CONTENTS

1 Introduction

2 Chain 1amwA
   2.1 P02829 overview
   2.2 Multiple sequence alignment for 1amwA
   2.3 Residue ranking in 1amwA
   2.4 Top ranking residues in 1amwA and their position on the structure
      2.4.1 Clustering of residues at 25% coverage.
      2.4.2 Overlap with known functional surfaces at 25% coverage.
      2.4.3 Possible novel functional surfaces at 25% coverage.

3 Notes on using trace results
   3.1 Coverage
   3.2 Known substitutions
   3.3 Surface
   3.4 Number of contacts
   3.5 Annotation
   3.6 Mutation suggestions

4 Appendix
   4.1 File formats
   4.2 Color schemes used
   4.3 Credits

4.3.1 Alistat
4.3.2 CE
4.3.3 DSSP
4.3.4 HSSP
4.3.5 LaTeX
4.3.6 Muscle
4.3.7 Pymol

4.4 Note about ET Viewer
4.5 Citing this work
4.6 About report_maker
4.7 Attachments

1 INTRODUCTION

From the original Protein Data Bank entry (PDB id 1amw):
Title: Adp binding site in the hsp90 molecular chaperone
Compound: Mol id: 1; molecule: heat shock protein 90; chain: a; fragment: n-terminal residues; synonym: hsp90; engineered: yes; other details: adp complex
Organism, scientific name: Saccharomyces Cerevisiae;
1amw contains a single unique chain 1amwA (213 residues long).

2 CHAIN 1AMWA

2.1 P02829 overview

From SwissProt, id P02829, 100% identical to 1amwA:
Description: ATP-dependent molecular chaperone HSP82 (Heat shock protein Hsp90 heat inducible isoform) (82 kDa heat shock protein).
Organism, scientific name: Saccharomyces cerevisiae (Baker’s yeast).
Taxonomy: Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomyces; Saccharomyces.
Function: Molecular chaperone that promotes the maturation, structural maintenance and proper regulation of specific target proteins involved in cell cycle control and signal transduction. Undergoes a functional cycle that is linked to its ATPase activity. The nucleotide-free form of the dimer is found in an open conformation in which the N-termini are not dimerized and the complex is ready for client protein binding. Binding of ATP induces large conformational changes, resulting in the formation of a ring-like closed structure in which the N-terminal domains associate intramolecularly with the middle domain and also dimerize with each other, stimulating their intrinsic ATPase activity and acting as a clamp on the substrate. Finally, ATP hydrolysis results in the release of the substrate. This cycle probably induces conformational changes in the client proteins, thereby causing their activation. Interacts dynamically with various co-chaperones that modulate its substrate recognition, ATPase cycle and chaperone function. Required for growth at high temperatures.

1 Lichtarge lab 2006
Enzyme regulation: Inhibited by geldanamycin, macbectin I and radicicol, which bind to the ATP-binding pocket. Co-chaperones CDC37, SBA1 and STI1 reduce ATPase activity. Co-chaperones AHA1 and HCH1 increase ATPase activity.

Subunit: Homodimer. Interacts with the co-chaperones AHA1, CDC37, CNS1, CPR6, CPR7, HCH1, SBA1, SSE1 and STI1. CNS1, CPR6, CPR7 and STI1 bind with their TPR repeats to the N-terminal pentapeptide MEEVD. Interacts directly with the substrates GCN2, HAP1 and STE11.

Interaction:
Subcellular location: Cytoplasmic.
Induction: Expressed constitutively and induced by high temperatures dependent on transcription factor HSF1. According to Ref.3, it is constitutively expressed at low levels, however, due to the specificity of the antibody, this result is unsure.

Miscellaneous: Present with 444943 molecules/cell.
Similarity: Belongs to the heat shock protein 90 family.
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.

