

# The structure of a resuscitation-promoting factor domain from *Mycobacterium tuberculosis* shows homology to lysozymes

Martin Cohen-Gonsaud<sup>1</sup>, Philippe Barthe<sup>2</sup>, Claire Bagn ris<sup>1</sup>, Brian Henderson<sup>3</sup>, John Ward<sup>4</sup>, Christian Roumestand<sup>2</sup> & Nicholas H Keep<sup>1</sup>

**Resuscitation-promoting factor (RPF) proteins reactivate stationary-phase cultures of (G+C)-rich Gram-positive bacteria including the causative agent of tuberculosis, *Mycobacterium tuberculosis*. We report the solution structure of the RPF domain from *M. tuberculosis* Rv1009 (RpfB) solved by heteronuclear multidimensional NMR. Structural homology with various glycoside hydrolases suggested that RpfB cleaved oligosaccharides. Biochemical studies indicate that a conserved active site glutamate is important for resuscitation activity. These data, as well as the presence of a clear binding pocket for a large molecule, indicate that oligosaccharide cleavage is probably the signal for revival from dormancy.**

One of the keys to *M. tuberculosis*' success as a pathogen is its ability to persist in its host organism in a latent state for years after the first phase of infection<sup>1</sup>. The World Health Organization estimates that one-third of the world's population harbor a latent tuberculosis infection. The reactivation of the bacteria and, as a consequence, the development of clinical tuberculosis occur in 5–10% of non-HIV-infected people. The reactivation rate is much higher in HIV patients (up to 30%), with tuberculosis accounting for 13% of HIV-related deaths (World Health Organization Fact Sheet No.104; <http://www.who.int/mediacentre/factsheets/fs104/en/>). Actively dividing *M. tuberculosis* infection can be cleared by a combination of drug therapy and the patient's immune response, but once the dormant state has been entered the bacteria are no longer cleared. Understanding and controlling the entry and exit from dormancy is therefore important in the development of new anti-tubercular therapies.

A secreted protein from *Micrococcus luteus* that decreases the growth lag time and increases the number of colony-forming units when added to dormant cultures of this organism has been purified and its gene has been sequenced. This 17-kDa protein, Rpf, is active at very low concentrations (picomolar)<sup>2</sup>. Homologs of this RPF are found in several (G+C)-rich Gram-positive bacteria including *M. tuberculosis*, which has five genes (*rpfA–E*) that contain a region homologous to the *M. luteus* Rpf. The growth-stimulating activity of the five tuberculosis RPFs has been demonstrated on dormant cultures of *Mycobacterium bovis* BCG<sup>3</sup>. It has been proposed that these proteins are bacterial 'cytokines' that signal the bacteria to exit dormancy by binding to a receptor. By determining the molecular structure of an RPF, we aimed to gain insight into the mechanism of action of these important molecules.

## RESULTS

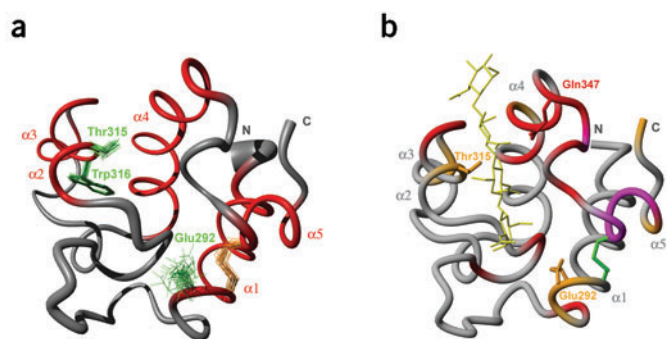
### NMR structure of an active RPF domain

Recently, we published a proposal, based on bioinformatic analysis, that the common core domain of all RPF proteins is structurally related to the c-type lysozyme family (CAZy family: GH22) and the active site glutamate is conserved<sup>4</sup>. We have subcloned, overexpressed and purified the C-terminal 108-residue core domain of Rv1009 (RpfB), called RpfBc. This domain corresponds to the Rpf homologous region found in all RPFs (residues 280–362) and an additional 25 N-terminal residues<sup>5</sup>. In our analysis the conserved sequence region was structurally related to the residues between helices  $\alpha 2$  and  $\alpha 5$  of the c-type lysozymes but enough residues to form an equivalent to helix  $\alpha 1$  were included.

First, the ability of the RpfBc region to resuscitate dormant culture of *M. luteus* was tested, showing that this protein was active at picomolar concentrations (data not shown). The RPF regions from Rv1884c (RpfC) and Rv2389c (RpfD) were also tested and showed a similarly high activity. These experiments formally demonstrate that the conserved domain of the RPF is responsible for the resuscitation of the bacteria, and is fully active as an isolated domain, in agreement with experiments on the conserved region of the *M. luteus* Rpf<sup>6</sup>.

