
Cohen-Gonsaud, Martin; Keep, Nicholas H.; Davies, Angharad P.; Ward, John; Henderson, Brian and Labesse, Gilles (2004). Resuscitation-promoting factors possess a lysozyme-like domain. *Trends in Biochemical Sciences* 29 (1) 7-10.

Resuscitation promoting factors possess a lysozyme-like domain

The novel bacterial cytokine family, termed resuscitation promoting factors (Rpfs), share a conserved domain of uncharacterised function. Predicting the structure of this domain suggests that Rpfs possess a lysozyme-like domain. The model highlights the good conservation of residues involved in catalysis and substrate binding. A lysozyme-like function makes sense for this domain in the light of experimental characterisation of the biological function of Rpfs.

A protein capable of resuscitating dormant *Micrococcus luteus* cultures – resuscitation promoting factor (Rpf) has been defined as the first bacterial 'cytokine' [1,2]. A number of high G+C bacteria contain homologues of this gene with the *M. tuberculosis* genome containing five *rpf* genes [3]. Recently this Rpf domain has been found in a mycobacteriophage [4]. All these proteins share a common sequence region – the Rpf domain [1,2] of unknown function.

Proteins containing the Rpf domain are able to reactivate dormant bacteria at very low concentration [1,3]. The *M.tuberculosis* and *M.luteus* proteins have been shown to be secreted [3,5]. The Rpf domain region contains some well-conserved residues. However, no significant similarities have been previously described with any other proteins of known biochemical activity and the precise biochemical function and mechanism of action of the Rpf domain remain unclear.

Most Rpf proteins contain potential peptide signal and/or transmembrane regions indicating that they are likely to be membrane associated or extracellular in agreement with the experimental evidence [3,5]. Furthermore, the micrococcal and two of the *Streptomyces coelicolor* Rpfs posses a 40 residues long LysM domain which is reported to bind to peptidoglycan [6,7]. This indicates that the Rpf domain is likely to be involved with the cell wall and the cell membrane.

Various Rpf domain and full-length sequences were used as queries to search the SWISSPROT [8] and the non-redundant GenPept databases [9] using PSI-BLAST [10]. This revealed significant similarities among known Rpfs and some uncharacterised proteins. After convergence, a consensus sequence was derived showing several short stretches of strictly conserved residues. No protein of known structure was detected, although some weak similarities were observed well above the E-value cut-off threshold of 0.002 with c-type lysozymes. PSI-BLAST was continued to convergence or 8 iterations.

Fold compatibility of the RPF domain region of mycobacterial Rpfs with known 3D structures was analysed using 3D-PSSM [11], FUGUE [12], GenTHREADER [13], PDB-BLAST (http://bioinformatics.burnham-inst.org/pdb_blast/), SAM-T99 [14] through the dedicated meta-server @TOME (http://bioserv.cbs.cnrs.fr); [15]). With most queries (including those in Fig. 1), mGenTHREADER ranked first various c-type and g-type lysozymes [23-26] with e-values ranging from 0.2 to 1.0, while 3D-PSSM suggested weak fold compatibility with both cytochrome C and lysozymes. However, the consensus motif for haem attachment ('CxxCH') present in cytochrome C was absent in Rpf. For two of the mycobacterial Rpf sequences (Rv2450c and Rv2389c), SAM-T99 and FUGUE showed significant threading scores (e-value of 0.001443 and Z-score of 5.29, respectively) for c-type lysozymes. Using the same software, the other mycobacterial Rpf sequences showed weaker compatibility with lysozymes. However, eukaryotic c-type and g-type lysozymes (PDB3LZT and PDB1LMQ or PDB153L) showed relatively high sequence conservation (24-28% over a common core of roughly 110 residues) with any of the Rpf sequences, whatever the threader used and despite some overall low threading scores. Use of the full length sequences gives similar conclusions.

In order to further validate the sequence-structure alignment, various models were built and assessed by molecular structure and amino acid environment evaluation tools. For each top-ranking templates (up to 14 per query), TITO [16], SCWRL [17] and MODELLER [18] were run. Assessments of these crude models were performed using Verify3D [19] and PROSA [20]. Good models (scoring above 0.2 and up to 0.4 with Verify3D with pseudo-energy below -0.3 and down to –0.9 as computed by PROSA) were obtained with the most Rpfs. While c-type

