1a6v
Evolutionary trace report by report_maker
February 3, 2010

CONTENTS

1 Introduction

2 Chain 1a6vJ
  2.1 Q924Q7 overview
  2.2 Multiple sequence alignment for 1a6vJ
  2.3 Residue ranking in 1a6vJ
  2.4 Top ranking residues in 1a6vJ 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 1a6vL
  3.1 P01724 overview
  3.2 Multiple sequence alignment for 1a6vL
  3.3 Residue ranking in 1a6vL
  3.4 Top ranking residues in 1a6vL 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 1a6v):

**Title:** B1-8 fv fragment complexed with a (4-hydroxy-3-nitrophenyl) acetate compound

**Compound:** Mol id: 1; molecule: b1-8 fv (light chain); chain: l, m, n; fragment: fv fragment; engineered: yes; mol id: 2; molecule: b1-8 fv (heavy chain); chain: h, i, j; fragment: fv fragment; engineered: yes

**Organism, scientific name:** Mus Musculus;
1a6v contains unique chains 1a6vJ (120 residues) and 1a6vL (110 residues) 1a6vH and 1a6vI are homologues of chain 1a6vJ. 1a6vM and 1a6vN are homologues of chain 1a6vL.

2 CHAIN 1A6VJ

2.1 Q924Q7 overview

From SwissProt, id Q924Q7, 99% identical to 1a6vJ:

**Description:** VH186.2-D-J-C mu protein (Fragment).

**Organism, scientific name:** Mus musculus (Mouse).

**Taxonomy:** Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
2.2 Multiple sequence alignment for 1a6vJ

For the chain 1a6vJ, the alignment 1a6vJ.msf (attached) with 846 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 1a6vJ.msf. Its statistics, from the alistat program are the following:

<table>
<thead>
<tr>
<th>Format:</th>
<th>MSF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sequences:</td>
<td>846</td>
</tr>
<tr>
<td>Total number of residues:</td>
<td>89480</td>
</tr>
<tr>
<td>Smallest:</td>
<td>30</td>
</tr>
<tr>
<td>Largest:</td>
<td>120</td>
</tr>
<tr>
<td>Average length:</td>
<td>105.8</td>
</tr>
<tr>
<td>Alignment length:</td>
<td>120</td>
</tr>
<tr>
<td>Average identity:</td>
<td>40%</td>
</tr>
<tr>
<td>Most related pair:</td>
<td>99%</td>
</tr>
<tr>
<td>Most unrelated pair:</td>
<td>0%</td>
</tr>
<tr>
<td>Most distant seq:</td>
<td>32%</td>
</tr>
</tbody>
</table>

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

The alignment consists of 49% eukaryotic (49% 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 1a6vJ.descr.

2.3 Residue ranking in 1a6vJ

The 1a6vJ sequence is shown in Fig. 1, with each residue colored according to its estimated importance. The full listing of residues in 1a6vJ can be found in the file called 1a6vJ.ranks_sorted in the attachment.

2.4 Top ranking residues in 1a6vJ 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 1a6vJ 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>
<thead>
<tr>
<th>cluster</th>
<th>size</th>
<th>member residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>continued in next column</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Clusters of top ranking residues in 1a6vJ.

<table>
<thead>
<tr>
<th>cluster color</th>
<th>size</th>
<th>member residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>blue</td>
<td>2</td>
<td>325, 326</td>
</tr>
<tr>
<td>yellow</td>
<td>2</td>
<td>341, 342</td>
</tr>
</tbody>
</table>

Table 1. Continued

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 1a6vI.** By analogy with 1a6vH – 1a6vI interface. Table 2 lists the top 25% of residues at the interface with 1a6vI. The following table (Table 3) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 2. The top 25% of residues in 1a6vJ at the interface with 1a6vI. (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>356</td>
<td>G</td>
<td>90.(12) 08</td>
<td>0.12</td>
<td>4/4</td>
<td>4.03</td>
</tr>
</tbody>
</table>

Table 3. List of disruptive mutations for the top 25% of residues in 1a6vJ, that are at the interface with 1a6vI.

Figure 4 shows residues in 1a6vJ colored by their importance, at the interface with 1a6vI.

