1ap2
Evolutionary trace report by report_maker
October 31, 2009

CONTENTS
1 Introduction
2 Chain 1ap2C
   2.1 Q6KB05 overview
   2.2 Multiple sequence alignment for 1ap2C
   2.3 Residue ranking in 1ap2C
   2.4 Top ranking residues in 1ap2C 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 Chain 1ap2D
   3.1 Q505N9 overview
   3.2 Multiple sequence alignment for 1ap2D
   3.3 Residue ranking in 1ap2D
   3.4 Top ranking residues in 1ap2D and their position on the structure
      3.4.1 Clustering of residues at 25% coverage.
      3.4.2 Overlap with known functional surfaces at 25% coverage.
      3.4.3 Possible novel functional surfaces at 25% coverage.
4 Notes on using trace results
   4.1 Coverage
   4.2 Known substitutions
   4.3 Surface
   4.4 Number of contacts
   4.5 Annotation
   4.6 Mutation suggestions
5 Appendix
   5.1 File formats
   5.2 Color schemes used
   5.3 Credits
      5.3.1 Alistat
      5.3.2 CE
      5.3.3 DSSP
      5.3.4 HSSP
      5.3.5 LaTex
      5.3.6 Muscle
      5.3.7 Pymol
   5.4 Note about ET Viewer
   5.5 Citing this work
   5.6 About report_maker
   5.7 Attachments

1 INTRODUCTION
From the original Protein Data Bank entry (PDB id 1ap2):
Title: Single chain fv of c219
Compound: Mol id: 1; molecule: monoclonal antibody c219; chain: a, c; fragment: fv; synonym: variable domain; engineered: yes; other details: light and heavy chains linked with a synthetic (ggggs)3 linker; mol id: 2; molecule: monoclonal antibody c219; chain: b, d; fragment: fv; synonym: variable domain; engineered: yes; other details: light and heavy chains linked with a synthetic (ggggs)3 linker
Organism, scientific name: Mus Musculus;
1ap2 contains unique chains 1ap2C (113 residues) and 1ap2D (123 residues) 1ap2A is a homologue of chain 1ap2C. 1ap2B is a homologue of chain 1ap2D.

2 CHAIN 1AP2C
2.1 Q6KB05 overview
From SwissProt, id Q6KB05, 91% identical to 1ap2C:
Description: ScFv B8E5 protein (Fragment).
Organism, scientific name: Mus musculus (Mouse).
Taxonomy: Eukaryota; Metazoa; Chordata; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Lichtarge lab 2006
2.2 Multiple sequence alignment for 1ap2C

For the chain 1ap2C, the alignment 1ap2C.msf (attached) with 853 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 1ap2C.msf. Its statistics, from the alstat program are the following:

- Format: MSF
- Number of sequences: 853
- Total number of residues: 84184
- Smallest: 47
- Largest: 113
- Average length: 98.7
- Alignment length: 113
- Average identity: 36%
- Most related pair: 99%
- Most unrelated pair: 6%
- Most distant seq: 35%

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

The alignment consists of 42% eukaryotic (42% vertebrata), and <1% viral sequences. (Descriptions of some sequences were not readily available.) The file containing the sequence descriptions can be found in the attachment, under the name 1ap2C.descr.

2.3 Residue ranking in 1ap2C

The 1ap2C sequence is shown in Fig. 1, with each residue colored according to its estimated importance. The full listing of residues in 1ap2C can be found in the file called 1ap2C.ranks, sorted in the attachment.

2.4 Top ranking residues in 1ap2C 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 1ap2C colored by their importance: bright red and yellow indicate more conserved/important residues (see Appendix for the coloring scheme). A Pymol script for producing this figure can be found in the attachment.

2.4.1 Clustering of residues at 25% coverage. Fig. 3 shows the top 25% of all residues, this time colored according to clusters they belong to. The clusters in Fig.3 are composed of the residues listed in Table 1.
Table 1. Clusters of top ranking residues in 1ap2C.

<table>
<thead>
<tr>
<th>color</th>
<th>size</th>
<th>residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td>26</td>
<td>6,16,19,21,23,41,42,43,44,50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>53,67,68,69,77,78,79,81,84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>87,88,89,90,92,93,94</td>
</tr>
<tr>
<td>blue</td>
<td>2</td>
<td>46,47</td>
</tr>
</tbody>
</table>

Table 1. Clusters of top ranking residues in 1ap2C.

