1ag4
Evolutionary trace report by report_maker
December 18, 2009

CONTENTS

1 Introduction 1

2 Chain 1ag4A 1

2.1 P09353 overview 1

2.2 Multiple sequence alignment for 1ag4A 1

2.3 Residue ranking in 1ag4A 1

2.4 Top ranking residues in 1ag4A and their position on the structure 1

2.4.1 Clustering of residues at 25% coverage. 2

2.4.2 Possible novel functional surfaces at 25% coverage. 2

3 Notes on using trace results 4

3.1 Coverage 4

3.2 Known substitutions 4

3.3 Surface 4

3.4 Number of contacts 4

3.5 Annotation 4

3.6 Mutation suggestions 4

4 Appendix 4

4.1 File formats 4

4.2 Color schemes used 4

4.3 Credits 4

4.3.1 Alistat 4

4.3.2 CE 4

4.3.3 DSSP 5

4.3.4 HSSP 5

4.3.5 LaTex 5

4.3.6 Muscle 5

4.3.7 Pymol 5

4.4 Note about ET Viewer 5

4.5 Citing this work 5

4.6 About report_maker 5

4.7 Attachments 5

1 INTRODUCTION

From the original Protein Data Bank entry (PDB id 1ag4):

Title: Nmr structure of spherulin 3a (s3a) from physarum polycephalum, minimized average structure

Compound: Mol id: 1; molecule: spherulin 3a; chain: a; synonym: s3a; engineered: yes; other details: reduced form of s3a, with calcium

Organism, scientific name: Physarum Polycephalum;

1ag4 contains a single unique chain 1ag4A (103 residues long).

This is an NMR-determined structure – in this report the first model in the file was used.

2 CHAIN 1AG4A

2.1 P09353 overview

From SwissProt, id P09353, 100% identical to 1ag4A:

Description: Spherulin 3A (S3A).

Organism, scientific name: Physarum polycephalum (Slime mold).

Function: Structural protein.

Subcellular location: Cytoplasmic.

Developmental stage: Major encystment-specific protein.

Similarity: Belongs to the beta/gamma-crystallin family.

Similarity: Contains 2 beta/gamma crystallin ‘Greek key’ domains.

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 1ag4A

For the chain 1ag4A, the alignment 1ag4A.msf (attached) with 15 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 1ag4A.msf. Its statistics, from the alistat program are the following:

Lichtarge lab 2006
Format: MSF  
Number of sequences: 15  
Total number of residues: 1390  
Smallest: 79  
Largest: 103  
Average length: 92.7  
Alignment length: 103  
Average identity: 34%  
Most related pair: 99%  
Most unrelated pair: 20%  
Most distant seq: 31%  
Furthermore, 5% of residues show as conserved in this alignment.

The alignment consists of 6% eukaryotic, and 6% 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 1ag4A.descr.

2.3 Residue ranking in 1ag4A

The 1ag4A sequence is shown in Fig. 1, with each residue colored according to its estimated importance. The full listing of residues in 1ag4A can be found in the file called 1ag4A.ranks in the attachment.

2.4 Top ranking residues in 1ag4A 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 1ag4A 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>17</td>
<td>16,19,23,24,26,28,31,45,46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>49,50,52,53,83,89,90,97</td>
</tr>
<tr>
<td>blue</td>
<td>6</td>
<td>61,62,64,69,93,94</td>
</tr>
</tbody>
</table>

Table 1. Clusters of top ranking residues in 1ag4A.

2.4.2 Possible novel functional surfaces at 25% coverage.

One group of residues is conserved on the 1ag4A surface, away from (or substantially larger than) other functional sites and interfaces recognizable in PDB entry 1ag4. It is shown in Fig. 4. 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. 4.** A possible active surface on the chain 1ag4A. The larger cluster it belongs to is shown in blue.

While Table 3 suggests possible disruptive replacements for these residues (see Section 3.6).

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>G</td>
<td>G(100)</td>
<td>0.06</td>
</tr>
<tr>
<td>23</td>
<td>N</td>
<td>N(80)S(13)D(6)</td>
<td>0.12</td>
</tr>
<tr>
<td>52</td>
<td>K</td>
<td>R(20)K(80)</td>
<td>0.14</td>
</tr>
<tr>
<td>19</td>
<td>Y</td>
<td>S(6)Y(80)F(13)</td>
<td>0.17</td>
</tr>
<tr>
<td>53</td>
<td>V</td>
<td>V(40)I(53)L(6)</td>
<td>0.19</td>
</tr>
<tr>
<td>28</td>
<td>S</td>
<td>T(6)S(73)C(6)</td>
<td>0.23</td>
</tr>
<tr>
<td>24</td>
<td>F</td>
<td>Y(46)F(46)H(6)</td>
<td>0.24</td>
</tr>
</tbody>
</table>