**Interface with 1a6vN.** Table 4 lists the top 25% of residues at the interface with 1a6vN. The following table (Table 5) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 4. Continued

<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>359</td>
<td>K</td>
<td>113.(8) 77</td>
<td>0.09</td>
<td>17/0</td>
<td>3.51</td>
</tr>
</tbody>
</table>

Table 4. Continued

<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>337</td>
<td>V</td>
<td>A(1)M</td>
<td>0.14</td>
<td>2/0</td>
<td>4.33</td>
</tr>
</tbody>
</table>

<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>395</td>
<td>Y</td>
<td>.(2)</td>
<td>0.20</td>
<td>45/0</td>
<td>3.14</td>
</tr>
</tbody>
</table>

<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>346</td>
<td>E</td>
<td>E(76) K(6)</td>
<td>0.24</td>
<td>1/1</td>
<td>4.70</td>
</tr>
</tbody>
</table>

continued in next column
Table 4. The top 25% of residues in 1a6vJ at the interface with 1a6vN. (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 5. List of disruptive mutations for the top 25% of residues in 1a6vJ, that are at the interface with 1a6vN.

Table 6. The top 25% of residues in 1a6vJ at the interface with NPC. (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 7. List of disruptive mutations for the top 25% of residues in 1a6vJ, that are at the interface with NPC.

NPC binding site. Table 6 lists the top 25% of residues at the interface with 1a6vNPC430 (npc). The following table (Table 7) suggests possible disruptive replacements for these residues (see Section 4.6).

Figure 5 shows residues in 1a6vJ colored by their importance, at the interface with 1a6vN. Figure 6 shows residues in 1a6vJ colored by their importance, at the interface with NPC.
Figure 6 shows residues in 1a6vJ colored by their importance, at the interface with 1a6vINPC430.

**NPC binding site.** By analogy with 1a6vI – 1a6vINPC430 interface. Table 8 lists the top 25% of residues at the interface with 1a6vINPC430 (npc). The following table (Table 9) suggests possible disruptive replacements for these residues (see Section 4.6).

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>subst’s</th>
<th>cvg</th>
<th>noc/ bb</th>
<th>dist (Å)</th>
<th>antn</th>
</tr>
</thead>
<tbody>
<tr>
<td>333</td>
<td>W</td>
<td>117.(7)</td>
<td>0.07</td>
<td>32/0</td>
<td>2.35</td>
<td>site</td>
</tr>
</tbody>
</table>

**Table 8.** The top 25% of residues in 1a6vJ at the interface with NPC.(Field names: res: residue number in the PDB entry; type: amino acid type; subst’s: substitutions seen in the alignment; cvg: percentage of each type in the bracket; noc/bb: number of contacts with the ligand, with the number of contacts realized through backbone atoms given in the bracket; dist: distance of closest approach to the ligand.)

**Table 9.** List of disruptive mutations for the top 25% of residues in 1a6vJ, that are at the interface with NPC.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>333</td>
<td>W</td>
<td>(KE)(TQD)(R)(SN)</td>
</tr>
</tbody>
</table>

2.4.3 Possible novel functional surfaces at 25% coverage. One group of residues is conserved on the 1a6vJ surface, away from (or substantially larger than) other functional sites and interfaces recognizable in PDB entry 1a6v. It is shown in Fig. 8. The right panel shows (in blue) the rest of the larger cluster this surface belongs to.

**Fig. 8.** A possible active surface on the chain 1a6vJ. The larger cluster it belongs to is shown in blue.

The residues belonging to this surface “patch” are listed in Table 10, while Table 11 suggests possible disruptive replacements for these residues (see Section 4.6).

**Fig. 7.** Residues in 1a6vJ, at the interface with NPC, colored by their relative importance. The ligand (NPC) 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 1a6vJ.)