lysozymes provided us with satisfactory models, other templates rarely produced acceptable models. After manual alignment refinement, the scores computed with the programs Verify3D (above 0.4; [19]), PROSA (below -0.9; [20]) and ERRAT [21] (above 82 %) validated the structural alignment with PDB1LMQ or PDB3LZT and the resulting three-dimensional models built for several members of this protein superfamily. The hydrophobic core observed in the current models corresponds with the positions of the better-conserved residues in the Rpfs (see Fig 1b). This structural analysis revealed sensible structures for the insertions/deletions, as well as a compact and hydrophobic core at the interface of the α-helices and the β-sheets running from helix 2 to helix 3 in the c-type lysozymes. No conservation is found in helix 1, but sufficient residues are present in all sequences for this helix to exist. Furthermore, the strict conservation of the residues lying in the catalytic site, including the structurally required glycine, the catalytic proton donor in the glycoside hydrolysis (the glutamate E61 in Rv2389c) [22] and the saccharide-binding groove especially three tryptophans (W54, W85 and W118 in Rv2389c) suggested a common function is shared by the template and the modelled proteins. Although none of the cysteines that form the disulphide bonds in c–type lysozymes are conserved, within the Rpf domain family, there is a fully conserved pair of cysteines, which are placed fairly close in most models such that forming a disulphide bridge is possible with only a small rearrangement of the model. What is not conserved is the second acidic residue (aspartate), which forms the covalent intermediate in the lysozymes that operate with retention of configuration [22]. This indicates that if the Rpf domains have catalytic activity, the reaction might proceed with inversion of configuration as do a number of already characterised lysozymes such as T4 or goose lysozymes (PDB153L) [22-24], which lack a second acidic residue.

The significance of the strictly conserved residue matches was checked using PATTINPROT and a deduced consensus sequence (see Fig 1). An E-value of 1.5×10^{-8} was obtained for a pattern that selected 81 c-type lysozymes, 41 alpha-lactalbumin and 16 Rpf and only 11 false positive sequences from SwissProt +SPtTrEMBL. These results suggested that the common domain of Rpfs would adopt the c type lysozyme fold in spite of the undetected sequence similarity by PSI-BLAST. This failure, despite an overall good sequence identity (25-30% over ~110 aa; Fig. 1a), is probably due to the domain length being short, as well as different number of amino acids between the conserved motifs.

Our structure prediction strongly indicates that the Rpf domain shares the c-type lysozyme fold. This fold is also shared by lactalbumin, which does not naturally cleave saccharides, but can be converted to a lysozyme by a limited number of mutations [27]. The Rpf domain does not have the tyrosine that blocks the sugar binding site in lactalbumin and does have the active site glutamate that is lacking in lactalbumin, but it does not have all the lysozymes-like residues introduced. The homology between Rpf and lysozyme is too weak to be certain that Rpf has lysozyme like activity or even binds saccharides, but these are testable hypothesis that others and we will investigate. However given that Rpf domains may release a small molecule by hydrolysis, it is intriguing that a molecule of less than 1400 Daltons isolated from a culture supernatant of *M.tuberculosis* also increases the viability of cultures [28].

Martin Cohen-Gonsaud, Nicholas H. Keep*, School of Crystallography and BBRSC Centre for Structural Biology, Birkbeck College, University of London, Malet Street, London WC1E 7HX – United Kingdom.
Angharad P. Davies, Department of Medical Microbiology, Royal Free and University College Medical School, Rowland Hill Street, London NW3 – United Kingdom.
John Ward, Department of Biochemistry and Molecular Biology, University College London, Gower St, London WC1E
Brian Henderson, Division of Infection and Immunity, Eastman Dental Institute, University College London, 256 Gray's Inn Road, London WC1X 8LD – United Kingdom.
Gilles Labesse, Centre de Biochimie Structurale, INSERM U554 - CNRS UMR5048 - Université Montpellier I,

15, Av. Charles Flahault, 34060 MONTPELLIER Cedex – France
* To whom correspondence should be addressed n.keep@mail.cryst.bbk.ac.uk Tel 020-7631-6852 Fax 020-7641-6803.
This work was supported by EU Grant QLK2-2001-02018 (MC-G and NHK). The PSI-BLAST and structural prediction meta server outputs for the mycobacterial Rpf domain and full length sequences are available at http://www.infobiosud.cnrs.fr/bioserver/RPF/suppl.html.