2.2 Multiple sequence alignment for 1amwA

For the chain 1amwA, the alignment 1amwA.msf (attached) with 849 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 1amwA.msf. Its statistics, from the alistat program are the following:

Format: MSF
Number of sequences: 849
Total number of residues: 174119
Smallest: 163
Largest: 213
Average length: 205.1
Alignment length: 213
Average identity: 51%
Most related pair: 99%
Most unrelated pair: 16%
Most distant seq: 49%

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

The alignment consists of 24% eukaryotic (3% vertebrata, 2% arthropoda, 2% fungi, 2% plantae), and 5% prokaryotic sequences. (Descriptions of some sequences were not readily available.) The file containing the sequence descriptions can be found in the attachment, under the name 1amwA.descr.

2.3 Residue ranking in 1amwA

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

2.4 Top ranking residues in 1amwA and their position on the structure

In the following we consider residues ranking among top 25% of residues in the protein. Figure 3 shows residues in 1amwA 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. 4 shows the top 25% of all residues, this time colored according to clusters they belong to. The clusters in Fig. 4 are composed of the residues listed in Table 1.

![Fig. 4. Residues in 1amwA, 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.](image)

<table>
<thead>
<tr>
<th>Cluster color</th>
<th>Size</th>
<th>Member Residues</th>
</tr>
</thead>
</table>

### Table 1.

Clustering of top-ranking residues in 1amwA.

### 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.

**ADP binding site.** Table 2 lists the top 25% of residues at the interface with 1amwADP300 (adp). The following table (Table 3) suggests possible disruptive replacements for these residues (see Section 3.6).

<table>
<thead>
<tr>
<th>Residue</th>
<th>Type</th>
<th>Substitutions</th>
<th>Cvg (%)</th>
<th>Number of Contacts</th>
<th>Distance (Å)</th>
<th>Antin.</th>
</tr>
</thead>
<tbody>
<tr>
<td>121</td>
<td>G</td>
<td>G(99)LS</td>
<td>0.01</td>
<td>21/21</td>
<td>2.82</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>G</td>
<td>G(99)YS</td>
<td>0.02</td>
<td>10/10</td>
<td>3.51</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>D</td>
<td>D(99)TN</td>
<td>0.04</td>
<td>8/0</td>
<td>2.97 site</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>G</td>
<td>G(99).V</td>
<td>0.04</td>
<td>7/7</td>
<td>4.32</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>G</td>
<td>G(99)SR</td>
<td>0.05</td>
<td>1/1</td>
<td>4.43</td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>F</td>
<td>F(99)PL</td>
<td>0.05</td>
<td>23/8</td>
<td>2.98</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>Y</td>
<td>Y(98)FL</td>
<td>0.06</td>
<td>2/0</td>
<td>4.12</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>D</td>
<td>D(98).LY</td>
<td>0.07</td>
<td>1/0</td>
<td>4.55</td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>V</td>
<td>V(97)I</td>
<td>0.07</td>
<td>19/17</td>
<td>3.15</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>N</td>
<td>N(97)K</td>
<td>0.08</td>
<td>37/10</td>
<td>2.98 site</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>M</td>
<td>M(98)L</td>
<td>0.09</td>
<td>23/0</td>
<td>3.58</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>A</td>
<td>A(98).S</td>
<td>0.12</td>
<td>7/2</td>
<td>3.36</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>L</td>
<td>L(92)I</td>
<td>0.13</td>
<td>9/1</td>
<td>3.38</td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>T</td>
<td>T(95)S</td>
<td>0.18</td>
<td>11/0</td>
<td>3.58</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** The top 25% of residues in 1amwA at the interface with ADP. (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>Residue</th>
<th>Type</th>
<th>Disruptive Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>121</td>
<td>G</td>
<td>(R) (K) (E) (H)</td>
</tr>
<tr>
<td>123</td>
<td>G</td>
<td>(K) (R) (E) (Q)</td>
</tr>
<tr>
<td>79</td>
<td>D</td>
<td>(R) (FWH) (Y) (K)</td>
</tr>
<tr>
<td>83</td>
<td>G</td>
<td>(KR) (E) (Q) (H)</td>
</tr>
<tr>
<td>118</td>
<td>G</td>
<td>(E) (K) (FMWR) (QH)</td>
</tr>
<tr>
<td>124</td>
<td>F</td>
<td>(KE) (T) (QDR) (SCG)</td>
</tr>
<tr>
<td>125</td>
<td>Y</td>
<td>(K) (Q) (R) (E)</td>
</tr>
<tr>
<td>40</td>
<td>D</td>
<td>(R) (FWH) (VA) (K)</td>
</tr>
<tr>
<td>122</td>
<td>V</td>
<td>(R) (Y) (KE) (H)</td>
</tr>
<tr>
<td>37</td>
<td>N</td>
<td>(Y) (FTW) (H) (VCA)</td>
</tr>
<tr>
<td>84</td>
<td>M</td>
<td>(Y) (T) (H) (R)</td>
</tr>
<tr>
<td>41</td>
<td>A</td>
<td>(KR) (YE) (H) (Q)</td>
</tr>
<tr>
<td>93</td>
<td>L</td>
<td>(R) (Y) (T) (KEH)</td>
</tr>
</tbody>
</table>