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>390</td>
<td>D</td>
<td>(2)D(95)NBGVQH</td>
<td>0.03</td>
</tr>
<tr>
<td>339</td>
<td>Q</td>
<td>Q(91)K(4)HRI.LP EMZTSA</td>
<td>0.05</td>
</tr>
<tr>
<td>342</td>
<td>G</td>
<td>G(84)VD(3)AE(5) S(2)R(1).KPTN CLFVDX</td>
<td>0.07</td>
</tr>
<tr>
<td>392</td>
<td>A</td>
<td>A(2)A(86)G(9)ST</td>
<td>0.07</td>
</tr>
<tr>
<td>338</td>
<td>K</td>
<td>K(13)R(67)Q(12) V(2)CFHW.SNMPYL G</td>
<td>0.08</td>
</tr>
<tr>
<td>345</td>
<td>L</td>
<td>L(75)N(3)P(15)T R(1)VAI.EFXQHMS</td>
<td>0.09</td>
</tr>
<tr>
<td>386</td>
<td>L</td>
<td>A(2)L(75)A(1) S(1)V(14)M(2)I P(1)TRFGQ</td>
<td>0.10</td>
</tr>
<tr>
<td>341</td>
<td>P</td>
<td>P(84)H(2)S(4) A(1) L(3)TRE.IGN V</td>
<td>0.11</td>
</tr>
<tr>
<td>314</td>
<td>P</td>
<td>P(79)T(4)Q.(3) E(4)V(1)A(1) L(2)SKHRIX</td>
<td>0.12</td>
</tr>
<tr>
<td>315</td>
<td>G</td>
<td>G(75)K(1).R(2) T(6)S(11)NARDEP YNNX</td>
<td>0.12</td>
</tr>
<tr>
<td>391</td>
<td>S</td>
<td>S(2)S(22)E(7)</td>
<td>0.13</td>
</tr>
</tbody>
</table>

continued in next column
Table 10. Residues forming surface "patch" in 1a6vJ.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>337</td>
<td>V</td>
<td>T(59)IMA(2)F(3)</td>
<td>DCGKLR</td>
</tr>
<tr>
<td>321</td>
<td>S</td>
<td>S(56)N(6)QT(25)L(1)NH</td>
<td>.WQD</td>
</tr>
<tr>
<td>367</td>
<td>K</td>
<td>.(7)K(13)P(5)G</td>
<td>R(67)S(1)QDXELT</td>
</tr>
<tr>
<td>395</td>
<td>Y</td>
<td>.(2)Y(70)R(3)</td>
<td>F(19)CL(1)H(1)T</td>
</tr>
<tr>
<td>318</td>
<td>V</td>
<td>V(26)L(56)H(1)S</td>
<td>E(1)I(2)M(1)</td>
</tr>
<tr>
<td>320</td>
<td>L</td>
<td>I(19)L(70)V(4)</td>
<td>M(2)P(1)RSF</td>
</tr>
<tr>
<td>346</td>
<td>E</td>
<td>E(76)K(6)Q(4)</td>
<td>V(1)IDR(4)T(1).</td>
</tr>
</tbody>
</table>

Table 11. Disruptive mutations for the surface patch in 1a6vJ.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>390</td>
<td>D</td>
<td>(R)(FWH)(Y)(K)</td>
</tr>
<tr>
<td>394</td>
<td>Q</td>
<td>(Y)(H)(T)(FW)</td>
</tr>
<tr>
<td>342</td>
<td>G</td>
<td>(R)(K)(EH)(FW)</td>
</tr>
<tr>
<td>392</td>
<td>A</td>
<td>(R)(K)(E)(Y)</td>
</tr>
<tr>
<td>338</td>
<td>K</td>
<td>(Y)(T)(FW)(S)</td>
</tr>
<tr>
<td>345</td>
<td>L</td>
<td>(Y)(R)(H)(T)</td>
</tr>
<tr>
<td>386</td>
<td>L</td>
<td>(Y)(R)(H)(T)</td>
</tr>
<tr>
<td>341</td>
<td>P</td>
<td>(R)(Y)(H)(T)</td>
</tr>
<tr>
<td>314</td>
<td>P</td>
<td>(Y)(R)(H)(T)</td>
</tr>
<tr>
<td>315</td>
<td>G</td>
<td>(R)(K)(E)(H)</td>
</tr>
<tr>
<td>391</td>
<td>S</td>
<td>(R)(K)(H)(FW)</td>
</tr>
<tr>
<td>337</td>
<td>V</td>
<td>(R)(KE)(Y)(D)</td>
</tr>
<tr>
<td>321</td>
<td>S</td>
<td>(R)(K)(H)(FW)</td>
</tr>
<tr>
<td>367</td>
<td>K</td>
<td>(Y)(FW)(T)(CG)</td>
</tr>
<tr>
<td>395</td>
<td>Y</td>
<td>(K)(Q)(EM)(R)</td>
</tr>
<tr>
<td>318</td>
<td>V</td>
<td>(R)(Y)(K)(E)</td>
</tr>
<tr>
<td>320</td>
<td>L</td>
<td>(Y)(R)(T)(H)</td>
</tr>
<tr>
<td>346</td>
<td>E</td>
<td>(H)(FW)(Y)(R)</td>
</tr>
</tbody>
</table>