References

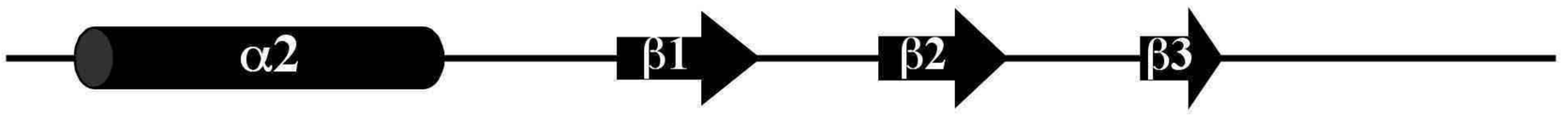
1 Mukamolova, G.V *et al.* (1998). A bacterial cytokine *Proc. Natl. Acad. Sci (USA)* 95, 8916-8921
2 Kell, D.B. and Young, M. (2000). Bacterial dormancy and culturability: the role of autocrine growth factors. *Curr. Op. Micro.* 3, 238-243.
3 Mukamolova, G.V *et al.* (2002). A family of autocrine growth factors in Mycobacterium tuberculosis *Mol Microbiol* 46, 623-635.
4 Pedulla ML *et al.* (2003) Origins of highly mosaic mycobacteriophage genomes. *Cell* 113,171-182.
5 Mukamolova, G.V *et al.* (2002). The rpf gene of Micrococcus luteus encodes an essential secreted growth factor. *Mol Microbiol* 46, 611-622.
6 Birkeland, N.-K. (1994). Cloning and molecular characterization and expression of the genes encoding the lytic function of lactococcal bacteriophage φLC3 : a dual lysis system of modular design. *Can. J. Microbiol.* 40, 658-665.
7 Bateman, A and Bycroft, M. (2000). The structure of a LysM domain from *E.coli* membrane-bound lytic murein transglycosylase D (MltD). *J.Mol. Biol.* 299, 1113-1119.
8 Bairoch A. and Apweiler R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000 *Nucleic Acids Res.* 28, 45-48
9 Benson D.A. *et al.* (2002) GenBank. *Nucleic Acids Res.* 30, 17-20
10 Altschul, S.F. *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389-3402
11 Kelley, L.A. *et al.* (2000). Enhanced Genome Annotation using Structural Profiles in the Program 3D-PSSM. *J. Mol. Biol.* 299, 499-520
12 Shi, J. *et al.* (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.* 310, 243-257
13 McGuffin, L.J. *et al.* (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404-405
14 Karplus, K. *et al.* (1998) Hidden Markov models for detecting remote protein homologies. *Bioinformatics* 14, 846-856
15 Douguet, D. and Labesse G. (2001) Easier threading through web-based comparisons and cross-validations. *BioInformatics* 17, 752-753
16 Labesse, G. and Mornon, J.-P. (1998). A Tool for Incremental Threading Optimization (TITO) to help alignment and modelling of remote homologs. *BioInformatics* 14, 206-211
17 Dunbrack, R.L. Jr, and Karplus, M. (1993) Backbone-dependent rotamer library for proteins. Application to side-chain prediction. *J Mol Biol*. 230, 543-74
18 Sali, A. and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815
19 Eisenberg, D. *et al.* (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277, 396-404
20 Sippl, M.J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins* 17, 355-362.
21 Colovos, C. and Yeates, T.O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 2, 1511-1519
22 Vocadlo D. *et al*. (2001) Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. *Nature* 412, 835-838.
23 Kuroki R. *et al*. (2001) Structural basis of the conversion of T4 lysozyme into a

transglycosidase by reengineering the active site. *Proc. Natl. Acad. Sci. USA* 96, 8949-8954.
24 Weaver *et al.* (1995) The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the "goose-type" lysozymes lack a catalytic aspartate residue. *J Mol. Biol.* 24, 554-568
25 Vollan *et al.* (1999) Structural studies on the binding of 4-methylumbelliferone glycosides of chitin to rainbow trout lysozyme. *Acta Crystallogr* D 55, 60-6
26 Walsh *et al.* (1998) Refinement of triclinic hen egg-white lysozyme at atomic resolution. *Acta Crystallogr* D 54, 522-46
27 Xue *et al.* (2000). α-Lactalbumin mutant acting as Lysozyme *Proteins*: *Structure, Function and Genetics* 42, 17-22.
28 Sun and Zhang (1999). Spent culture supernatant of Mycobacterium tuberculosis H37Ra improves viability of aged cultures of this strain and allows small inocula to initiate growth. *J.Bac.* 181, 7626-7628.
29 Combet, C. *et al.* (2000) NPS@: network protein sequence analysis. *Trends Biochem. Sci.* 25, 147-150

