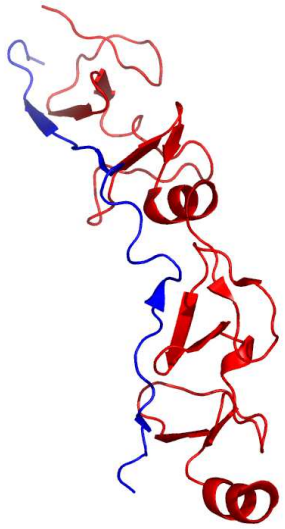
2xjy

Evolutionary trace report by **report_maker**

July 26, 2010



CONTENTS

1 INTRODUCTION

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

Title: Crystal structure of the lmo2:ldb1-lid complex, p21 crystal form

Compound: Mol id: 1; molecule: rhombotin-2; chain: a; fragment: residues 26-156; synonym: lim domain only protein 2, cysteine-rich protein t t-cell translocation protein 2, lmo-2; engineered: yes; mol id: 2; molecule: lim domain-binding protein 1; chain: b; fragment: residues 334-368; synonym: nuclear lim interactor, carboxyl-terminal lim domain-binding protein 2, lim domain-binding factor clim2, clim-2, hldb1; engineered: yes

Organism, scientific name: Homo Sapiens;

2xjy contains unique chains 2xjyA (131 residues) and 2xjyB (35 residues)

2 CHAIN 2XJYA

2.1 Q71LH8 overview

From SwissProt, id Q71LH8, 100% identical to 2xjyA:

Description: LIM domain-containing transcription factor.

Organism, scientific name: Gallus gallus (Chicken).

Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

Similarity: Contains 2 LIM zinc-binding domains.

2.2 Multiple sequence alignment for 2xjyA

For the chain 2xjyA, the alignment 2xjyA.msf (attached) with 411 sequences was used. The alignment was downloaded from the HSSP

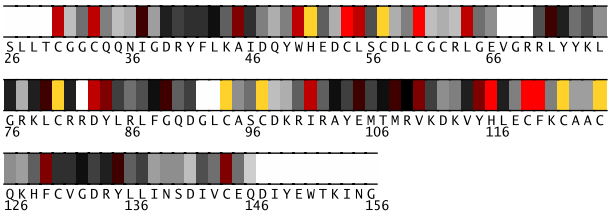


Fig. 1. Residues 26-156 in 2xjyA colored by their relative importance. (See Appendix, Fig.14, for the coloring scheme.)

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 2xjyA.msf. Its statistics, from the *alistat* program are the following:

```
Format:              MSF
Number of sequences: 411
Total number of residues:    47985
Smallest:            99
Largest:             131
Average length:      116.8
Alignment length:    131
Average identity:    33%
Most related pair:   98%
Most unrelated pair: 12%
Most distant seq:    33%
```

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

The alignment consists of 30% eukaryotic (22% vertebrata, 3% arthropoda, <1% fungi) sequences. (Descriptions of some sequences were not readily available.) The file containing the sequence descriptions can be found in the attachment, under the name 2xjyA.descr.

2.3 Residue ranking in 2xjyA

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

2.4 Top ranking residues in 2xjyA 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 2xjyA 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.

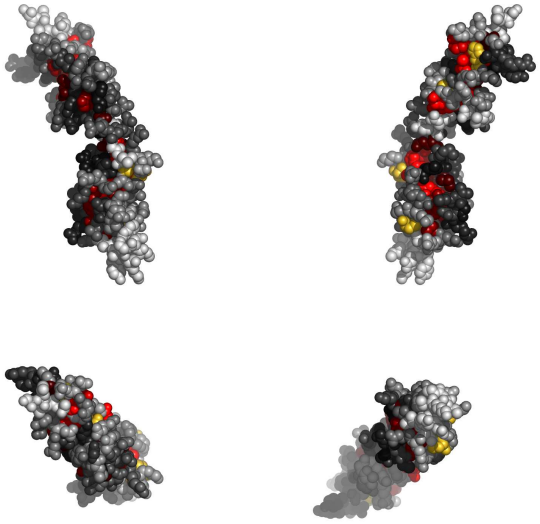


Fig. 2. Residues in 2xjyA, colored by their relative importance. Clockwise: front, back, top and bottom views.