**Table 3.**

Disruptive mutations

---

continued in next column

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

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>171</td>
<td>T</td>
<td>(R) (K) (H) (FQW)</td>
</tr>
</tbody>
</table>

Table 4. The top 25% of residues in 1amwA at the interface with 1amwA1. (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>209</td>
<td>V</td>
<td>(KER) (Y) (Q) (D)</td>
<td>0.08</td>
<td>39/32</td>
<td>2.60</td>
</tr>
<tr>
<td>208</td>
<td>V</td>
<td>(KER) (Y) (Q) (D)</td>
<td>0.09</td>
<td>25/13</td>
<td>3.21</td>
</tr>
<tr>
<td>204</td>
<td>P</td>
<td>(Y) (R) (H) (T)</td>
<td>0.21</td>
<td>17/8</td>
<td>3.35</td>
</tr>
</tbody>
</table>

Table 5. List of disruptive mutations for the top 25% of residues in 1amwA, that are at the interface with 1amwA1.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>209</td>
<td>V</td>
<td>(KER) (Y) (Q) (D)</td>
</tr>
<tr>
<td>208</td>
<td>V</td>
<td>(KER) (Y) (Q) (D)</td>
</tr>
<tr>
<td>204</td>
<td>P</td>
<td>(Y) (R) (H) (T)</td>
</tr>
</tbody>
</table>

Fig. 5. Residues in 1amwA, at the interface with ADP, colored by their relative importance. The ligand (ADP) is colored green. Atoms further than 30 Å away from the geometric center of the ligand, as well as on the line of sight to the ligand were removed. (See Appendix for the coloring scheme for the protein chain 1amwA.)

Figure 5 shows residues in 1amwA colored by their importance, at the interface with 1amwADP300.

Interface with 1amwA1. Table 4 lists the top 25% of residues at the interface with 1amwA1. The following table (Table 5) suggests possible disruptive replacements for these residues (see Section 3.6).

Fig. 6. Residues in 1amwA, at the interface with 1amwA1, colored by their relative importance. 1amwA1 is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 1amwA.)

Figure 6 shows residues in 1amwA colored by their importance, at the interface with 1amwA1.

2.4.3 Possible novel functional surfaces at 25% coverage. One group of residues is conserved on the 1amwA surface, away from (or substantially larger than) other functional sites and interfaces recognizable in PDB entry 1amw. It is shown in Fig. 7. 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
Fig. 7. A possible active surface on the chain 1amwA. The larger cluster it belongs to is shown in blue.

