43/16	2.84
358	X	M(93)KH .(1)E L(2) V(1)AIX	0.17	33/0	2.80
361	S	S(92) T(4) .(1)GDA	0.17	2/0	4.39
59	E	K(11) Q(73) E(12) .(1)HSN	0.18	14/0	2.73
344	L	L(96)K .(1)MVG AI	0.21	1/0	4.72
85	Q	Q(93) D(1) .(1) H(1)VNA KLGR	0.23	1/0	4.88

Table 10. The top 25% of residues in 1a7aA at the interface with ADC.(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 11.

res	type	disruptive mutations
190	D	(R)(FWH)(KY)(VCAG)
186	K	(Y)(FW)(T)(H)
156	E	(H)(FW)(Y)(CG)
157	T	(KR)(QH)(FMW)(NE)

continued in next column

Table 11. *continued*		
res	**type**	**disruptive mutations**
181	N	(Y)(TH)(FW)(E)
191	N	(Y)(FW)(H)(TR)
346	N	(Y)(T)(H)(FW)
131	D	(R)(H)(FW)(Y)
347	L	(R)(Y)(TH)(K)
352	G	(E)(KR)(FWH)(MD)
362	F	(K)(E)(Q)(R)
55	H	(E)(Q)(T)(D)
57	T	(R)(K)(H)(FW)
60	T	(KR)(Q)(H)(FMW)
353	H	(E)(T)(D)(SQMCG)
358	X	(Y)(R)(E)(K)
361	S	(R)(K)(H)(Q)
59	E	(FW)(H)(Y)(VCAG)
344	L	(Y)(R)(H)(T)
85	Q	(Y)(T)(FWH)(SCG)

Table 11. List of disruptive mutations for the top 25% of residues in 1a7aA, that are at the interface with ADC.

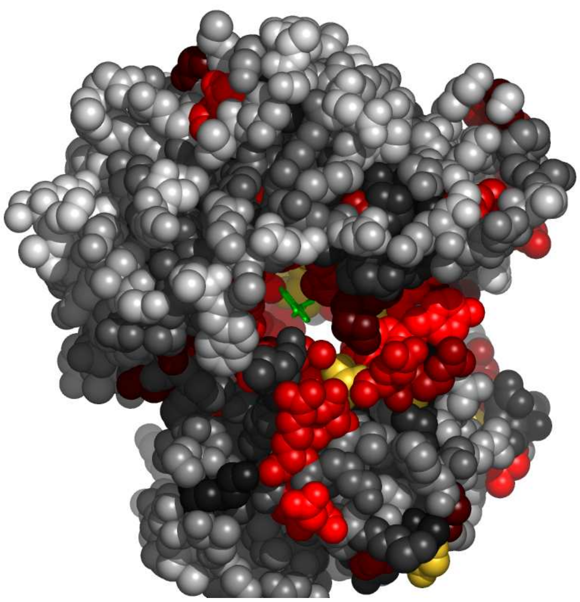


Fig. 9. Residues in 1a7aA, at the interface with ADC, colored by their relative importance. The ligand (ADC) is colored green. Atoms further than 30$\AA$ away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme for the protein chain 1a7aA.)

Figure 9 shows residues in 1a7aA colored by their importance, at the interface with 1a7aADC435.

Interface with 1a7aB. Table 12 lists the top 25% of residues at the interface with 1a7aB. 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 ($\AA$)
222	G	G(99).	0.01	4/4	3.47
277	G	G(98)F .(1)D	0.02	7/7	3.67
189	F	F(90) N(5).VM ICLY(1)	0.06	38/0	3.56
193	Y	Y(96). H(2)IRF E	0.06	52/0	2.95
302	F	F(92)L S(3) .(1)YVG HIAT	0.08	1/0	4.58
243	E	E(98)L. DSH	0.10	28/26	2.83
249	A	A(97)T. RGSIPYL	0.12	5/1	3.63
399	L	L(88) .(2) V(8)WSI	0.13	19/0	4.01
430	Y	.(24) Y(75)V	0.14	24/7	3.35
391	D	D(96)T .(2)NEH GM	0.15	8/4	4.34
395	A	A(96)R .(2)SIG	0.15	23/12	3.70
426	K	.(24) K(75)D	0.16	20/0	2.94
182	D	D(95). E(1) N(2)RLH SG	0.17	19/5	2.80
196	R	G(21) R(66). K(11)M	0.17	14/0	2.93
251	Q	E(19)A Q(78).K LMPNDR	0.17	35/11	3.11
246	P	P(96)M. NECASFH	0.18	62/20	3.56
179	N	N(80) A(15)MD EVLIKT	0.19	3/0	4.04
432	Y	S(1). .(25) Y(73)W	0.19	72/18	3.06
185	T	T(90).L C(2) M(5)SAI VP	0.20	4/0	4.05
250	L	L(84).F M(6) I(4) N(1) V(1)TA	0.21	39/5	3.69
259	V	V(94). L(2) I(1)TPA N	0.22	25/16	3.38
357	V	V(85)L .(1) I(11)AT	0.22	1/1	4.96
188	K	M(4) K(77) L(12).E F(2)IDP YNGQA	0.23	18/0	3.07
221	Y	Y(87) F(11).L	0.23	1/1	5.00
416	Y	Y(93)I .(3) F(1)L	0.23	56/3	3.54
413	Q	Q(95)K .(3)ERD	0.24	46/6	2.70