3 CHAIN 1A6VL

3.1 P01724 overview

From SwissProt, id P01724, 99% identical to 1a6vL:

Description: Ig lambda-1 chain V regions MOPC 104E/RPC20/J558/S104 precursor.

Organism, scientific name: Mus musculus (Mouse).

Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Mus.

Miscellaneous: Compositions and partial sequences of RPC 20 show no differences from MOPC 104E. The sequences of J558 and S104 seem identical with that shown.

Miscellaneous: These proteins were isolated from serum or urine of tumor-bearing mice.

Miscellaneous: The MOPC 104E precursor was synthesized in a cell-free system directed by mRNA isolated from MOPC 104E myeloma polysomes. Met-1 was lacking in 90 rapidly cleaved after synthesis.

Similarity: Contains 1 Ig-like (immunoglobulin-like) domain.

About: This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

3.2 Multiple sequence alignment for 1a6vL

For the chain 1a6vL, the alignment 1a6vL.msf (attached) with 706 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 1a6vL.msf. Its statistics, from the alstat program are the following:

Format: MSF
Number of sequences: 706
Total number of residues: 71042
Smallest: 59
Largest: 110
Average length: 100.6
Alignment length: 110
Average identity: 39%
Most related pair: 99%
Most unrelated pair: 10%
Most distant seq: 33%

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

The alignment consists of 44% eukaryotic (44% vertebrata), and <1% 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 1a6vL.descr.

3.3 Residue ranking in 1a6vL

The 1a6vL sequence is shown in Fig. 9, with each residue colored according to its estimated importance. The full listing of residues in 1a6vL can be found in the file called 1a6vL.ranks_sorted in the attachment.

6
3.4 Top ranking residues in 1a6vL and their position on the structure

In the following we consider residues ranking among top 25% of residues in the protein. Figure 10 shows residues in 1a6vL 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. 11 shows the top 25% of all residues, this time colored according to clusters they belong to. The clusters in Fig. 11 are composed of the residues listed in Table 12.

Table 12. Clusters of top ranking residues in 1a6vL.

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.

NPC binding site. Table 13 lists the top 25% of residues at the interface with 1a6vHNPC431 (npc). The following table (Table 14) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 13. The top 25% of residues in 1a6vL at the interface with NPC. (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; cvg: number of contacts with the ligand; noc/ bb: number of contacts realized through backbone atoms given in the bracket; dist: distance of closest approach to the ligand.)
Table 14. List of disruptive mutations for the top 25% of residues in 1a6vL, that are at the interface with NPC.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>Y</td>
<td>(K) (QM) (E) (R)</td>
</tr>
</tbody>
</table>

Table 15. The top 25% of residues in 1a6vL at the interface with NPC. (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; cvg: number of contacts with the ligand; noc/bb: 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>
<th>antn</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>Y</td>
<td>107.(2)</td>
<td>0.02</td>
<td>36/0</td>
<td>3.32</td>
<td>site</td>
</tr>
</tbody>
</table>

NPC binding site. Table 15 lists the top 25% of residues at the interface with 1a6vHNPC430 (npc). The following table (Table 16) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 16. List of disruptive mutations for the top 25% of residues in 1a6vL, that are at the interface with NPC.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>Y</td>
<td>(K) (QM) (E) (R)</td>
</tr>
</tbody>
</table>

Fig. 12. Residues in 1a6vL, at the interface with NPC, colored by their relative importance. The ligand (NPC) 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 1a6vL.)