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 1ap2D. Table 2 lists the top 25% of residues at the interface with 1ap2D. The following table (Table 3) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 2.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>subst’s (%)</th>
<th>cvg</th>
<th>noc/bb</th>
<th>dist (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>T</td>
<td>113.96</td>
<td>0.01</td>
<td>2/0</td>
<td>3.98</td>
</tr>
<tr>
<td>50</td>
<td>P</td>
<td>L(21)IR</td>
<td>0.07</td>
<td>46/12</td>
<td>3.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F(2)AN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>V(1)T.S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Q</td>
<td>Q(91).R</td>
<td>0.09</td>
<td>22/0</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Z(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H(1)VS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R(1)LPW</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Y</td>
<td>Y(70)</td>
<td>0.12</td>
<td>29/0</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>V(15)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I(2)NHG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>WAQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RI(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>continued in next column</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>Y</td>
<td>L(4)</td>
<td>0.19</td>
<td>30/0</td>
<td>3.42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y(68)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F(22)RI</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>H(1)X.C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ST</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Y</td>
<td>191.23</td>
<td>0.24</td>
<td>51/0</td>
<td>2.40</td>
</tr>
</tbody>
</table>

Table 2. The top 25% of residues in 1ap2C at the interface with 1ap2D. (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. List of disruptive mutations for the top 25% of residues in 1ap2C, that are at the interface with 1ap2D.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>T</td>
<td>(KR)(Q)(M)(FWH)</td>
</tr>
<tr>
<td>50</td>
<td>P</td>
<td>(Y)(R)(H)(T)</td>
</tr>
<tr>
<td>44</td>
<td>Q</td>
<td>(Y)(T)(H)(FW)</td>
</tr>
<tr>
<td>42</td>
<td>Y</td>
<td>(K)(E)(Q)(R)</td>
</tr>
<tr>
<td>93</td>
<td>Y</td>
<td>(K)(Q)(M)(E)</td>
</tr>
<tr>
<td>100</td>
<td>Y</td>
<td>(K)(QM)(E)(R)</td>
</tr>
</tbody>
</table>

Table 3. List of disruptive mutations for the top 25% of residues in 1ap2C, that are at the interface with 1ap2D.

Figure 4 shows residues in 1ap2C colored by their importance, at the interface with 1ap2D.

Figure 4. Residues in 1ap2C, at the interface with 1ap2D, colored by their relative importance. 1ap2D is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 1ap2C.)

Interface with 1ap2A. Table 4 lists the top 25% of residues at the interface with 1ap2A. The following table (Table 5) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 4.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>subst’s (%)</th>
<th>cvg</th>
<th>noc/bb</th>
<th>dist (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>69</td>
<td>T</td>
<td>LK(4)</td>
<td>0.25</td>
<td>1/0</td>
<td>4.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S(62)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T(20)A.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Q(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>E(2)WPF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>XR(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. The top 25% of residues in 1ap2C at the interface with 1ap2D. (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.)

continued in next column
Table 4. The top 25% of residues in 1ap2C at the interface with 1ap2A. (Field names: res: residue number in the PDB entry; type: amino acid type; subs: 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. List of disruptive mutations for the top 25% of residues in 1ap2C, that are at the interface with 1ap2A.

Table 6. Residues forming surface “patch” in 1ap2C.

Table 7. Disruptive mutations for the surface patch in 1ap2C.

Figure 5 shows residues in 1ap2C colored by their relative importance, at the interface with 1ap2A.

2.4.3 Possible novel functional surfaces at 25% coverage. One group of residues is conserved on the 1ap2C surface, away from (or substantially larger than) other functional sites and interfaces recognizable in PDB entry 1ap2. It is shown in Fig. 6. The residues belonging to this surface “patch” are listed in Table 6, while Table 7 suggests possible disruptive replacements for these residues (see Section 4.6).
Fig. 7. Another possible active surface on the chain 1ap2C. The larger cluster it belongs to is shown in blue.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>T</td>
<td>113.96</td>
<td>0.01</td>
</tr>
<tr>
<td>86</td>
<td>A</td>
<td>87.82</td>
<td>0.01</td>
</tr>
<tr>
<td>88</td>
<td>D</td>
<td>D(93)PB(1)Q(2)G</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N.SHTV</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>A</td>
<td>A(81)G(12)CS(3)</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td>QV.TLRD</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>P</td>
<td>P(68)L(21)IR</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F(2)ANM(1)V(1)T</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.SX</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>Q</td>
<td>Q(91).RE(2)Z(1)</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H(1)VSK(1)LPWI</td>
<td></td>
</tr>
<tr>
<td>68</td>
<td>F</td>
<td>F(84)V(1)L(4)</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y(4).I(2)S(1)HA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DNRW</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>G</td>
<td>G(91).1S(1)D</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R(1)AKNPE(1)</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Y</td>
<td>Y(70)L(1).V(15)</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F(7)I(2)NHGWAQ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>.SX</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>R</td>
<td>S(1)R(83)K(2)EF</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.VG(6)NT(1)DQAI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HMYL</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>Q</td>
<td>Q(55)L(5).R(23)</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K(7)V(2)WZXHDS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y(1)ITMF</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>G</td>
<td>G(82)E(2)HN(1)Q</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D(5)RS(3)KMPWA.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>V</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>L</td>
<td>L(60)M(3)V(15)</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F(1)I(15)W(2)E.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>AHPSY</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>P</td>
<td>A(3)P(80)S(5)M</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L(5)FQKT(1)VHGR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EI</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>L</td>
<td>S(17)A(10)E(21)</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L(4)T(19)V(6)</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. continued