Fig. 1 a) Multiple sequence alignment of Rpf domain region and C-type lysozymes. The alignment was refined manually from pairwise alignment obtained from our meta-server [15]. Structure codes are from the PDB (http://www.rcsb.org/pdb/) and sequence codes are from the *Mycobacterum tuberculosis* genome for Rv2389c, Rv2450c, Rv1884c, Rv1009 and Rv0867c. Sequence from NCBI Entrez database (http://www.ncbi.nlm.nih.gov/entrez) for the Rpfs of *Micrococcus luteus* (NP_613076) named M_LUTE_RPF, *Streptomyces avermitilis* (BAC74961) named S_AVER_RPF, *Streptomyces coelicolor* (NP_733519) named S_COEL_RPF and the gp33 protein from the Mycobacteriophage Barnyard (NP_818571). Secondary structure assignment for the crystal structure, PDB1LMQ [25] and PDB3LZT [26] was performed using P-SEA (http://bioserv.cbs.cnrs.fr/HTML_BIO/frame_sea.html). Residues in conserved position (mainly with hydrophobic properties) are highlighted in orange, the catalytic conserved glutamate in red. The sequence pattern specifically derived using PATTINPROT [29] for both c-type lysozymes and Rpfs is as follows: [GQEKNRFY]-[GACVILFYWHND]-[STNE]-[ASDNHQEKR]-[ANGST]-X(11,15)-[YHL]-G-X-[VILMFY]-[EQR]-[VILMFY]-X(4)-W. The relevant columns in the alignment are indicated with an X.
b) Schematic representation of the c-type lysozyme structure 1LMQ. The side-chains of the strictly conserved residues from the alignment are coloured in orange (mainly in contact with the substrate), and red for the catalytic glutamate. The saccharide substrate is coloured in grey. Regions where an insertion or deletion is found between the 1LMQ and the Rpf sequences are coloured in blue, showing that these all lie in loops.

a)

```
                                  .    .  XXXXX  .    .    .    .  XX·XXX .  X  .    .
1LMQ          23 NSLPNWVCLSKWESSYNTQATNRNTDGSTDYGIFQINSRYWCDDGRTP 70
3LZT          23 YSLGNWVCAAKFESNFNTQATNRNTDGSTDYGILQINSRWWCNDGRTP 70
Rv2389c       49 ADDIDWDAIAQCESGGNWAANTGN----GLYGGLQISQATWD------ 86
Rv2450c       96 AYSVNWDAIAQCESGGNWSINTGN----GYYGGLRFTAGTWR------ 133
Rv1884c       80 GPSPNWDAVAQCESGGNWAANTGN----GKYGGLQFKPATWA------ 117
Rv1009       280 IDGSIWDAIAGCEAGGNWAINTGN----GYYGGVQFDQGTWE------ 317
Rv0867c       39 ATDGEWDQVARCESGGNWSINTGN----GYLGGLQFTQSTWA------ 76
M_LUTE_RPF    47 ATVDTWDRLAECESNGTWDINTGN----GFYGGVQFTLSSWQ------ 84
S_COEL_RPF    42 ADAATWDKVAACESTDDWDINTGN----GYYGGLQFTQSTWE------ 79
S_AVER_RPF    47 ADADTWNKVAACESSDNWSINTGN----GYYGGLQFTQSTWE------ 84
gp33        1511 KAGADWDAIAQKESGGNWAINTGN----GYYGGLQFAQSSWE------ 1548
```




```
                                  .    .    .    .    .    .    .    .    .
1LMQ          71 GAKNVCGIRC---SQLLTDDLTVAIRCAKRVVLDPNGIGAWVAWRLHC 115
3LZT          71 GSRNLCNIPC---SALLSSDITASVNCAKKIVSDGNGMNAWVAWRNRC 115
Rv2389c       87 ---SNGGVGS---PAAAS--PQQQIEVADNIMKT-QGPGAWPKCSSCS 125
Rv2450c      134 ---ANGGSGS---AANAS--REEQIRVAENVLRS-QGIRAWPVCGRRG 172
Rv1884c      118 ---AFGGVGN---PAAAS--REQQIAVANRVLAE-QGLDAWPTCGAAS 156
Rv1009       318 ---ANGGLRYAPRADLAT--REEQIAVAEVTRLR-QGWGAWPVCAARA 359
Rv0867c       77 ---AHGGGEFAPSAQLAS--REQQIAVGERVLAT-QGRGAWPVCGRGL 118
M_LUTE_RPF    85 ---AVGGEGY---PHQAS--KAEQIKRAEILQDL-QGWGAWPLCSQKL 123
S_COEL_RPF    80 ---AFGGTRYAPRADLAT--REQQIAGAEKVLDT-QGPGAWPVCSERA 118
S_AVER_RPF    85 ---AYGGKVYAQRADLAT--RDQQIAVAEKVLEG-QGPGAWPVCSVRA 123
gp33        1549 ---AAGGLAYASRADLAS--KEQQIAAAEELLKQ-QGPGAWPNTFVAA 1587
```

b)