## 1 INTRODUCTION

From the original Protein Data Bank entry (PDB id 1abv):

**Title:** N-terminal domain of the delta subunit of the f1f0-atp synthase from escherichia coli, nmr, minimized average structure

**Compound:** Mol id: 1; molecule: delta subunit of the f1f0-atp synthase; chain: a; fragment: n-terminal domain, residues 1 - 134; ec: 3.6.1.34; engineered: yes

**Organism, scientific name:** Escherichia Coli;

1abv contains a single unique chain 1abvA (105 residues long). This is an NMR-determined structure – in this report the first model in the file was used.

### 2 CHAIN 1ABVA

#### 2.1 Q9RFL6 overview

From SwissProt, id Q9RFL6, 100% identical to 1abvA:

**Description:** ATP synthase delta subunit.

**Organism, scientific name:** Salmonella typhimurium.

**Taxonomy:** Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.

#### 2.2 Multiple sequence alignment for 1abvA

For the chain 1abvA, the alignment 1abvA.msf (attached) with 181 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 1abvA.msf. Its statistics, from the alstat program are the following:

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSSP</td>
<td>4</td>
</tr>
<tr>
<td>HSSP</td>
<td>4</td>
</tr>
<tr>
<td>LaTeX</td>
<td>4</td>
</tr>
<tr>
<td>Muscle</td>
<td>4</td>
</tr>
<tr>
<td>Pymol</td>
<td>4</td>
</tr>
<tr>
<td>Note about ET Viewer</td>
<td>4</td>
</tr>
<tr>
<td>Citing this work</td>
<td>4</td>
</tr>
<tr>
<td>About report_maker</td>
<td>5</td>
</tr>
<tr>
<td>Attachments</td>
<td>5</td>
</tr>
</tbody>
</table>

### CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Chain 1abvA</td>
<td>2</td>
</tr>
<tr>
<td>Notes on using trace results</td>
<td>3</td>
</tr>
<tr>
<td>Appendix</td>
<td>4</td>
</tr>
</tbody>
</table>

#### 1 INTRODUCTION

From the original Protein Data Bank entry (PDB id 1abv):

**Title:** N-terminal domain of the delta subunit of the f1f0-atp synthase from escherichia coli, nmr, minimized average structure

**Compound:** Mol id: 1; molecule: delta subunit of the f1f0-atp synthase; chain: a; fragment: n-terminal domain, residues 1 - 134; ec: 3.6.1.34; engineered: yes

**Organism, scientific name:** Escherichia Coli;

1abv contains a single unique chain 1abvA (105 residues long). This is an NMR-determined structure – in this report the first model in the file was used.

#### 2 CHAIN 1ABVA

#### 2.1 Q9RFL6 overview

From SwissProt, id Q9RFL6, 100% identical to 1abvA:

**Description:** ATP synthase delta subunit.

**Organism, scientific name:** Salmonella typhimurium.

**Taxonomy:** Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.

#### 2.2 Multiple sequence alignment for 1abvA

For the chain 1abvA, the alignment 1abvA.msf (attached) with 181 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 1abvA.msf. Its statistics, from the alstat program are the following:
2.3 Residue ranking in 1abvA
The 1abvA sequence is shown in Fig. 1, with each residue colored according to its estimated importance. The full listing of residues in 1abvA can be found in the file called 1abvA.ranks, sorted in the attachment.

2.4 Top ranking residues in 1abvA 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 1abvA 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 color</th>
<th>size</th>
<th>member residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td>24</td>
<td>5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18, 27, 31, 74, 75, 79, 82, 84, 85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>89, 91, 95, 98</td>
</tr>
<tr>
<td>blue</td>
<td>2</td>
<td>1, 2</td>
</tr>
</tbody>
</table>

continued in next column

Table 1. Clusters of top ranking residues in 1abvA.
2.4.2 Possible novel functional surfaces at 25% coverage. One group of residues is conserved on the 1abvA surface, away from (or substantially larger than) other functional sites and interfaces recognizable in PDB entry 1abv. It is shown in Fig. 4. The right panel shows (in blue) the rest of the larger cluster this surface belongs to.

Fig. 4. A possible active surface on the chain 1abvA. The larger cluster it belongs to is shown in blue.

The residues belonging to this surface “patch” are listed in Table 2, while Table 3 suggests possible disruptive replacements for these residues (see Section 3.6).

