1a64
Evolutionary trace report by report_maker
December 27, 2009

CONTENTS
1 Introduction 1
2 Chain 1a64A 1
  2.1 P08921 overview 1
  2.2 Multiple sequence alignment for 1a64A 1
  2.3 Residue ranking in 1a64A 1
  2.4 Top ranking residues in 1a64A and their position on the structure 2
  2.4.1 Clustering of residues at 26% coverage. 2
  2.4.2 Overlap with known functional surfaces at 26% coverage. 2
3 Notes on using trace results 3
  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
  4.1 File formats 4
  4.2 Color schemes used 4
  4.3 Credits 4
    4.3.1 Alistat 5
    4.3.2 CE 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 1a64):
Title: Engineering a misfolded form of rat cd2
Compound: Mol id: 1; molecule: cd2; chain: a, b; fragment: domain 1; engineered: yes; mutation: yes
Organism, scientific name: Rattus Norvegicus;
1a64 contains a single unique chain 1a64A (94 residues long) and its homologue 1a64B.

2 CHAIN 1A64A
2.1 P08921 overview
From SwissProt, id P08921, 97% identical to 1a64A:
1 Description: T-cell surface antigen CD2 precursor (T-cell surface antigen T11/Leu-5) (LFA-2) (LFA-3 receptor) (OX-34 antigen).
1 Organism, scientific name: Rattus norvegicus (Rat).
1 Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 Function: CD2 interacts with lymphocyte function-associated antigen (LFA-3) and CD48/BCM1 to mediate adhesion between T cells and other cell types. CD2 is implicated in the triggering of T- cells, the cytoplasmic domain is implicated in the signaling function.
1 Subunit: Interacts with CD2AP (By similarity).
1 Subcellular location: Type I membrane protein.
3 Similarity: Contains 1 Ig-like C2-type (immunoglobulin-like) domain.
3 Similarity: Contains 1 Ig-like V-type (immunoglobulin-like) domain.
3 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 2.2 Multiple sequence alignment for 1a64A
For the chain 1a64A, the alignment 1a64A.msf (attached) with 18 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 1a64A.msf. Its statistics, from the aalistat program are the following:

Format: MSF  
Number of sequences: 18  
Total number of residues: 1602  
Smallest: 74  
Largest: 94  
Average length: 89.0  
Alignment length: 94  
Average identity: 42%  
Most related pair: 99%  
Most unrelated pair: 11%  
Most distant seq: 31%

Furthermore, 3% of residues show as conserved in this alignment. The alignment consists of 77% eukaryotic (77% 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 1a64A.descr.

2.3 Residue ranking in 1a64A

The 1a64A sequence is shown in Fig. 1, with each residue colored according to its estimated importance. The full listing of residues in 1a64A can be found in the file called 1a64A.ranks_sorted in the attachment.

2.4 Top ranking residues in 1a64A and their position on the structure

In the following we consider residues ranking among top 26% of residues in the protein (the closest this analysis allows us to get to 25%). Figure 2 shows residues in 1a64A 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 26% coverage

Fig. 3 shows the top 26% 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>9</td>
<td>60,61,63,65,66,68,72,76,93</td>
</tr>
<tr>
<td>blue</td>
<td>7</td>
<td>16,17,18,19,32,38,40</td>
</tr>
</tbody>
</table>

Table 1. continued

<table>
<thead>
<tr>
<th>cluster color</th>
<th>size</th>
<th>member residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>yellow</td>
<td>2</td>
<td>78,89</td>
</tr>
<tr>
<td>green</td>
<td>2</td>
<td>7,8</td>
</tr>
</tbody>
</table>

Table 1. continued
Table 1. Clusters of top ranking residues in 1a64A.

2.4.2 Overlap with known functional surfaces at 26% coverage. The name of the ligand is composed of the source PDB identifier and the heteroatom name used in that file.

Interface with 1a64B. Table 2 lists the top 26% of residues at the interface with 1a64B. The following table (Table 3) suggests possible disruptive replacements for these residues (see Section 3.6).

Table 2.