**Table 2.** Residues forming surface "patch" in 1ag4A.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td>G</td>
<td>(KER)(FQMWHD) (NYLPI) (SVA)</td>
</tr>
<tr>
<td>23</td>
<td>N</td>
<td>(Y) (FWH) (R) (T)</td>
</tr>
<tr>
<td>52</td>
<td>K</td>
<td>(Y) (T) (FW) (SVCAG)</td>
</tr>
<tr>
<td>19</td>
<td>Y</td>
<td>(K) (Q) (M) (ER)</td>
</tr>
<tr>
<td>53</td>
<td>V</td>
<td>(YR) (KE) (H) (QD)</td>
</tr>
<tr>
<td>28</td>
<td>S</td>
<td>(R) (K) (H) (FW)</td>
</tr>
<tr>
<td>24</td>
<td>F</td>
<td>(KE) (Q) (D) (T)</td>
</tr>
</tbody>
</table>

**Table 3.** Disruptive mutations for the surface patch in 1ag4A.

Another group of surface residues is shown in Fig. 5. The residues belonging to this surface "patch" are listed in Table 4, while Table 5 suggests possible disruptive replacements for these residues (see Section 3.6).

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>69</td>
<td>G</td>
<td>G(100)</td>
<td>0.06</td>
</tr>
<tr>
<td>93</td>
<td>S</td>
<td>S(93)T(6)</td>
<td>0.08</td>
</tr>
<tr>
<td>67</td>
<td>F</td>
<td>F(66)Y(33)</td>
<td>0.17</td>
</tr>
<tr>
<td>62</td>
<td>F</td>
<td>Y(46)F(46)C(6)</td>
<td>0.18</td>
</tr>
</tbody>
</table>

**Table 4. continued**

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>64</td>
<td>D</td>
<td>H(33)D(59)N(6)</td>
</tr>
</tbody>
</table>

**Table 4.** Residues forming surface "patch" in 1ag4A.

Another group of surface residues 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 3.6).

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>N</td>
<td>N(100)</td>
<td>0.06</td>
</tr>
<tr>
<td>46</td>
<td>D</td>
<td>D(100)</td>
<td>0.06</td>
</tr>
<tr>
<td>89</td>
<td>N</td>
<td>N(93)D(6)</td>
<td>0.08</td>
</tr>
<tr>
<td>90</td>
<td>D</td>
<td>D(66)N(33)</td>
<td>0.10</td>
</tr>
<tr>
<td>83</td>
<td>L</td>
<td>L(73)K(6)F(6)</td>
<td>0.22</td>
</tr>
</tbody>
</table>

**Table 5.** Disruptive mutations for the surface patch in 1ag4A.

**Table 6.**

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>substitutions(%)</th>
<th>cvg</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>N</td>
<td>N(100)</td>
<td>0.06</td>
</tr>
<tr>
<td>46</td>
<td>D</td>
<td>D(100)</td>
<td>0.06</td>
</tr>
<tr>
<td>89</td>
<td>N</td>
<td>N(93)D(6)</td>
<td>0.08</td>
</tr>
<tr>
<td>90</td>
<td>D</td>
<td>D(66)N(33)</td>
<td>0.10</td>
</tr>
<tr>
<td>83</td>
<td>L</td>
<td>L(73)K(6)F(6)</td>
<td>0.22</td>
</tr>
</tbody>
</table>

**Table 5. continued**

continued in next column

continued in next column
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 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 [LPNQDEM1K], large [WFYHR], hydrophobic [LPVAMWF], 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_sorted” are the actual trace results. The fields in the table in this file:
4.3.1 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.

4.3.2 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.

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 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).


4.4 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

The method used to rank residues and make predictions in this report can be found in Mihalek, I., I. Reš, O. Lichtarge. (2004). "A Family of


4.6 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.

4.7 Attachments

The following files should accompany this report:

- 1ag4A.complex.pdb - coordinates of 1ag4A with all of its interacting partners
- 1ag4A.etvx - ET viewer input file for 1ag4A
- 1ag4A.cluster_report.summary - Cluster report summary for 1ag4A
- 1ag4A.ranks - Ranks file in sequence order for 1ag4A
- 1ag4A.clusters - Cluster descriptions for 1ag4A
- 1ag4A.msf - the multiple sequence alignment used for the chain 1ag4A
- 1ag4A.descr - description of sequences used in 1ag4A msf
- 1ag4A.ranks_sorted - full listing of residues and their ranking for 1ag4A