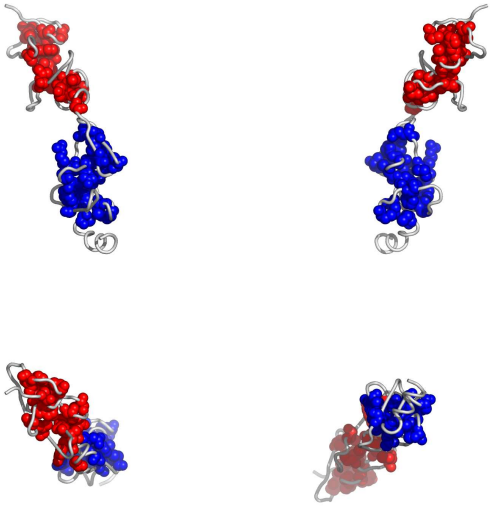


Fig. 3. Residues in 2xjyA, colored according to the cluster they belong to: red, followed by blue and yellow are the largest clusters (see Appendix for the coloring scheme). Clockwise: front, back, top and bottom views. The corresponding Pymol script is attached.

Table 1.		
cluster color	**size**	**member residues**
continued in next column		

Table 1. *continued*		
cluster color	**size**	**member residues**
red	17	30,33,37,45,50,51,54,55,57 60,64,71,79,80,83,84,89
blue	16	94,97,101,105,108,109,110 115,116,119,120,122,125,129 135,144

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

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.

Zinc ion binding site. Table 2 lists the top 25% of residues at the interface with 2xjyAZN201 (zinc 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)**	**antn**
51	H	H(98)TA .M	0.05	8/2	2.34	site
54	C	C(89) H(9)GEV	0.09	4/2	2.18	
30	C	C(97) .(2)	0.10	3/1	2.46	site
33	C	C(97) .(2)IQ	0.11	5/3	2.18	site

Table 2. The top 25% of residues in 2xjyA at the interface with zinc 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**
51	H	(E)(Q)(D)(K)
54	C	(K)(R)(E)(Q)
30	C	(KER)(FQMWHD)(NLPI)(Y)
33	C	(R)(EH)(FKW)(YD)

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

Figure 4 shows residues in 2xjyA colored by their importance, at the interface with 2xjyAZN201.

Zinc ion binding site. Table 4 lists the top 25% of residues at the interface with 2xjyAZN204 (zinc ion). The following table (Table 5) suggests possible disruptive replacements for these residues (see Section 4.6).

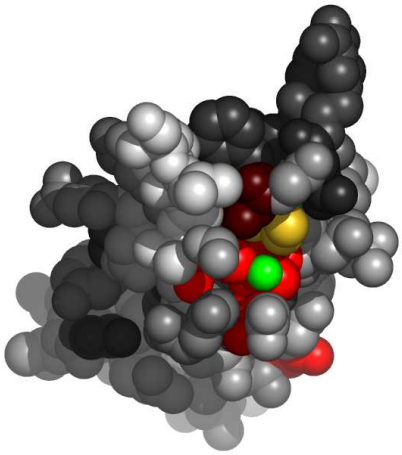


Fig. 4. Residues in 2xjyA, at the interface with zinc ion, colored by their relative importance. The ligand (zinc ion) is colored green. Atoms further than 30Å away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme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

Table 4.						
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist (Å)**	**antn**
122	C	C(99)YR L	0.01	3/1	2.27	site
125	C	C(99)TP A	0.02	5/3	2.20	site
144	C	C(84) .(10) V(1) H(1)IYK GT	0.18	4/2	2.33	site

Table 4. The top 25% of residues in 2xjyA at the interface with zinc 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 5.		
res	**type**	**disruptive mutations**
122	C	(E)(K)(R)(D)
125	C	(R)(K)(E)(H)
144	C	(E)(KR)(Q)(D)

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

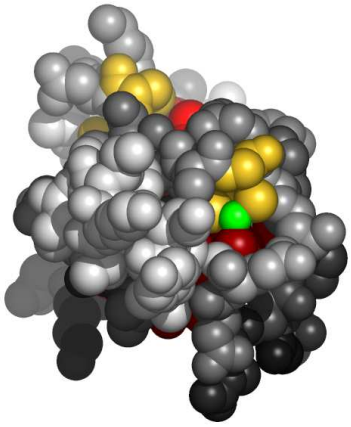


Fig. 5. Residues in 2xjyA, at the interface with zinc ion, colored by their relative importance. The ligand (zinc ion) is colored green. Atoms further than 30Å away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme for the protein chain 2xjyA.)