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>D(1)QXC.K</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M(3)S(4)T(3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I(1)RFP.XC</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>Y</td>
<td>L(4)Y(68)F(22)R</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IH(1)X.CST</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Q</td>
<td>DQ(67).17E(3)</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ZT(7)AFMHPVGLQXI</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>V</td>
<td>V(52)A(23).1</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L(14)EI(3)HM(1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GTPKSF</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>E</td>
<td>KA(5)E(65)S(9)</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G(7)D(5)BQMV</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R(1)NT(1)Z.X</td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>T</td>
<td>LK(4)S(62)T(20)</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A.Q(1)E(2)WPFXR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I(1)N(1)D VHGM</td>
<td></td>
</tr>
</tbody>
</table>

Table 8. Residues forming surface "patch" in 1ap2C.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>T</td>
<td>(KR) (Q) (M) (FWH)</td>
</tr>
<tr>
<td>86</td>
<td>A</td>
<td>(KER) (Y) (Q) (D)</td>
</tr>
<tr>
<td>88</td>
<td>D</td>
<td>(R) (H) (FW) (Y)</td>
</tr>
<tr>
<td>90</td>
<td>A</td>
<td>(R) (Y) (K) (E)</td>
</tr>
<tr>
<td>50</td>
<td>P</td>
<td>(Y) (R) (H) (T)</td>
</tr>
<tr>
<td>44</td>
<td>Q</td>
<td>(Y) (T) (H) (FW)</td>
</tr>
<tr>
<td>68</td>
<td>F</td>
<td>(K) (E) (Q) (T)</td>
</tr>
<tr>
<td>16</td>
<td>G</td>
<td>(R) (FWH) (KE) (Y)</td>
</tr>
<tr>
<td>42</td>
<td>Y</td>
<td>(K) (E) (Q) (R)</td>
</tr>
<tr>
<td>67</td>
<td>R</td>
<td>(T) (D) (Y) (E)</td>
</tr>
<tr>
<td>43</td>
<td>Q</td>
<td>(Y) (T) (H) (FW)</td>
</tr>
<tr>
<td>47</td>
<td>G</td>
<td>(R) (E) (K) (H)</td>
</tr>
<tr>
<td>53</td>
<td>L</td>
<td>(R) (Y) (T) (KH)</td>
</tr>
<tr>
<td>46</td>
<td>P</td>
<td>(Y) (R) (H) (T)</td>
</tr>
<tr>
<td>89</td>
<td>L</td>
<td>(R) (Y) (H) (T)</td>
</tr>
<tr>
<td>84</td>
<td>V</td>
<td>(R) (Y) (E) (K)</td>
</tr>
<tr>
<td>93</td>
<td>Y</td>
<td>(K) (Q) (M) (E)</td>
</tr>
<tr>
<td>6</td>
<td>Q</td>
<td>(Y) (H) (FW) (T)</td>
</tr>
<tr>
<td>19</td>
<td>V</td>
<td>(Y) (R) (E) (K)</td>
</tr>
<tr>
<td>87</td>
<td>E</td>
<td>(H) (FW) (Y) (R)</td>
</tr>
<tr>
<td>69</td>
<td>T</td>
<td>(R) (K) (H) (FW)</td>
</tr>
</tbody>
</table>

Table 9. Disruptive mutations for the surface patch in 1ap2C.