Figure 12 shows residues in 1a6vL colored by their importance, at the interface with 1a6vHNPC431.

Fig. 13. Residues in 1a6vL, at the interface with NPC, colored by their relative importance. The ligand (NPC) 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 1a6vL.)

Interface with 1a6vH. Table 17 lists the top 25% of residues at the interface with 1a6vH. The following table (Table 18) suggests possible disruptive replacements for these residues (see Section 4.6).

Table 17. The top 25% of residues in 1a6vL at the interface with NPC. (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; cvg: number of contacts with the ligand; noc/bb: 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>
<th>antn</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>A</td>
<td>62.81</td>
<td>0.01</td>
<td>15/11</td>
<td>3.66</td>
<td>site</td>
</tr>
<tr>
<td>34</td>
<td>Y</td>
<td>107.(2)</td>
<td>0.02</td>
<td>17/1</td>
<td>3.09</td>
<td>site</td>
</tr>
<tr>
<td>36</td>
<td>N</td>
<td>107.(1)</td>
<td>0.02</td>
<td>47/0</td>
<td>2.90</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>F</td>
<td>P (65) M (2) E L (22) I (1)</td>
<td>0.10</td>
<td>72/15</td>
<td>3.41</td>
<td></td>
</tr>
</tbody>
</table>

continued in next column
Table 17. The top 25% of residues in 1a6vL at the interface with 1a6vH. (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 18. List of disruptive mutations for the top 25% of residues in 1a6vL, that are at the interface with 1a6vH.

Figure 14 shows residues in 1a6vL colored by their importance, at the interface with 1a6vH.

3.4.3 Possible novel functional surfaces at 25% coverage. One group of residues is conserved on the 1a6vL surface, away from (or substantially larger than) other functional sites and interfaces recognizable in PDB entry 1a6v. It is shown in Fig. 15. 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 19, while Table 20 suggests possible disruptive replacements for these residues (see Section 4.6).
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 10 Å², 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 Å 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 4 Å.

4.5 Annotation
If the residue annotation is available (either from the pdb file or from other sources), another column, with the header “annotation” 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%.

4.6 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 [LPNQDEM1K], large

Table 19. continued

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>S</td>
<td>K(1) S(73) T(15) E</td>
<td>0.14</td>
</tr>
<tr>
<td>43</td>
<td>D</td>
<td>G(81) K(3) H(1) R</td>
<td>0.15</td>
</tr>
<tr>
<td>66</td>
<td>G</td>
<td>G(70) V(5) K(4) I</td>
<td>0.16</td>
</tr>
<tr>
<td>39</td>
<td>Q</td>
<td>Q(62) L(7) R(19) K</td>
<td>0.17</td>
</tr>
<tr>
<td>85</td>
<td>E</td>
<td>A(11) E(25) L(4) S</td>
<td>0.21</td>
</tr>
<tr>
<td>71</td>
<td>N</td>
<td>T(48) N(25) G(1) S</td>
<td>0.22</td>
</tr>
<tr>
<td>69</td>
<td>I</td>
<td>S(67) Q(1) D(9) N</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 19. Residues forming surface “patch” in 1a6vL.