Table 6.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
<th>antn</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>S</td>
<td>S(99)F</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>G</td>
<td>G(99)Ls.</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>G</td>
<td>G(99)YS.A</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>A</td>
<td>A(99)TG</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>D</td>
<td>D(99)TN</td>
<td>0.04</td>
<td>site</td>
</tr>
<tr>
<td>83</td>
<td>G</td>
<td>G(99).VXSA</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>G</td>
<td>G(99)SR.</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>F</td>
<td>F(99)PL.</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>F</td>
<td>F(99)Ls.</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>Y</td>
<td>Y(98)FIL.</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>D</td>
<td>D(98),(1)Y</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>V</td>
<td>V(97)I(1)LG.</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>N</td>
<td>N(97)K.(2)</td>
<td>0.08</td>
<td>site</td>
</tr>
<tr>
<td>84</td>
<td>M</td>
<td>M(98)L(1).FV</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>G</td>
<td>G(99)QEDSA</td>
<td>0.09</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>S</td>
<td>S(96),(2)A(1)P</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>R</td>
<td>R(96),(2)WLQF</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>Q</td>
<td>Q(96)K(1)HMRE.A</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>E</td>
<td>E(96),(2)DGVK</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>A</td>
<td>A(98),(1)S</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>L</td>
<td>L(92)I(7)VF</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td>198</td>
<td>S</td>
<td>S(96),(1)CVP</td>
<td>0.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A(1)TY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>I</td>
<td>I(95)L(1)Y(2)T</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>K</td>
<td>K(95)R(2),(1)EN</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>T</td>
<td>T(95)V(2)QSIGKA</td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>T</td>
<td>T(95)S(2)VK.WAN</td>
<td>0.18</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>L</td>
<td>L(94),(2)I(2)VS</td>
<td>0.20</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>D</td>
<td>D(85)E(10),(1)L</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>R</td>
<td>R(87)HY(3),(1)</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>K(2)Q(1)N(2)VLS</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

continued in next column

Table 7.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>99</td>
<td>S</td>
<td>(K)(R)(Q)(EM)</td>
</tr>
<tr>
<td>121</td>
<td>G</td>
<td>(R)(K)(E)(H)</td>
</tr>
<tr>
<td>123</td>
<td>G</td>
<td>(K)(R)(E)(Q)</td>
</tr>
<tr>
<td>97</td>
<td>A</td>
<td>(KR)(E)(QH)(Y)</td>
</tr>
<tr>
<td>79</td>
<td>D</td>
<td>(R)(FWH)(Y)(K)</td>
</tr>
<tr>
<td>83</td>
<td>G</td>
<td>(KR)(E)(Q)(H)</td>
</tr>
<tr>
<td>118</td>
<td>G</td>
<td>(E)(K)(FMWDR)(QH)</td>
</tr>
<tr>
<td>124</td>
<td>F</td>
<td>(KE)(T)(QDR)(SCG)</td>
</tr>
<tr>
<td>120</td>
<td>F</td>
<td>(K)(E)(Q)(TDR)</td>
</tr>
<tr>
<td>125</td>
<td>Y</td>
<td>(Y)(FTW)(H)(VCAG)</td>
</tr>
<tr>
<td>37</td>
<td>N</td>
<td>(Y)(FTW)(H)(VCAG)</td>
</tr>
<tr>
<td>84</td>
<td>M</td>
<td>(Y)(T)(H)(R)</td>
</tr>
<tr>
<td>100</td>
<td>G</td>
<td>(R)(H)(K)(FW)</td>
</tr>
<tr>
<td>36</td>
<td>S</td>
<td>(R)(K)(H)(Q)</td>
</tr>
<tr>
<td>32</td>
<td>R</td>
<td>(T)(D)(S)(ECG)</td>
</tr>
<tr>
<td>119</td>
<td>Q</td>
<td>(Y)(T)(FW)(H)</td>
</tr>
<tr>
<td>33</td>
<td>E</td>
<td>(FWH)(Y)(R)(VCAG)</td>
</tr>
<tr>
<td>41</td>
<td>A</td>
<td>(KR)(YE)(H)(Q)</td>
</tr>
<tr>
<td>93</td>
<td>L</td>
<td>(R)(Y)(T)(KEH)</td>
</tr>
<tr>
<td>198</td>
<td>S</td>
<td>(KR)(Q)(H)(M)</td>
</tr>
<tr>
<td>96</td>
<td>I</td>
<td>(R)(Y)(H)(K)</td>
</tr>
<tr>
<td>44</td>
<td>K</td>
<td>(Y)(FW)(T)(VCAG)</td>
</tr>
<tr>
<td>95</td>
<td>T</td>
<td>(R)(H)(K)(FW)</td>
</tr>
<tr>
<td>171</td>
<td>T</td>
<td>(R)(K)(H)(FW)</td>
</tr>
<tr>
<td>34</td>
<td>L</td>
<td>(R)(Y)(H)(K)</td>
</tr>
<tr>
<td>43</td>
<td>D</td>
<td>(R)(H)(FW)(Y)</td>
</tr>
<tr>
<td>46</td>
<td>R</td>
<td>(TD)(E)(Y)(CG)</td>
</tr>
<tr>
<td>197</td>
<td>H</td>
<td>(E)(T)(D)(S)</td>
</tr>
<tr>
<td>39</td>
<td>S</td>
<td>(K)(R)(Q)(EM)</td>
</tr>
<tr>
<td>116</td>
<td>M</td>
<td>(Y)(H)(TR)(S)</td>
</tr>
<tr>
<td>200</td>
<td>F</td>
<td>(E)(K)(T)(D)</td>
</tr>
</tbody>
</table>