<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>32</td>
<td>W</td>
<td>W(100)</td>
<td>0.03</td>
<td>119/9</td>
<td>2.91</td>
</tr>
<tr>
<td>63</td>
<td>L</td>
<td>L(100)</td>
<td>0.03</td>
<td>92/48</td>
<td>2.72</td>
</tr>
<tr>
<td>76</td>
<td>Y</td>
<td>Y(100)</td>
<td>0.03</td>
<td>63/14</td>
<td>3.30</td>
</tr>
<tr>
<td>72</td>
<td>D</td>
<td>D(94)</td>
<td>0.05</td>
<td>3/3</td>
<td>4.55</td>
</tr>
<tr>
<td>8</td>
<td>G</td>
<td>G(77)</td>
<td>0.06</td>
<td>39/39</td>
<td>2.65</td>
</tr>
<tr>
<td>78</td>
<td>V</td>
<td>V(77)</td>
<td>0.07</td>
<td>55/18</td>
<td>3.10</td>
</tr>
<tr>
<td>65</td>
<td>I</td>
<td>I(94)</td>
<td>0.09</td>
<td>34/23</td>
<td>2.91</td>
</tr>
<tr>
<td>19</td>
<td>P</td>
<td>P(88)</td>
<td>0.10</td>
<td>12/3</td>
<td>4.13</td>
</tr>
<tr>
<td>18</td>
<td>I</td>
<td>I(83)</td>
<td>0.11</td>
<td>39/9</td>
<td>2.91</td>
</tr>
<tr>
<td>7</td>
<td>W</td>
<td>W(72)</td>
<td>0.12</td>
<td>64/12</td>
<td>3.06</td>
</tr>
<tr>
<td>13</td>
<td>G</td>
<td>G(38)</td>
<td>0.14</td>
<td>26/26</td>
<td>3.10</td>
</tr>
<tr>
<td>38</td>
<td>L</td>
<td>L(38)</td>
<td>0.14</td>
<td>29/13</td>
<td>3.26</td>
</tr>
<tr>
<td>10</td>
<td>L</td>
<td>L(72)</td>
<td>0.16</td>
<td>51/14</td>
<td>3.30</td>
</tr>
<tr>
<td>99</td>
<td>E</td>
<td>E(72)</td>
<td>0.16</td>
<td>6/1</td>
<td>4.21</td>
</tr>
<tr>
<td>40</td>
<td>A</td>
<td>A(77)</td>
<td>0.17</td>
<td>25/17</td>
<td>3.21</td>
</tr>
<tr>
<td>68</td>
<td>L</td>
<td>L(88)</td>
<td>0.18</td>
<td>82/35</td>
<td>2.68</td>
</tr>
<tr>
<td>89</td>
<td>L</td>
<td>L(77)</td>
<td>0.19</td>
<td>9/0</td>
<td>3.85</td>
</tr>
</tbody>
</table>

continued in next column

Table 2. 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>61</td>
<td>G</td>
<td>G(94)</td>
<td>0.20</td>
<td>24/24</td>
<td>2.83</td>
</tr>
<tr>
<td>93</td>
<td>L</td>
<td>L(33)</td>
<td>0.21</td>
<td>29/5</td>
<td>3.38</td>
</tr>
<tr>
<td>17</td>
<td>N</td>
<td>N(16)</td>
<td>0.22</td>
<td>28/18</td>
<td>2.88</td>
</tr>
<tr>
<td>66</td>
<td>K</td>
<td>K(83)</td>
<td>0.23</td>
<td>11/11</td>
<td>3.23</td>
</tr>
<tr>
<td>60</td>
<td>N</td>
<td>N(77)</td>
<td>0.24</td>
<td>20/19</td>
<td>3.81</td>
</tr>
</tbody>
</table>

Table 3.