3 CHAIN 1AP2D

3.1 Q505N9 overview
From SwissProt, id Q505N9, 82% identical to 1ap2D:
Description: Hypothetical protein.
3.2 Multiple sequence alignment for 1ap2D

For the chain 1ap2D, the alignment 1ap2D.msf (attached) with 819 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 1ap2D.msf. Its statistics, from the alisat program are the following:

- Format: MSF
- Number of sequences: 819
- Total number of residues: 87595
- Smallest: 49
- Largest: 123
- Average length: 107.0
- Alignment length: 123
- Average identity: 41%
- Most related pair: 99%
- Most unrelated pair: 0%
- Most distant seq: 33%

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

The alignment consists of 46% eukaryotic (46% 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 1ap2D.descr.

3.3 Residue ranking in 1ap2D

The 1ap2D sequence is shown in Fig. 8, with each residue colored according to its estimated importance. The full listing of residues in 1ap2D can be found in the file called 1ap2D.ranks sorted in the attachment.

3.4 Top ranking residues in 1ap2D and their position on the structure

In the following we consider residues ranking among top 25% of residues in the protein. Figure 9 shows residues in 1ap2D colored by their importance: bright red and yellow indicate more conserved/important residues (see Appendix for the coloring scheme). A Pymol script for producing this figure can be found in the attachment.

3.4.1 Clustering of residues at 25% coverage.  Fig. 10 shows the top 25% of all residues, this time colored according to clusters they belong to. The clusters in Fig.10 are composed of the residues listed in Table 10.
Table 10. Clusters of top ranking residues in 1ap2D.

<table>
<thead>
<tr>
<th>cluster color</th>
<th>size</th>
<th>member residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td>27</td>
<td>7, 14, 15, 18, 20, 21, 22, 36, 37, 38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39, 41, 42, 45, 46, 48, 49, 67, 81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>83, 86, 90, 91, 92, 94, 95, 96</td>
</tr>
<tr>
<td>blue</td>
<td>2</td>
<td>25, 26</td>
</tr>
</tbody>
</table>

3.4.2 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 1ap2C. Table 11 lists the top 25% of residues at the interface with 1ap2C. The following table (Table 12) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 11. The top 25% of residues in 1ap2D at the interface with 1ap2C. (Field names: res: residue number in the PDB entry; type: amino acid type; subst’s: 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>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>subst’s (%)</th>
<th>cvg</th>
<th>noc/bb</th>
<th>dist (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>L</td>
<td>L(76)</td>
<td>0.04</td>
<td>60/18</td>
<td>3.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P(18)T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R(1)VA.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EFQKMS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Q</td>
<td>Q(94).H</td>
<td>0.07</td>
<td>17/0</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K(1)FRL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>EMZDS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>V</td>
<td>AMV(61)</td>
<td>0.19</td>
<td>2/0</td>
<td>4.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y(17).I(13)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F(4)LXD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>HQ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>E</td>
<td>E(74)</td>
<td>0.20</td>
<td>3/3</td>
<td>4.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>K(7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R(6)XID</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>V(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>T(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Q(2)LGZ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N.AMPSW</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>Y</td>
<td>Y(74)</td>
<td>0.25</td>
<td>40/0</td>
<td>3.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F(20)CT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LMHSW</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12. Disruptive mutations

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>L</td>
<td>(Y) (R) (H) (T)</td>
</tr>
</tbody>
</table>

continued in next column
**Table 12.** List of disruptive mutations for the top 25% of residues in 1ap2D, that are at the interface with 1ap2C.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>Q</td>
<td>(Y) (T) (FWH) (CG)</td>
</tr>
<tr>
<td>37</td>
<td>V</td>
<td>(R) (K) (Y) (E)</td>
</tr>
<tr>
<td>46</td>
<td>E</td>
<td>(H) (FW) (R) (CG)</td>
</tr>
<tr>
<td>95</td>
<td>Y</td>
<td>(K) (Q) (EMR) (N)</td>
</tr>
</tbody>
</table>

Table 12. continued

---

**Figure 11.** Residues in 1ap2D, at the interface with 1ap2C, colored by their relative importance. 1ap2C is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 1ap2D.)

Figure 11 shows residues in 1ap2D colored by their importance, at the interface with 1ap2C.

### 3.4.3 Possible novel functional surfaces at 25% coverage.

One group of residues is conserved on the 1ap2D surface, away from (or substantially larger than) other functional sites and interfaces recognizable in PDB entry 1ap2. 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 13, while Table 14 suggests possible disruptive replacements for these residues (see Section 4.6).