Figure 5 shows residues in 2xjyA colored by their importance, at the interface with 2xjyAZN204.

Interface with 2xjyB. Table 6 lists the top 25% of residues at the interface with 2xjyB. 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 (Å)
120	F	F(93) L(1)S Y(2) A(2)	0.06	5/0	3.71
55	L	L(52) F(39) V(6)SAP C	0.12	4/0	3.84
101	I	I(78) L(6) M(1) F(7) V(6)T	0.13	15/6	3.51
64	L	L(82)M K(1) I(2) F(8). V(1)WQY	0.14	5/4	3.52

continued in next column

Table 6. *continued*

res	type	subst's (%)	cvg	noc/ bb	dist (Å)
		CS			
84	Y	Y(73) F(22)S. GHNW	0.15	32/0	2.55
129	F	L(56) F(26) I(13)Q M(2)CV.	0.16	7/1	3.84
45	A	A(54) V(29) S(5)TL Y(1)M I(1) H(2).DE GKRNF	0.18	12/5	3.65
110	V	A(74) G(3)K V(9)L S(1)P Y(2) C(1) T(1)MR I(1)HFE	0.19	2/0	3.97
135	Y	Y(29) F(51) T(2)H L(3) M(2) C(1)S .(1)A V(2)KEI Q	0.20	83/29	2.84
37	I	I(86) V(3) .(1)H L(4)MAF CRKQ	0.21	4/3	3.69
108	M	A(4) R(20) M(41)P I(11) L(9)WY V(7)TH .(1) K(1)DQC S	0.21	23/1	3.78
89	G	G(55) S(8) D(1) A(20) E(2)HQP TV(2)L	0.22	7/7	3.82

continued in next column

Table 6. *continued*

res	type	subst's (%)	cvg	noc/ bb	dist (Å)
		K(1)IMN			
		R(1)F			
105	E	E(45)	0.23	16/16	3.24
		D(22)			
		Y(2)A			
		G(13)			
		S(1)			
		Q(7)MF			
		T(2)			
		V(1)HCN			
		K			
71	L	C(42)	0.24	52/28	2.81
		.(8)D			
		Y(8)			
		F(25)			
		L(8)			
		I(1)MTV			
		NSA(1)G			
109	R	K(7)	0.25	15/1	3.42
		R(53)			
		L(2)			
		S(9)			
		V(8)F			
		E(5)			
		Q(1)			
		H(1)G			
		T(1)			
		.(2)			
		Y(1)PAC			
		MNI			

Table 6. The top 25% of residues in 2xjyA at the interface with 2xjyB. (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
120	F	(K)(E)(Q)(R)
55	L	(R)(Y)(KH)(E)
101	I	(R)(Y)(H)(TK)
64	L	(R)(Y)(TH)(E)
84	Y	(K)(Q)(M)(E)
129	F	(E)(K)(T)(D)
45	A	(R)(Y)(K)(E)
110	V	(R)(KYE)(H)(D)
135	Y	(K)(QR)(M)(E)
37	I	(Y)(R)(T)(H)
108	M	(Y)(H)(T)(R)

continued in next column

Table 7. *continued*

res	type	disruptive mutations
89	G	(R)(E)(K)(H)
105	E	(H)(FW)(R)(Y)
71	L	(R)(Y)(H)(K)
109	R	(D)(T)(Y)(E)

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

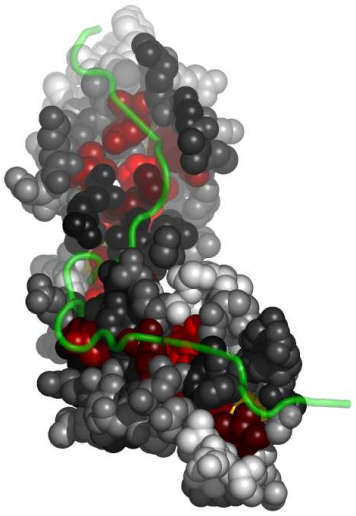


Fig. 6. Residues in 2xjyA, at the interface with 2xjyB, colored by their relative importance. 2xjyB is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 2xjyA.)