continued in next column
3 NOTES ON USING TRACE RESULTS

3.1 Coverage

Trace results are commonly expressed in terms of coverage: the residue is important if its “coverage” is small - that is if it belongs to some small top percentage of residues [100% is all of the residues in a chain], according to trace. The ET results are presented in the form of a table, usually limited to top 25% percent of residues (or to some nearby percentage), sorted by the strength of the presumed evolutionary pressure. (i.e., the smaller the coverage, the stronger the pressure on the residue.) Starting from the top of that list, mutating a couple of residues should affect the protein somehow, with the exact effects to be determined experimentally.

3.2 Known substitutions

One of the table columns is “substitutions” - other amino acid types seen at the same position in the alignment. These amino acid types may be interchangeable at that position in the protein, so if one wants to affect the protein by a point mutation, they should be avoided. For example if the substitutions are “RVK” and the original protein has an R at that position, it is advisable to try anything, but RVK. Conversely, when looking for substitutions which will not affect the protein, one may try replacing R with K, or (perhaps more surprisingly), with V. The percentage of times the substitution appears in the alignment is given in the immediately following bracket. No percentage is given in the cases when it is smaller than 1%. This is meant to be a rough guide - due to rounding errors these percentages often do not add up to 100%.

3.3 Surface

To detect candidates for novel functional interfaces, first we look for residues that are solvent accessible (according to DSSP program) by at least 10 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.)

3.4 Number of contacts

Another column worth noting is denoted “noc/bb”; it tells the number of contacts heavy atoms of the residue in question make across the interface, as well as how many of them are realized through the backbone atoms (if all or most contacts are through the backbone, mutation presumably won’t have strong impact). Two heavy atoms are considered to be “in contact” if their centers are closer than 5 Å.

3.5 Annotation

If the residue annotation is available (either from the pdb file or from other sources), another column, with the header “annotation” appears. Annotations carried over from PDB are the following: site (indicating existence of related site record in PDB ), S-S (disulfide bond forming residue), hb (hydrogen bond forming residue), jb (james bond forming residue), and sb (for salt bridge forming residue).

3.6 Mutation suggestions

Mutation suggestions are completely heuristic and based on complementarity with the substitutions found in the alignment. Note that they are meant to be disruptive to the interaction of the protein with its ligand. The attempt is made to complement the following properties: small [AVGSTC], medium [LPQDEM1K], large [WYHR], hydrophobic [LPVAMWF], polar [GYCT]; positively [KDR], or negatively [DE] charged, aromatic [FYWH], long aliphatic chain [EQRQ], OH-group possession [DETY], and NH2 group possession [NQRK]. The suggestions are listed according to how different they appear to be from the original amino acid, and they are grouped in round brackets if they appear equally disruptive. From left to right, each bracketed group of amino acid types resembles more strongly the original (i.e. is, presumably, less disruptive) These suggestions are tentative - they might prove disruptive to the fold rather than to the interaction. Many researcher will choose, however, the straightforward alanine mutations, especially in the beginning stages of their investigation.

