1am2
Evolutionary trace report by report_maker
January 19, 2010

CONTENTS

1 Introduction 1

2 Chain 1am2A
2.1 P72065 overview 1
2.2 Multiple sequence alignment for 1am2A 1
2.3 Residue ranking in 1am2A 1
2.4 Top ranking residues in 1am2A and their position on
the structure 1
2.4.1 Clustering of residues at 45% coverage 2

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

4 Appendix
4.1 File formats 4
4.2 Color schemes used 4
4.3 Credits
4.3.1 Alistat 4
4.3.2 CE 4
4.3.3 DSSP 4
4.3.4 H SSP 4
4.3.5 LaTeX 4
4.3.6 Muscle 4
4.3.7 Pymol 4
4.4 Note about ET Viewer 4
4.5 Citing this work 4
4.6 About report_maker 4
4.7 Attachments 4

1 INTRODUCTION
From the original Protein Data Bank entry (PDB id 1am2):
Title: Gyra intein from mycobacterium xenopi
Compound: Mol id: 1; molecule: mxe gyrA intein; chain: a;
enengineered; yes; mutation; yes
Organism, scientific name: Mycobacterium Xenopi;
1am2 contains a single unique chain 1am2A (181 residues long).

2 CHAIN 1AM2A
2.1 P72065 overview
From SwissProt, id P72065, 89% identical to 1am2A:
Description: DNA gyrase subunit A (EC 5.99.1.3) [Contains: Mxe
gyrA intein] (Fragment).
Organism, scientific name: Mycobacterium xenopi.
Taxonomy: Bacteria; Actinobacteria; Actinobacteridae; Actinomy-
cetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
Function: DNA gyrase negatively supercoils closed circular double-
stranded DNA in an ATP-dependent manner and also catalyzes the
interconversion of other topological isomers of double-stranded DNA
rings, including catenanes and knotted rings.
Catalytic activity: ATP-dependent breakage, passage and rejoining
of double-stranded DNA.
Subunit: Made up of two chains. The A chain is responsible for
DNA breakage and rejoining; the B chain catalyzes ATP hydrolysis.
The enzyme forms an A2B2 tetramer.
PTM: This protein undergoes a protein self splicing that involves a
post-translational excision of the intervening region (intein) followed
by peptide ligation.
Similarity: Belongs to the topoisomerase gyrA/parC subunit 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 1am2A
For the chain 1am2A, the alignment 1am2A.msf (attached) with 11
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 1am2A.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>11</td>
</tr>
<tr>
<td>Total number of residues:</td>
<td>1935</td>
</tr>
<tr>
<td>Smallest:</td>
<td>173</td>
</tr>
<tr>
<td>Largest:</td>
<td>181</td>
</tr>
<tr>
<td>Average length:</td>
<td>175.9</td>
</tr>
<tr>
<td>Alignment length:</td>
<td>181</td>
</tr>
<tr>
<td>Average identity:</td>
<td>78%</td>
</tr>
<tr>
<td>Most related pair:</td>
<td>99%</td>
</tr>
<tr>
<td>Most unrelated pair:</td>
<td>68%</td>
</tr>
<tr>
<td>Most distant seq:</td>
<td>77%</td>
</tr>
</tbody>
</table>

Furthermore, 45% of residues show as conserved in this alignment.

The alignment consists of 63% 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 1am2A.descr.

2.3 Residue ranking in 1am2A

The 1am2A sequence is shown in Fig. 1, with each residue colored according to its estimated importance. The full listing of residues in 1am2A can be found in the file called 1am2A.ranks_sorted in the attachment.

2.4 Top ranking residues in 1am2A and their position on the structure

In the following we consider residues ranking among top 45% of residues in the protein (the closest this analysis allows us to get to 25%). Figure 2 shows residues in 1am2A 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 45% coverage. Fig. 3 shows the top 45% 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>color</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.

contined in next column

---

Fig. 1. Residues 0-198 in 1am2A colored by their relative importance. (See Appendix, Fig. 4, for the coloring scheme.)

Fig. 2. Residues in 1am2A, colored by their relative importance. Clockwise: front, back, top and bottom views.

Fig. 3. Residues in 1am2A, 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.
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.

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 5Å.

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 (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 [LPQDEM]K, large [WFYH], hydrophobic [LPVAMWFI], polar [GTCY]; positively [KHR], or negatively [DE] charged, aromatic [WFYH], long aliphatic chain [EKRQM], OH-group possession [SDETY], and NH2 group possession [NQRK]. The suggestions are listed according to how different they appear to be from the original amino acid, and they are grouped in round brackets if they appear equally disruptive. From left to right, each bracketed group of amino acid types resembles more strongly the original (i.e. is, presumably, less disruptive) These suggestions are tentative - they might prove disruptive to the fold rather than to the interaction. Many researcher will choose, however, the straightforward alanine mutations, especially in the beginning stages of their investigation.

4 APPENDIX

4.1 File formats
Files with extension “ranks_sort” 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

Table 1. Clusters of top ranking residues in 1am2A.

<table>
<thead>
<tr>
<th>cluster color</th>
<th>size</th>
<th>member residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>red</td>
<td>78</td>
<td>8,10,15,25,29,30,31,36,37,40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41,42,44,45,47,50,51,52,53</td>
</tr>
<tr>
<td></td>
<td></td>
<td>54,56,63,65,66,70,71,72,74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75,76,78,79,80,81,85,87,88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>89,90,91,92,96,97,101,106</td>
</tr>
<tr>
<td></td>
<td></td>
<td>107,133,134,135,136,139,156</td>
</tr>
<tr>
<td></td>
<td></td>
<td>157,158,159,160,161,162,163</td>
</tr>
<tr>
<td></td>
<td></td>
<td>164,166,169,173,175,176,177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>179,180,182,183,187,189,192</td>
</tr>
<tr>
<td>blue</td>
<td>2</td>
<td>193,194,196,197,198</td>
</tr>
</tbody>
</table>

Table 1. continued
4.3.3 DSSP (CE) of the optimal path

"Protein structure alignment by incremental combinatorial extension (CE) of the optimal path." Protein Engineering 11(9) 739-747.


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).


http://www.drive5.com/muscle/

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


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:

- 1am2A.complex.pdb - coordinates of 1am2A with all of its interacting partners
- 1am2A.etvx - ET viewer input file for 1am2A
- 1am2A.cluster_report.summary - Cluster report summary for 1am2A
- 1am2A.ranks - Ranks file in sequence order for 1am2A
- 1am2A.clusters - Cluster descriptions for 1am2A
- 1am2A.msf - the multiple sequence alignment used for the chain 1am2A
- 1am2A.descr - description of sequences used in 1am2A msf
- 1am2A.ranks_sorted - full listing of residues and their ranking for 1am2A