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

Zinc ion binding site. Table 8 lists the top 25% of residues at the interface with 2xjyAZN202 (zinc ion). 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
57	C	C(99)AQ	0.03	3/1	2.40	site
80	C	C(99).V	0.05	4/2	2.29	site
60	C	C(97).A	0.08	4/2	2.38	site
		G(1)				
83	D	D(83)	0.12	6/2	2.07	
		H(4)				
		C(10)E.				

continued in next column

Table 8. *continued*						
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist (Å)**	**antn**
		Y				

Table 8. The top 25% of residues in 2xjyA at the interface with zinc 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 9.		
res	**type**	**disruptive mutations**
57	C	(E)(R)(H)(K)
80	C	(KER)(HD)(Q)(FMW)
60	C	(KER)(QHD)(FMW)(NY)
83	D	(R)(FW)(KH)(VA)

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

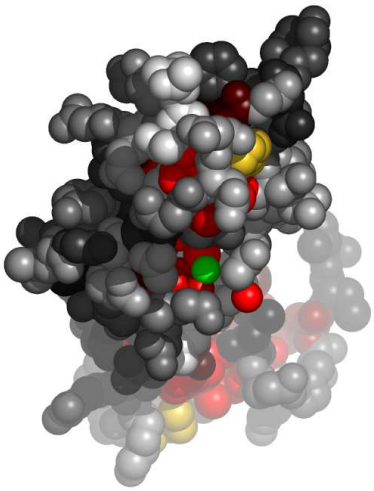


Fig. 7. Residues in 2xjyA, at the interface with zinc ion, colored by their relative importance. The ligand (zinc ion) is colored green. Atoms further than 30Å away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme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

Figure 7 shows residues in 2xjyA colored by their importance, at the interface with 2xjyAZN202.

Zinc ion binding site. Table 10 lists the top 25% of residues at the interface with 2xjyAZN203 (zinc ion). The following table (Table 11) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 10.						
res	**type**	**subst's (%)**	**cvg**	**noc/ bb**	**dist (Å)**	**antn**
97	C	C(99).T	0.01	5/3	2.19	site
94	C	C(99).AR	0.04	3/1	2.46	site
119	C	C(94) H(4)GI	0.07	4/2	2.34	
116	H	H(98)C. NDK	0.08	8/2	2.21	site

Table 10. The top 25% of residues in 2xjyA at the interface with zinc 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 11.		
res	**type**	**disruptive mutations**
97	C	(KR)(E)(FMW)(H)
94	C	(E)(KD)(R)(FQMWH)
119	C	(E)(KR)(QD)(M)
116	H	(E)(T)(M)(D)

Table 11. List of disruptive mutations for the top 25% of residues in 2xjyA, that are at the interface with zinc ion.

Figure 8 shows residues in 2xjyA colored by their importance, at the interface with 2xjyAZN203.

2.4.3 Possible novel functional surfaces at 25% coverage. One group of residues is conserved on the 2xjyA surface, away from (or susbtantially larger than) other functional sites and interfaces recognizable in PDB entry 2xjy. It is shown in Fig. 9. The residues belonging to this surface "patch" are listed in Table 12, while Table 13 suggests possible disruptive replacements for these residues (see Section 4.6).