**Table 13.** Residues forming surface “patch” in 1ap2D.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
<th>antn</th>
</tr>
</thead>
<tbody>
<tr>
<td>94</td>
<td>Y</td>
<td>. (1) Y (96) LFHC</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Q</td>
<td>Q (94) HK (1) PRLE MZDS</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>A</td>
<td>A (90) G (5) DS TVCFXL</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>K</td>
<td>K (11) R (67) Q (17) CXHLINMPVYWG</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>P</td>
<td>P (88) H (2) A (1) L (1) S (2) T (1) R. I EGNVM</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>E</td>
<td>G (84) VAE (6) D (1) S (3) PN. RKTOH</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>V</td>
<td>V (28) L (55) . (1) A (7) E (1) I (2) M (1) HRFG</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>L</td>
<td>L (1) L (31) M (39) V (2) I (22) GKF7WA</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>T</td>
<td>T (1) S (12) T (64) E (9) IM (1) A (2) V (1) F (4) LG EAMV61</td>
<td>0.17</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>V</td>
<td>V (13) F (4) LXDHHQ I (13) E (4) LDDGK</td>
<td>0.19</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>E</td>
<td>E (13) K (7) R (6) XI DV (2) T (1) Q (2) LG 2N. AMPSW</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>K</td>
<td>K (3) K (12) R (75) P (3) S (2) NGHQLXEL MTV</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>Y</td>
<td>Y (1) Y (74) F (20) C TLMHSW</td>
<td>0.25</td>
<td></td>
</tr>
</tbody>
</table>

Table 13. continued

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
<th>antn</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>C</td>
<td>. (1) C (98) S</td>
<td>0.02</td>
<td>S–S</td>
</tr>
<tr>
<td>90</td>
<td>D</td>
<td>. (1) D (96) NBGVHQ E</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>L</td>
<td>L (76) P (18) TR (1)</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

**Table 14.** Possible disruptive replacements for these residues (see Section 4.6).

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>96</td>
<td>C</td>
<td>(KR) (E) (FWH) (Q)</td>
</tr>
</tbody>
</table>
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 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%.

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 10Å², 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.)

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 5Å.

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 [LPNQDEMVK], 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 [NQKR]. 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
5.3.2 CE To map ligand binding sites from different source structures, report-maker uses the CE program:

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 10 Å², which is roughly the area needed for one water molecule to come in 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://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).


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


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:

- 1ap2C.complex.pdb - coordinates of 1ap2C with all of its interacting partners
- 1ap2C.etvx - ET viewer input file for 1ap2C
- 1ap2C.cluster_report.summary - Cluster report summary for 1ap2C
- 1ap2C.ranks - Ranks file in sequence order for 1ap2C
- 1ap2C.clusters - Cluster descriptions for 1ap2C
- 1ap2C.msf - the multiple sequence alignment used for the chain 1ap2C
- 1ap2C.descr - description of sequences used in 1ap2C msf
- 1ap2C.ranks_sorted - full listing of residues and their ranking for 1ap2C
- 1ap2D.complex.pdb - coordinates of 1ap2D with all of its interacting partners
- 1ap2D.etvx - ET viewer input file for 1ap2D
- 1ap2D.cluster_report.summary - Cluster report summary for 1ap2D
- 1ap2D.ranks - Ranks file in sequence order for 1ap2D
- 1ap2D.clusters - Cluster descriptions for 1ap2D
- 1ap2D.msf - the multiple sequence alignment used for the chain 1ap2D
- 1ap2D.descr - description of sequences used in 1ap2D msf
- 1ap2D.ranks_sorted - full listing of residues and their ranking for 1ap2D
- 1ap2D.1ap2C.if.pml - Pymol script for Figure 11
- 1ap2D.1ap2D.if.pml - Pymol script for Figure 4
- 1ap2C.cbcvg - used by other 1ap2C – related pymol scripts
- 1ap2C.1ap2A.if.pml - Pymol script for Figure 5
- 1ap2D.complex.pdb - coordinates of 1ap2D with all of its interacting partners
- 1ap2D.etvx - ET viewer input file for 1ap2D
- 1ap2D.cluster_report.summary - Cluster report summary for 1ap2D
- 1ap2D.ranks - Ranks file in sequence order for 1ap2D
- 1ap2D.clusters - Cluster descriptions for 1ap2D
- 1ap2D.msf - the multiple sequence alignment used for the chain 1ap2D
- 1ap2D.descr - description of sequences used in 1ap2D msf
- 1ap2D.ranks_sorted - full listing of residues and their ranking for 1ap2D
- 1ap2D.cbcvg - used by other 1ap2D – related pymol scripts