<table>
<thead>
<tr>
<th>res</th>
<th>type</th>
<th>disruptive mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>W</td>
<td>(KE) (TQD) (SNCRG) (M)</td>
</tr>
<tr>
<td>63</td>
<td>L</td>
<td>(YR) (TH) (SKECG) (FQWD)</td>
</tr>
<tr>
<td>76</td>
<td>Y</td>
<td>(K) (QM) (NEVLAPIR) (D)</td>
</tr>
<tr>
<td>72</td>
<td>D</td>
<td>(R) (FKCAWG) (TYQM) (SNLPI)</td>
</tr>
<tr>
<td>8</td>
<td>G</td>
<td>(KER) (FQMWD) (NLPI) (Y)</td>
</tr>
<tr>
<td>78</td>
<td>V</td>
<td>(R) (Y) (KE) (H)</td>
</tr>
<tr>
<td>65</td>
<td>I</td>
<td>(YR) (TH) (KE) (SCG)</td>
</tr>
<tr>
<td>19</td>
<td>P</td>
<td>(R) (Y) (T) (H)</td>
</tr>
<tr>
<td>18</td>
<td>I</td>
<td>(YR) (TH) (SKECG) (FQWD)</td>
</tr>
<tr>
<td>7</td>
<td>W</td>
<td>(KE) (T) (QD) (CG)</td>
</tr>
<tr>
<td>13</td>
<td>G</td>
<td>(R) (KF) (FWH) (EQM)</td>
</tr>
<tr>
<td>38</td>
<td>L</td>
<td>(Y) (T) (HR) (SCG)</td>
</tr>
<tr>
<td>10</td>
<td>L</td>
<td>(Y) (R) (H) (T)</td>
</tr>
<tr>
<td>99</td>
<td>E</td>
<td>(FW) (H) (VCAG) (R)</td>
</tr>
<tr>
<td>40</td>
<td>A</td>
<td>(Y) (HR) (KE) (FWD)</td>
</tr>
<tr>
<td>68</td>
<td>L</td>
<td>(Y) (R) (H) (T)</td>
</tr>
<tr>
<td>89</td>
<td>L</td>
<td>(R) (Y) (H) (K)</td>
</tr>
<tr>
<td>61</td>
<td>G</td>
<td>(KER) (QHD) (FYMW) (N)</td>
</tr>
<tr>
<td>93</td>
<td>L</td>
<td>(R) (Y) (T) (H)</td>
</tr>
<tr>
<td>17</td>
<td>N</td>
<td>(Y) (H) (FW) (TR)</td>
</tr>
<tr>
<td>66</td>
<td>K</td>
<td>(Y) (T) (FW) (CG)</td>
</tr>
<tr>
<td>60</td>
<td>N</td>
<td>(Y) (FWH) (R) (TE)</td>
</tr>
</tbody>
</table>

Table 2. The top 26% of residues in 1a64A at the interface with 1a64B. (Field names: res: residue number in the PDB entry; type: amino acid type; subst's: substitutions seen in the alignment; with the percentage 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 3. List of disruptive mutations for the top 26% of residues in 1a64A, that are at the interface with 1a64B.

Figure 4 shows residues in 1a64A colored by their importance, at the interface with 1a64B.
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 neighboring 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 [LPNQDEMIK], large [WFYHR], hydrophobic [LPVAMWFI], polar [GTCY]; positively [KHKR], or negatively [DE] charged, aromatic [WFYH], long aliphatic chain [EKRQM], 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.

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

---

**Fig. 4.** Residues in 1a64A, at the interface with 1a64B, colored by their relative importance. 1a64B is shown in backbone representation (See Appendix for the coloring scheme for the protein chain 1a64A.)
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)
   • \( \text{covg} \) coverage - percentage of the residues on the structure which have this rho or smaller
   • \( \text{gaps} \) percentage of gaps in this column

4.2 Color schemes used

The following color scheme is used in figures with residues colored by the estimated evolutionary pressure they experience can be seen in Fig. 5.

4.3 Credits

4.3.1 Alistat
   - \textit{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 \( \text{idents} / \text{MIN(len1, len2)} \) where \( \text{idents} \) is the number of exact identities and \( \text{len1, len2} \) 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). \textit{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/ 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 \( 10\AA^2 \), 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.

4.3.4 HSSP

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

4.3.6 \textit{Muscle}

4.3.7 \textit{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 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:

- 1a64A.complex.pdb - coordinates of 1a64A with all of its interacting partners
- 1a64A.etvx - ET viewer input file for 1a64A
- 1a64A.cluster_report.summary - Cluster report summary for 1a64A
- 1a64A.ranks - Ranks file in sequence order for 1a64A
- 1a64A.clusters - Cluster descriptions for 1a64A
- 1a64A.msf - the multiple sequence alignment used for the chain 1a64A
- 1a64A.descr - description of sequences used in 1a64A msf
- 1a64A.ranks_sorted - full listing of residues and their ranking for 1a64A
- 1a64A.1a64B.if.pml - Pymol script for Figure 4
- 1a64A.cbcvg - used by other 1a64A – related pymol scripts