Table 12.				
res	**type**	**substitutions(%)**	**cvg**	**antn**
57	C	C(99)AQ	0.03	site
51	H	H(98)TA.M	0.05	site
80	C	C(99).V	0.05	site
60	C	C(97).AG(1)	0.08	site
54	C	C(89)H(9)GEV	0.09	
33	C	C(97).(2)IQ	0.11	site
55	L	L(52)F(39)V(6)S APC	0.12	
83	D	D(83)H(4)C(10)E .Y	0.12	
50	W	W(81)Y(8)F(8)G.	0.14	

continued in next column

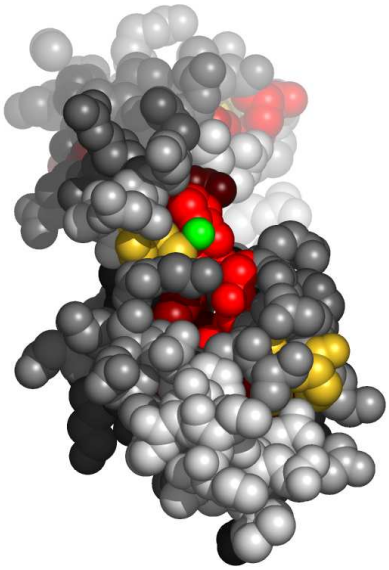


Fig. 8. Residues in 2xjyA, at the interface with zinc ion, colored by their relative importance. The ligand (zinc ion) is colored green. Atoms further than 30$\mathring{A}$ away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme for the protein chain 2xjyA.)

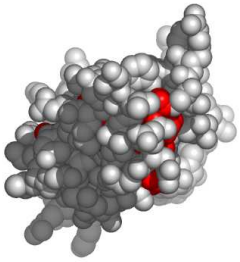


Fig. 9. A possible active surface on the chain 2xjyA.

Table 12. *continued*

res	type	substitutions(%)	cvg	antn
64	L	V L(82)MK(1)I(2) F(8).V(1)WQYCS	0.14	

continued in next column

Table 12. *continued*

res	type	substitutions(%)	cvg	antn
84	Y	Y(73)F(22)S.GHN W	0.15	
45	A	A(54)V(29)S(5)T LY(1)MI(1)H(2). DEGKRNF	0.18	
37	I	I(86)V(3).(1)H L(4)MAFCRKQ	0.21	
89	G	G(55)S(8)D(1) A(20)E(2)HQPT V(2)LK(1)IMN R(1)F	0.22	
71	L	C(42).(8)DY(8) F(25)L(8)I(1)MT VNSA(1)G	0.24	

Table 12. Residues forming surface "patch" in 2xjyA.

Table 13.

res	type	disruptive mutations
57	C	(E)(R)(H)(K)
51	H	(E)(Q)(D)(K)
80	C	(KER)(HD)(Q)(FMW)
60	C	(KER)(QHD)(FMW)(NY)
54	C	(K)(R)(E)(Q)
33	C	(R)(EH)(FKW)(YD)
55	L	(R)(Y)(KH)(E)
83	D	(R)(FW)(KH)(VA)
50	W	(K)(E)(Q)(D)
64	L	(R)(Y)(TH)(E)
84	Y	(K)(Q)(M)(E)
45	A	(R)(Y)(K)(E)
37	I	(Y)(R)(T)(H)
89	G	(R)(E)(K)(H)
71	L	(R)(Y)(H)(K)

Table 13. Disruptive mutations for the surface patch in 2xjyA.

Another group of surface residues is shown in Fig.10. The residues belonging to this surface "patch" are listed in Table 14, while Table 15 suggests possible disruptive replacements for these residues (see Section 4.6).

Table 14.

res	type	substitutions(%)	cvg	antn
97	C	C(99).T	0.01	site
125	C	C(99)TPA	0.02	site
94	C	C(99).AR	0.04	site
120	F	F(93)L(1)SY(2) A(2)	0.06	
119	C	C(94)H(4)GI	0.07	
116	H	H(98)C.NDK	0.08	site
101	I	I(78)L(6)M(1)	0.13	

continued in next column

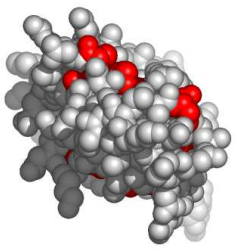


Fig. 10. Another possible active surface on the chain 2xjyA.