Table 2. Residues forming surface “patch” in 1abvA.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Y</td>
<td>Y(100)</td>
<td>0.02</td>
</tr>
<tr>
<td>13</td>
<td>A</td>
<td>A(100)</td>
<td>0.02</td>
</tr>
<tr>
<td>84</td>
<td>R</td>
<td>R(98)K(1)</td>
<td>0.03</td>
</tr>
<tr>
<td>27</td>
<td>W</td>
<td>W(98)LAF</td>
<td>0.04</td>
</tr>
<tr>
<td>9</td>
<td>P</td>
<td>P(91)TLI(6)ER</td>
<td>0.07</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>F(97)Y(2)</td>
<td>0.08</td>
</tr>
<tr>
<td>82</td>
<td>N</td>
<td>N(90)S(1)A(3)Q(2)R(1)</td>
<td>0.09</td>
</tr>
<tr>
<td>74</td>
<td>N</td>
<td>N(90)R(8)AG</td>
<td>0.10</td>
</tr>
<tr>
<td>5</td>
<td>T</td>
<td>T(86)N(1)V(2)D</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S(7).1</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>R</td>
<td>R(86)H(3)K(1)</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Q(3)N(4)S</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>A</td>
<td>A(62)L(16)S(1)I(10)V(9)</td>
<td>0.16</td>
</tr>
<tr>
<td>12</td>
<td>K</td>
<td>K(69)E(18)Q(1)N</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D(1)R(7)V(1)</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>L</td>
<td>L(50)L(48)IY</td>
<td>0.18</td>
</tr>
<tr>
<td>6</td>
<td>V</td>
<td>V(52)I(26)L(20)E</td>
<td>0.19</td>
</tr>
<tr>
<td>98</td>
<td>L</td>
<td>L(11)L(69)F(7)</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M(6)I(2).1WH</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Disruptive mutations for the surface patch in 1abvA.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Y</td>
<td>(K)(QM)(NEVLA)IPE(D)</td>
</tr>
<tr>
<td>13</td>
<td>A</td>
<td>(K)VERI(Q)(H) (N) (F)TMW</td>
</tr>
<tr>
<td>84</td>
<td>R</td>
<td>(T)YD(SVCG) (F)ELWI</td>
</tr>
<tr>
<td>27</td>
<td>W</td>
<td>(KE)TQDG (R) (S)NCG</td>
</tr>
<tr>
<td>9</td>
<td>P</td>
<td>(Y) (R) (H) (T)</td>
</tr>
<tr>
<td>15</td>
<td>F</td>
<td>(K)E (Q)D</td>
</tr>
<tr>
<td>82</td>
<td>N</td>
<td>(Y)FW (H) (T)</td>
</tr>
<tr>
<td>74</td>
<td>N</td>
<td>(Y)FWH (T) (DR)</td>
</tr>
<tr>
<td>5</td>
<td>T</td>
<td>(R) (K) (H) (FW)</td>
</tr>
<tr>
<td>8</td>
<td>R</td>
<td>(T)YD (E) (VCAG)</td>
</tr>
<tr>
<td>14</td>
<td>A</td>
<td>(R)Y (K) (E)</td>
</tr>
<tr>
<td>12</td>
<td>K</td>
<td>(Y)FW (T) (CG)</td>
</tr>
<tr>
<td>75</td>
<td>L</td>
<td>(R)TKY (E) (H)</td>
</tr>
<tr>
<td>6</td>
<td>V</td>
<td>(YR) (KH) (E) (Q)</td>
</tr>
<tr>
<td>98</td>
<td>L</td>
<td>(R)T (Y) (KE)</td>
</tr>
</tbody>
</table>

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 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.
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 5.3 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 [LPNQDEM1K], large [WFYHR], hydrophobic [LPVAMWF], polar [GTGY]; positively [KHRR], or negatively [DE] charged, aromatic [WFYH], long aliphatic chain [EKRQ], OH-group possession [SDETY], and NH2 group possession [NQRK]. The suggestions are listed according to how different they appear to be from the original amino acid, and they are grouped in round brackets if they appear equally disruptive. From left to right, each bracketed group of amino acid types resembles more strongly the original (i.e., 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 this column of the alignment
  2. their type

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 5.3 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 [LPNQDEM1K], large [WFYHR], hydrophobic [LPVAMWF], polar [GTGY]; positively [KHRR], or negatively [DE] charged, aromatic [WFYH], long aliphatic chain [EKRQ], OH-group possession [SDETY], and NH2 group possession [NQRK]. The suggestions are listed according to how different they appear to be from the original amino acid, and they are grouped in round brackets if they appear equally disruptive. From left to right, each bracketed group of amino acid types resembles more strongly the original (i.e., 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 this column of the alignment
  2. their type

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.)
4.3.3 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.


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/. (Note for Windows users: the attached package needs to be unzipped for Pymol to read the scripts and launch the viewer.)

4.4 Note about ET Viewer  
Dan Morgan from the Lichtarge lab has developed a visualization tool specifically for viewing trace results. If you are interested, please visit: http://mammoth bcm.tmc.edu/traceview/ The viewer is self-unpacking and self-installing. Input files to be used with ETV (extension .etvx) can be found in the attachment to the main report.

4.5 Citing this work  


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

4.7 Attachments  
The following files should accompany this report:

- labvA.complex.pdb - coordinates of labvA with all of its interacting partners
- labvA.etvx - ET viewer input file for labvA
- labvA.cluster_report.summary - Cluster report summary for labvA
- labvA.ranks - Ranks file in sequence order for labvA
- labvA.clusters - Cluster descriptions for labvA
- labvA.msf - the multiple sequence alignment used for the chain labvA
- labvA.descr - description of sequences used in labvA msf
- labvA.ranks_sorted - full listing of residues and their ranking for labvA