4 APPENDIX

4.1 File formats

Files with extension “ranks_sorted” are the actual trace results. The fields in the table in this file:

- alignment# number of the position in the alignment
- residue# residue number in the PDB file
- type amino acid type
- rank rank of the position according to older version of ET
- variability has two subfields:
  1. number of different amino acids appearing in in this column of the alignment
  2. their type
- rho ET score - the smaller this value, the lesser variability of this position across the branches of the tree (and, presumably, the greater the importance for the protein)
- cvg coverage - percentage of the residues on the structure which have this rho or smaller
- gaps percentage of gaps in this column

4.2 Color schemes used

The following color scheme is used in figures with residues colored by cluster size: black is a single-residue cluster; clusters composed of more than one residue colored according to this hierarchy (ordered by descending size): red, blue, yellow, green, purple, azure, turquoise, brown, coral, magenta, LightSalmon, SkyBlue, violet, gold, bisque, LightSlateBlue, orchid, RosyBrown, MediumAquamarine, DarkOliveGreen, CornflowerBlue, grey55, burlywood, LimeGreen, tan, DarkOrange, DeepPink, maroon, BlanchedAlmond.
The colors used to distinguish the residues by the estimated evolutionary pressure they experience can be seen in Fig. 8.

4.3 Credits

4.3.1 Alistat  *alistat* reads a multiple sequence alignment from the file and shows a number of simple statistics about it. These statistics include the format, the number of sequences, the total number of residues, the average and range of the sequence lengths, and the alignment length (e.g. including gap characters). Also shown are some percent identities. A percent pairwise alignment identity is defined as (idents / MIN(len1, len2)) where idents is the number of exact identities and len1, len2 are the unaligned lengths of the two sequences. The "average percent identity", "most related pair", and "most unrelated pair" of the alignment are the average, maximum, and minimum of all (N)(N-1)/2 pairs, respectively. The "most distant seq" is calculated by finding the maximum pairwise identity (best relative) for all N sequences, then finding the minimum of these N numbers (hence, the most outlying sequence). *alistat* is copyrighted by HHMI/Washington University School of Medicine, 1992-2001, and freely distributed under the GNU General Public License.

4.3.2 CE  To map ligand binding sites from different source structures, report_maker uses the CE program: [http://cl.sdsc.edu/](http://cl.sdsc.edu/). Shindyalov IN, Bourne PE (1998) "Protein structure alignment by incremental combinatorial extension (CE) of the optimal path". Protein Engineering 11(9) 739-747.

4.3.3 DSSP  In this work a residue is considered solvent accessible if the DSSP program finds it exposed to water by at least 10Å², 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](http://www.cmbi.kun.nl/gv/dssp/descrip.html).


4.3.5 LaTeX  The text for this report was processed using `LaTeX`; Leslie Lamport, “LaTeX: A Document Preparation System Addison-Wesley,” Reading, Mass. (1986).


4.3.7 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/](http://pymol.sourceforge.net/). (Note for Windows users: the attached package needs to be unzipped for Pymol to read the scripts and launch the viewer.)

4.4 Note about ET Viewer  Dan Morgan from the Lichtarge lab has developed a visualization tool specifically for viewing trace results. If you are interested, please visit: [http://mammoth.bcm.tmc.edu/traceview/](http://mammoth.bcm.tmc.edu/traceview/)

The viewer is self-unpacking and self-installing. Input files to be used with ETV (extension .etvx) can be found in the attachment to the main report.


4.6 About report_maker  report_maker was written in 2006 by Ivana Mihalek. The 1D ranking visualization program was written by Ivica Reˇs. report_maker is copyrighted by Lichtarge Lab, Baylor College of Medicine, Houston.

4.7 Attachments  The following files should accompany this report:

- 1amwA.complex.pdb - coordinates of 1amwA with all of its interacting partners
- 1amwA.etvx - ET viewer input file for 1amwA
- 1amwA.cluster_report.summary - Cluster report summary for 1amwA
- 1amwA.ranks - Ranks file in sequence order for 1amwA
- 1amwA.clusters - Cluster descriptions for 1amwA
- 1amwA.msf - the multiple sequence alignment used for the chain 1amwA
- 1amwA.deser - description of sequences used in 1amwA msf
- 1amwA.ranks_sorted - full listing of residues and their ranking for 1amwA
- 1amwA.1amwADP300.if.pml - Pymol script for Figure 5
- 1amwA.cbcvg - used by other 1amwA – related pymol scripts
- 1amwA.1amwA1.if.pml - Pymol script for Figure 6