Table 12. The top 25% of residues in 1a7aA at the interface with 1a7aB. (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 13.		
res	type	disruptive mutations
222	G	(KER)(FQMWHD)(NLPI)(Y)
277	G	(KR)(E)(Q)(MH)
189	F	(KE)(TDR)(Q)(S)
193	Y	(K)(Q)(M)(EVA)
302	F	(K)(E)(Q)(D)
243	E	(FWH)(R)(YCG)(VA)
249	A	(R)(K)(E)(Y)
399	L	(R)(Y)(H)(TK)
430	Y	(K)(Q)(M)(E)
391	D	(R)(FWH)(Y)(K)
395	A	(Y)(E)(KR)(H)
426	K	(Y)(FW)(T)(VCAG)
182	D	(R)(FW)(H)(Y)
196	R	(TD)(Y)(S)(E)
251	Q	(Y)(TH)(FW)(CG)

continued in next column

Table 13. *continued*		
res	**type**	**disruptive mutations**
246	P	(R)(Y)(H)(T)
179	N	(Y)(H)(FW)(R)
432	Y	(K)(Q)(M)(E)
185	T	(R)(K)(H)(Q)
250	L	(R)(Y)(H)(T)
259	V	(R)(Y)(KE)(H)
357	V	(R)(K)(Y)(E)
188	K	(Y)(T)(FW)(CG)
221	Y	(K)(Q)(EMR)(N)
416	Y	(K)(Q)(R)(E)
413	Q	(Y)(FW)(T)(H)

Table 13. List of disruptive mutations for the top 25% of residues in 1a7aA, that are at the interface with 1a7aB.

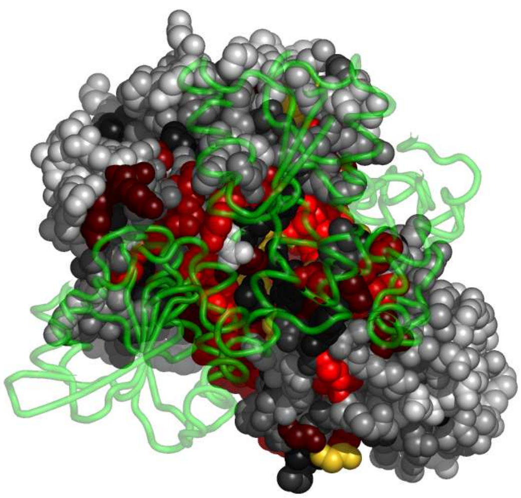


Fig. 10. Residues in 1a7aA, at the interface with 1a7aB, colored by their relative importance. 1a7aB is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 1a7aA.)

Figure 10 shows residues in 1a7aA colored by their importance, at the interface with 1a7aB.

2.4.3 Possible novel functional surfaces at 25% coverage. One group of residues is conserved on the 1a7aA surface, away from (or susbtantially larger than) other functional sites and interfaces recognizable in PDB entry 1a7a. It is shown in Fig. 11. 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 14, while Table 15 suggests possible disruptive replacements for these residues (see Section 3.6).

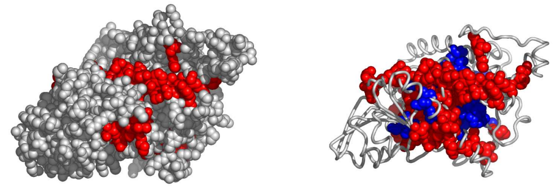


Fig. 11. A possible active surface on the chain 1a7aA. The larger cluster it belongs to is shown in blue.