Table 20. Disruptive mutations for the surface patch in 1a6vL.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>84</td>
<td>D</td>
<td>(R) (H) (FW) (Y)</td>
</tr>
<tr>
<td>63</td>
<td>R</td>
<td>(D) (TY) (E) (VCLAPIG)</td>
</tr>
<tr>
<td>86</td>
<td>A</td>
<td>(R) (K) (Y) (H)</td>
</tr>
<tr>
<td>64</td>
<td>F</td>
<td>(K) (E) (T) (Q)</td>
</tr>
<tr>
<td>46</td>
<td>F</td>
<td>(E) (K) (T) (D)</td>
</tr>
<tr>
<td>67</td>
<td>S</td>
<td>(R) (K) (FW) (M)</td>
</tr>
<tr>
<td>40</td>
<td>E</td>
<td>(H) (FW) (Y) (R)</td>
</tr>
<tr>
<td>42</td>
<td>P</td>
<td>(Y) (R) (H) (T)</td>
</tr>
<tr>
<td>65</td>
<td>S</td>
<td>(R) (K) (H) (Y)</td>
</tr>
<tr>
<td>43</td>
<td>D</td>
<td>(R) (H) (FW) (Y)</td>
</tr>
<tr>
<td>66</td>
<td>G</td>
<td>(R) (K) (E) (H)</td>
</tr>
<tr>
<td>39</td>
<td>Q</td>
<td>(Y) (T) (H) (FW)</td>
</tr>
<tr>
<td>85</td>
<td>E</td>
<td>(H) (R) (FW) (Y)</td>
</tr>
<tr>
<td>71</td>
<td>N</td>
<td>(Y) (H) (FW) (T)</td>
</tr>
<tr>
<td>69</td>
<td>I</td>
<td>(Y) (R) (H) (T)</td>
</tr>
</tbody>
</table>

Table 20. Disruptive mutations for the surface patch in 1a6vL.
5.2 Color schemes used
The following color scheme is used in figures with residues colored by cluster size: black is a single-residue cluster; clusters composed of more than one residue colored according to this hierarchy (ordered by descending size): red, blue, yellow, green, purple, azure, turquoise, brown, coral, magenta, LightSalmon, SkyBlue, violet, gold, bisque, LightSlateBlue, orchid, RosyBrown, MediumAquamarine, DarkOliveGreen, CornflowerBlue, grey55, burlywood, LimeGreen, tan, DarkOrange, maroon, BlanchedAlmond.

The colors used to distinguish the residues by the estimated evolutionary pressure they experience can be seen in Fig. 16.

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, 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://www.cmbi.kun.nl/gv/dssp/descrip.html.


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).


5.3.7 Pymol The figures in this report were produced using Pymol. The scripts can be found in the attachment. Pymol

Fig. 16. Coloring scheme used to color residues by their relative importance.

[WFYHR], hydrophobic [LPVAMWFTE], polar [GTCY]; positively [KHR], or negatively [DE] charged, aromatic [WFYH], long aliphatic chain [EKRMQ], OH-group possession [SDET], 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
- 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
- gap 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
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 reportmaker

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

5.7 Attachments

The following files should accompany this report:

- 1a6vJ.complex.pdb - coordinates of 1a6vJ with all of its interacting partners
- 1a6vJ.etvx - ET viewer input file for 1a6vJ
- 1a6vJ.cluster_report.summary - Cluster report summary for 1a6vJ
- 1a6vJ.ranks - Ranks file in sequence order for 1a6vJ
- 1a6vJ.clusters - Cluster descriptions for 1a6vJ
- 1a6vJ.msf - the multiple sequence alignment used for the chain 1a6vJ
- 1a6vJ.descr - description of sequences used in 1a6vJ msf
- 1a6vJ.ranks_sorted - full listing of residues and their ranking for 1a6vJ
- 1a6vJ.1a6vL.etvx - Pymol script for Figure 4
- 1a6vJ.cbcvg - used by other 1a6vJ – related pymol scripts
- 1a6vJ.1a6vN.etvx - Pymol script for Figure 5
- 1a6vJ.1a6vJNPC430.etvx - Pymol script for Figure 6
- 1a6vJ.1a6vL.cluster_report.summary - Cluster report summary for 1a6vL
- 1a6vJ.ranks - Ranks file in sequence order for 1a6vL
- 1a6vJ.clusters - Cluster descriptions for 1a6vL
- 1a6vL.msf - the multiple sequence alignment used for the chain 1a6vL
- 1a6vL.descr - description of sequences used in 1a6vL msf
- 1a6vL.ranks_sorted - full listing of residues and their ranking for 1a6vL
- 1a6vL.1a6vHNPC431.etvx - Pymol script for Figure 12
- 1a6vL.cbcvg - used by other 1a6vL – related pymol scripts
- 1a6vL.1a6vHNPC430.etvx - Pymol script for Figure 13
- 1a6vL.1a6vH.etvx - Pymol script for Figure 14