Table 14. *continued*

res	type	substitutions(%)	cvg	antn
		F(7)V(6)T		
129	F	L(56)F(26)I(13)QM(2)CV.	0.16	
115	Y	F(23)Y(55)W(19)L.IHE	0.17	
144	C	C(84).(10)V(1)H(1)IYKGT	0.18	site
110	V	A(74)G(3)KV(9)LS(1)PY(2)C(1)T(1)MRI(1)HFE	0.19	
135	Y	Y(29)F(51)T(2)HL(3)M(2)C(1)S.(1)AV(2)KEIQ	0.20	
108	M	A(4)R(20)M(41)PI(11)L(9)WYV(7)TH.(1)K(1)DQCS	0.21	
105	E	E(45)D(22)Y(2)AG(13)S(1)Q(7)MFT(2)V(1)HCNK	0.23	
109	R	K(7)R(53)L(2)S(9)V(8)FE(5)Q(1)H(1)GT(1).(2)Y(1)PACMNI	0.25	

Table 14. Residues forming surface ”patch” in 2xjyA.

Table 15.

res	type	disruptive mutations
97	C	(KR)(E)(FMW)(H)
125	C	(R)(K)(E)(H)
94	C	(E)(KD)(R)(FQMWH)
120	F	(K)(E)(Q)(R)
119	C	(E)(KR)(QD)(M)
116	H	(E)(T)(M)(D)
101	I	(R)(Y)(H)(TK)
129	F	(E)(K)(T)(D)
115	Y	(K)(Q)(R)(EM)
144	C	(E)(KR)(Q)(D)
110	V	(R)(KYE)(H)(D)
135	Y	(K)(QR)(M)(E)
108	M	(Y)(H)(T)(R)
105	E	(H)(FW)(R)(Y)
109	R	(D)(T)(Y)(E)

Table 15. Disruptive mutations for the surface patch in 2xjyA.

3 CHAIN 2XJYB

3.1 Q4T230 overview

From SwissProt, id Q4T230, 100% identical to 2xjyB:
Description: Chromosome undetermined SCAF10361, whole genome shotgun sequence. (Fragment).
Organism, scientific name: Tetraodon nigroviridis (Green puffer).
Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetradontoidea; Tetraodontidae; Tetraodon.
Caution: The sequence shown here is derived from an EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is preliminary data.

3.2 Multiple sequence alignment for 2xjyB

For the chain 2xjyB, the alignment 2xjyB.msf (attached) with 15 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 2xjyB.msf. Its statistics, from the *alistat* program are the following:

```
Format:              MSF
Number of sequences: 15
Total number of residues:    492
Smallest:            28
Largest:             35
Average length:      32.8
Alignment length:    35
Average identity:    78%
Most related pair:   97%
Most unrelated pair: 59%
Most distant seq:    77%
```

Furthermore, 40% of residues show as conserved in this alignment.

Fig. 11. Residues 334-368 in 2xjyB colored by their relative importance. (See Appendix, Fig.14, for the coloring scheme.)

The alignment consists of 26% eukaryotic (20% 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 2xjyB.descr.

3.3 Residue ranking in 2xjyB

The 2xjyB sequence is shown in Fig. 11, with each residue colored according to its estimated importance. The full listing of residues in 2xjyB can be found in the file called 2xjyB.ranks_sorted in the attachment.

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

In the following we consider residues ranking among top 40% of residues in the protein (the closest this analysis allows us to get to 25%). Figure 12 shows residues in 2xjyB 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.

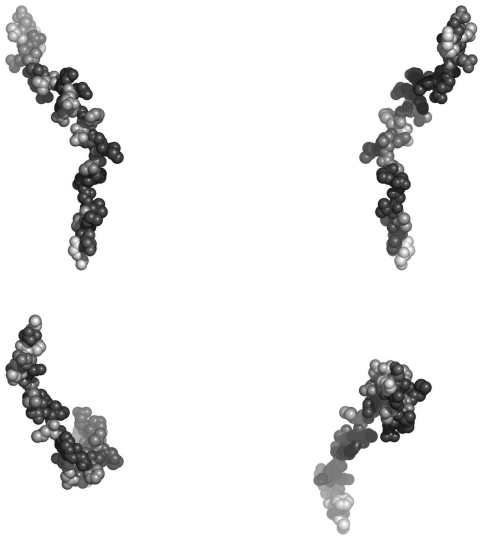


Fig. 12. Residues in 2xjyB, colored by their relative importance. Clockwise: front, back, top and bottom views.