Table 14.			
res	**type**	**substitutions(%)**	**cvg**
190	D	D(98).GEHP	0.00
186	K	K(98).TSQM	0.01
220	G	G(99).S	0.01
222	G	G(99).	0.01
300	G	G(98)P.(1)	0.01
277	G	G(98)F.(1)D	0.02
305	E	E(98)S.(1)L	0.02
156	E	E(98)KPF.QG	0.03
157	T	T(94)S(4)A.	0.03
173	L	L(98)FVW.CQ	0.04
191	N	N(97).DS(1)ETR	0.04
301	H	H(96)K.(1)RGQD	0.05
159	T	T(96)S(1)YDA.NI	0.06
189	F	F(90)N(5).VMICL Y(1)	0.06
193	Y	Y(96).H(2)IRFE	0.06
132	G	G(98).(1)RPAI	0.07
158	T	T(94)V(3)SIWRGL .Y	0.07
276	T	T(98).VEYS	0.07
346	N	N(97)M.(1)AK	0.07
131	D	D(97).(1)NSLEI	0.08
195	C	T(23)C(72).V I(1)GAQ	0.08
208	D	N(20)D(77).KSLE GI	0.08
302	F	F(92)LS(3).(1)Y VGHIAT	0.08
347	L	L(96).(1)IY	0.09
224	V	C(17)V(81).FPTI	0.10
243	E	E(98)L.DSH	0.10
80	N	N(96).(2)LPGEHA	0.11
134	D	D(96).(1)AGYCES QH	0.11
352	G	G(93)RT(3).(1)A	0.11
continued in next column			

Table 14. *continued*			
res	type	substitutions(%)	cvg
		SN	
164	L	L(95)KGYPEI.MFV A	0.12
249	A	A(97)T.RGSIPYL	0.12
343	R	R(96)T.(1)NMQWA DE	0.12
201	D	D(94).ES(1)A(1) TQH	0.13
205	R	R(94).KGHSIQAMP N(1)L	0.13
399	L	L(88).(2)V(8)WS I	0.13
55	H	H(96).(2)LPSIV	0.14
197	E	Q(18)H(22)E(57) .SYDL	0.14
207	T	T(95).M(1)AIVLC SG	0.14
430	Y	.(24)Y(75)V	0.14
57	T	E(11)T(84).(1)Y N(1)	0.15
353	H	H(92)AI.(1)K V(2)QN	0.15
391	D	D(96)T.(2)NEHGM	0.15
395	A	A(96)R.(2)SIG	0.15
163	N	R(85)N(6)YIH(5) QAL.EKSP	0.16
426	K	.(24)K(75)D	0.16
196	R	G(21)R(66). K(11)M	0.17
251	Q	E(19)AQ(78).KLM PNDR	0.17
358	X	M(93)KH.(1)E L(2)V(1)AIX	0.17
59	E	K(11)Q(73)E(12) .(1)HSN	0.18
81	I	P(20)I(75).(2)H RGSVL	0.18
246	P	P(96)M.NECASFH	0.18
424	P	.(23)RP(75)T	0.19
432	Y	.(25)Y(73)W	0.19
86	N	D(91)N(4).(1)RE GP	0.20
185	T	T(90).LC(2)M(5) SAIVP	0.20
79	C	C(80)S(15).(2)G PAQW	0.21
250	L	L(84).FM(6)I(4) N(1)V(1)TA	0.21
303	D	D(83)IN(13).(1) SQHCAE	0.21
344	L	L(96)K.(1)MVGAI	0.21
259	V	V(94).L(2)I(1)T	0.22
		PAN	
354	P	P(90)C.(1)GS(6)	0.22
357	V	V(85)L.(1)I(11) AT	0.22
85	Q	Q(93)D(1).(1) H(1)VNAKLGR	0.23
188	K	M(4)K(77)L(12). EF(2)IDPYNGQA	0.23
221	Y	Y(87)F(11).L	0.23
299	I	A(12)I(72)S(7) M(3).(1)CV(1)TH GL	0.23
104	G	G(92).(1)D(1) N(1)RT(1)EAI	0.24
110	Y	Y(81)F(10).(1) H(4)GAWLQE	0.25
176	P	P(93)LR(1)IA(1) CSQT(1)N.	0.25
423	G	G(89)V.(10)SN	0.25

Table 14. Residues forming surface "patch" in 1a7aA.

Table 15.		
res	type	disruptive mutations
190	D	(R)(FWH)(KY)(VCAG)
186	K	(Y)(FW)(T)(H)
220	G	(KR)(E)(FMWH)(Q)
222	G	(KER)(FQMWHD)(NLPI)(Y)
300	G	(R)(KE)(H)(FWD)
277	G	(KR)(E)(Q)(MH)
305	E	(H)(FW)(R)(Y)
156	E	(H)(FW)(Y)(CG)
157	T	(KR)(QH)(FMW)(NE)
173	L	(R)(Y)(T)(E)
191	N	(Y)(FW)(H)(TR)
301	H	(E)(T)(D)(M)
159	T	(R)(K)(H)(FQW)
189	F	(KE)(TDR)(Q)(S)
193	Y	(K)(Q)(M)(EVA)
132	G	(E)(R)(K)(HD)
158	T	(K)(R)(Q)(H)
276	T	(KR)(Q)(H)(FW)
346	N	(Y)(T)(H)(FW)
131	D	(R)(H)(FW)(Y)
195	C	(R)(KE)(H)(FWD)
208	D	(R)(H)(FW)(Y)
302	F	(K)(E)(Q)(D)
347	L	(R)(Y)(TH)(K)
224	V	(R)(K)(E)(Y)
243	E	(FWH)(R)(YCG)(VA)