Resonance assignments of RpfBc were carried out with a series of standard two- and three-dimensional experiments using <sup>15</sup>N- and <sup>15</sup>N, <sup>13</sup>C-labeled protein<sup>5</sup>. The first 22 residues from the N-terminal segment (Asn255–Val276) exhibited negative ratio values in the <sup>1</sup>H-<sup>15</sup>N heteronuclear NOE experiment corresponding to highly mobile residues<sup>5</sup>. This demonstrates that there is no equivalent in RpfBc of helix  $\alpha 1$  of the c-type lysozyme. For the remaining residues, structures were

<sup>1</sup>School of Crystallography and Institute for Structural Molecular Biology, Birkbeck College, University of London, Malet Street, London, WC1E 7HX, UK. <sup>2</sup>Centre de Biochimie Structurale, CNRS UMR5048, INSERM UMR554, UMI, Facult  de Pharmacie, BP 14491, 15 Avenue Charles Flahault, 34093 Montpellier Cedex 5, France. <sup>3</sup>Division of Microbial Diseases, Eastman Dental Institute, University College London, 256 Gray's Inn Road, London WC1X 8LD, UK. <sup>4</sup>Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK. Correspondence should be addressed to N.H.K. (n.keep@mail.cryst.bbk.ac.uk).



**Figure 1** Solution structure and oligosaccharide binding to RpfBc. (a) Sausage representation of 30 NMR structures of RpfBc (residues Thr274–Arg362). The backbone thickness is directly proportional to the r.m.s. deviation of the ensemble. The sausage is colored by secondary structure, with helices in red, coil and turn in gray and disulfide bond in orange. (b) Chemical shift mapping of RpfBc after addition of tri-NAG. Ribbon representation of the RpfBc structure with tetra-NAG superimposed based on a structural superposition with c-type lysozyme–tetra-NAG complex (PDB entry 1LMQ). Residues whose combined weighted N and NH chemical shift change on addition of tri-NAG exceeds 1.0 p.p.m. are red (for example, Glu347) and those whose chemical shift are between 0.85 and 1.0 p.p.m. are orange (for example, Glu315 and Thr317). Residues that disappear from the HSQC owing to chemical exchange line broadening are purple. This figure was generated with MolMol<sup>20</sup>.

calculated to be compatible with 1,613 distance restraints. The ensemble of the 30 best structures shows an r.m.s. deviation for the main chain of 0.57 Å (Fig. 1a). Experimental details, results and statistics for the structure determination are described in Table 1.

### RpfBc resembles several glycoside hydrolases

The experimentally determined structure of RpfBc is most similar to c-type lysozymes with the top hit in FSSP<sup>7</sup>, PDB entry 3LZT, having a Z-score of 6.4 (r.m.s. deviation of 2.9 Å over 79 residues). The oligosaccharide-binding site cleft of c-type lysozymes is well defined (PDB entry 1LZS<sup>8</sup>). An equivalent cleft is present in RpfBc and we show below that this does indeed bind polysaccharide.

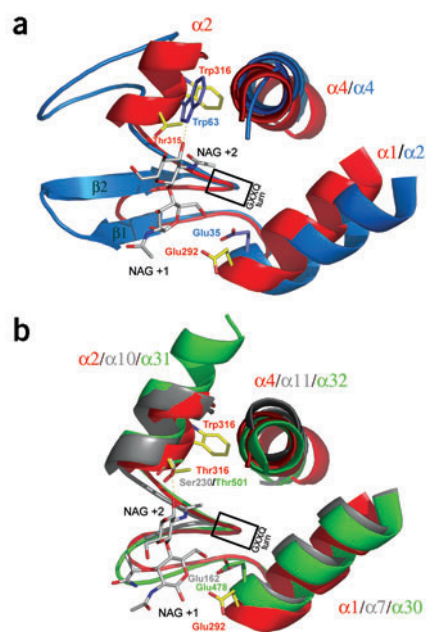
The peptidoglycan-binding cleft can be thought of as composed of two sides (Fig. 1b). The first side is formed predominantly by the two largest helices of the c-type lysozyme fold (helix  $\alpha 2$  and  $\alpha 4$ ). The RpfBc helices  $\alpha 1$  and  $\alpha 4$  match these two helices well. Consequently, the catalytic glutamate at the C-terminal end of the first matching helix is at the same position in those two proteins (Fig. 2a). The  $\alpha 5$  helices in the two structures are also equivalent but differ slightly in relative orientation to the larger helices, probably owing to the sole disulfide bridge in RpfBc (conserved in all RPFs) being at a different point between helices  $\alpha 1$  and  $\alpha 5$  compared with the disulfide between the equivalent helices  $\alpha 2$  and  $\alpha 5$  in the c-type lysozymes. In RpfBc, there is no equivalent of helices  $\alpha 1$  and  $\alpha 6$  in the c-type lysozyme fold; these helices are dispensable from the fold without altering the function of the protein as they lie away from the cleft. Consequently, the RPF fold is the most compact lysozyme-like structure determined.