3.4.1 Clustering of residues at 40% coverage. Fig. 13 shows the top 40% of all residues, this time colored according to clusters they belong to. The clusters in Fig.13 are composed of the residues listed in Table 16.

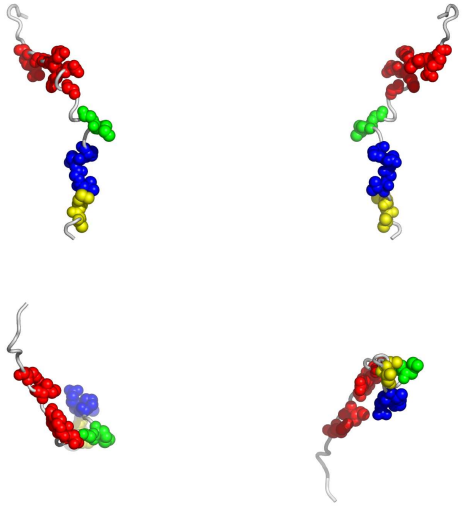


Fig. 13. Residues in 2xjyB, colored according to the cluster they belong to: red, followed by blue and yellow are the largest clusters (see Appendix for the coloring scheme). Clockwise: front, back, top and bottom views. The corresponding Pymol script is attached.

Table 16.		
cluster color	**size**	**member residues**
red	6	351,354,355,356,358,360
blue	4	340,341,342,343
yellow	2	337,338
green	2	346,347

Table 16. Clusters of top ranking residues in 2xjyB.

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,

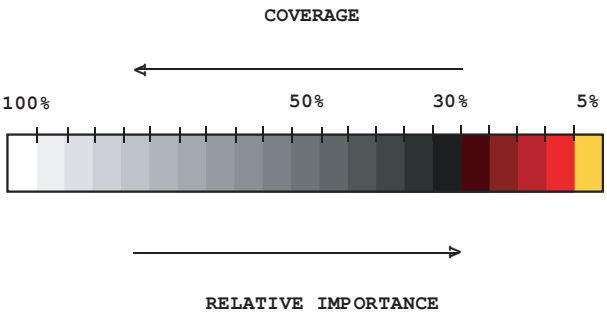


Fig. 14. Coloring scheme used to color residues by their relative importance.

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. 14.

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:

- 2xjyA.complex.pdb - coordinates of 2xjyA with all of its interacting partners
- 2xjyA.etvx - ET viewer input file for 2xjyA
- 2xjyA.cluster_report.summary - Cluster report summary for 2xjyA

- 2xjyA.ranks - Ranks file in sequence order for 2xjyA
- 2xjyA.clusters - Cluster descriptions for 2xjyA
- 2xjyA.msf - the multiple sequence alignment used for the chain 2xjyA
- 2xjyA.descr - description of sequences used in 2xjyA msf
- 2xjyA.ranks_sorted - full listing of residues and their ranking for 2xjyA
- 2xjyA.2xjyAZN201.if.pml - Pymol script for Figure 4
- 2xjyA.cbcvg - used by other 2xjyA – related pymol scripts
- 2xjyA.2xjyAZN204.if.pml - Pymol script for Figure 5
- 2xjyA.2xjyB.if.pml - Pymol script for Figure 6
- 2xjyA.2xjyAZN202.if.pml - Pymol script for Figure 7
- 2xjyA.2xjyAZN203.if.pml - Pymol script for Figure 8
- 2xjyB.complex.pdb - coordinates of 2xjyB with all of its interacting partners
- 2xjyB.etvx - ET viewer input file for 2xjyB
- 2xjyB.cluster_report.summary - Cluster report summary for 2xjyB
- 2xjyB.ranks - Ranks file in sequence order for 2xjyB
- 2xjyB.clusters - Cluster descriptions for 2xjyB
- 2xjyB.msf - the multiple sequence alignment used for the chain 2xjyB
- 2xjyB.descr - description of sequences used in 2xjyB msf
- 2xjyB.ranks_sorted - full listing of residues and their ranking for 2xjyB