continued in next column

Table 15. *continued*		
res	**type**	**disruptive mutations**
80	N	(Y)(H)(T)(FWR)
134	D	(R)(FWH)(K)(Y)
352	G	(E)(KR)(FWH)(MD)
164	L	(R)(Y)(H)(T)
249	A	(R)(K)(E)(Y)
343	R	(T)(Y)(D)(CG)
201	D	(R)(FWH)(KY)(VCAG)
205	R	(T)(Y)(D)(E)
399	L	(R)(Y)(H)(TK)
55	H	(E)(Q)(T)(D)
197	E	(FWH)(R)(VCAG)(Y)
207	T	(R)(K)(H)(Q)
430	Y	(K)(Q)(M)(E)
57	T	(R)(K)(H)(FW)
353	H	(E)(T)(D)(SQMCG)
391	D	(R)(FWH)(Y)(K)
395	A	(Y)(E)(KR)(H)
163	N	(Y)(T)(H)(FW)
426	K	(Y)(FW)(T)(VCAG)
196	R	(TD)(Y)(S)(E)
251	Q	(Y)(TH)(FW)(CG)
358	X	(Y)(R)(E)(K)
59	E	(FW)(H)(Y)(VCAG)
81	I	(Y)(R)(T)(H)
246	P	(R)(Y)(H)(T)
424	P	(Y)(R)(H)(T)
432	Y	(K)(Q)(M)(E)
86	N	(Y)(FWH)(T)(R)
185	T	(R)(K)(H)(Q)
79	C	(R)(KE)(H)(D)
250	L	(R)(Y)(H)(T)
303	D	(R)(H)(FW)(Y)
344	L	(Y)(R)(H)(T)
259	V	(R)(Y)(KE)(H)
354	P	(R)(Y)(H)(K)
357	V	(R)(K)(Y)(E)
85	Q	(Y)(T)(FWH)(SCG)
188	K	(Y)(T)(FW)(CG)
221	Y	(K)(Q)(EMR)(N)
299	I	(R)(Y)(H)(K)
104	G	(R)(K)(H)(FEW)
110	Y	(K)(Q)(MR)(E)
176	P	(Y)(R)(H)(T)
423	G	(R)(KE)(H)(FW)

Table 15. Disruptive mutations for the surface patch in 1a7aA.

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 *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%.

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 $10Å^2$, 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 $5Å$ 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 $5Å$.

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 properties: small $[AVGSTC]$, medium $[LPNQDEMIK]$, large

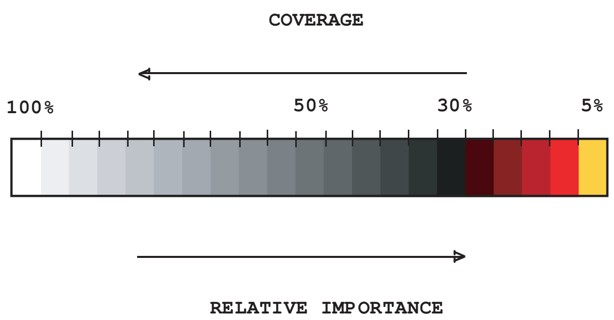


Fig. 12. Coloring scheme used to color residues by their relative importance.

$[WFYHR]$, 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.

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. 12.

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 (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.

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 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.`

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 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:

- 1a7aA.complex.pdb - coordinates of 1a7aA with all of its interacting partners
- 1a7aA.etvx - ET viewer input file for 1a7aA
- 1a7aA.cluster_report.summary - Cluster report summary for 1a7aA
- 1a7aA.ranks - Ranks file in sequence order for 1a7aA
- 1a7aA.clusters - Cluster descriptions for 1a7aA
- 1a7aA.msf - the multiple sequence alignment used for the chain 1a7aA
- 1a7aA.descr - description of sequences used in 1a7aA msf
- 1a7aA.ranks_sorted - full listing of residues and their ranking for 1a7aA
- 1a7aA.1a7aB1.if.pml - Pymol script for Figure 5
- 1a7aA.cbcvg - used by other 1a7aA – related pymol scripts
- 1a7aA.1a7aA1.if.pml - Pymol script for Figure 6
- 1a7aA.1a7aNAD433.if.pml - Pymol script for Figure 7
- 1a7aA.1a7aNAD434.if.pml - Pymol script for Figure 8
- 1a7aA.1a7aADC435.if.pml - Pymol script for Figure 9
- 1a7aA.1a7aB.if.pml - Pymol script for Figure 10