The second side of the substrate cleft differs between the RPFs and c-type lysozymes. The equivalent region to the section forming the two  $\beta$ -strands in c-type lysozyme is four residues shorter in the RPF fold, so although the backbone near the cleft superimposes well, the strands are not present (Fig. 2a). The two structures then diverge the most, with helix  $\alpha 2$  in RpfBc not matching the c-type lysozyme fold (Fig. 2a). However,

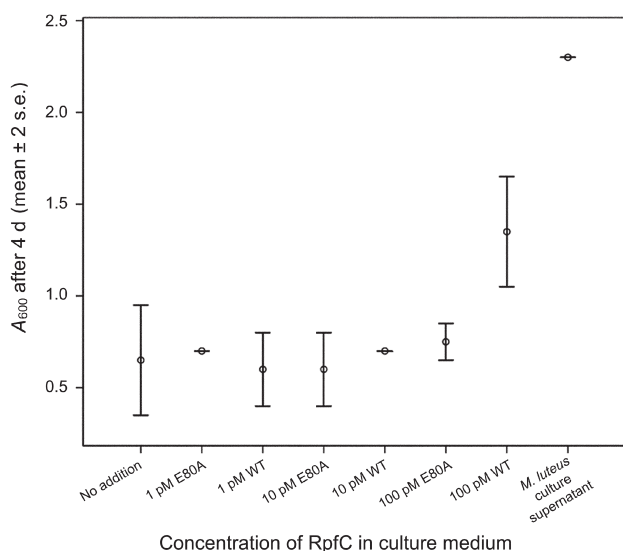
this side of the cleft closely resembles the bacterial soluble lytic transglycosidase (SLT) proteins<sup>9,10</sup> (CAZy family: GH23), particularly Slt70 (PDB entry 1QSA; Fig. 2b), as Slt35 (PDB entry 1D0K) has a 45-residue helical insertion between the helices equivalent to  $\alpha 1$  and  $\alpha 2$  of RpfBc. After the GXXQ turn (conserved among RPFs,  $\alpha$ -lactalbumins, SLTs, and c-type and g-type lysozymes) RpfBc and the SLTs have an additional helix ( $\alpha 2$  in RpfBc; Fig. 2b). At this position in the c-type lysozymes, the O4-hydroxyl of the N-acetylglucosamine (NAG) in position +2 forms a hydrogen bond with the indole part of a tryptophan (such as Trp316 in PDB entry 1LZS; Fig. 2a). In Slt35 and Slt70, this role is played, respectively, by a serine and a threonine conserved among the various SLTs. The superposition of RpfBc with the SLTs shows the presence of a threonine conserved among the RPFs at this position (Thr315; Fig. 2b).

### Oligosaccharide binds to the RPF domain

Using <sup>1</sup>H-<sup>15</sup>N-HSQC NMR experiments we tested the binding of RpfBc to NAG and its trimer, N,N',N''-triacetylchitotriose (tri-NAG). NAG is one of the two sugars that make up the polysaccharide of peptidoglycan and polymers of NAG have been bound to lysozyme in several crystal structures. With monomeric NAG no chemical shift changes were observed, whereas in the trimer case, chemical shift changes were observed for residues of the putative binding groove (Fig. 1b and Supplementary Fig. 1 online). All the residues we propose above to be



**Figure 2** Comparison of RpfBc with other glycoside hydrolases. (a) Structure superimposition in the cleft area of the RpfBc (red) and a c-type lysozyme (PDB entry 1LMQ; blue). The catalytic glutamate at the C-terminal end of the first helix is at the same position in RpfBc and lysozyme (Glu292 and Glu35, respectively). (b) Structure superimposition in the cleft area of the RpfBc (red), Slt35 (PDB entry 1D0K; gray) and Slt70 (PDB entry 1QSA; green). The catalytic glutamate is at the same position for RpfBc, Slt35 and Slt70 (Glu292, Glu162 and Glu478, respectively). At the equivalent position of Thr315 in RpfBc, a serine is present in Slt35 (Ser230) and a threonine in Slt70 (Thr501). These residues bind the O4-hydroxyl of the NAG in position +2 and play the same role as one of the conserved tryptophans, Trp316, in PDB entry 1LMQ. This figure was generated with the PyMOL (<http://pymol.sourceforge.net>).



**Figure 3** Resuscitation activity assays of recombinant *M. tuberculosis* RpfC and RpfC E80A. Only 100 pM unmutated RpfC and the endogenous *M. luteus* Rpf control show enhanced resuscitation. Error bars represent twice the s.e. of duplicate samples.

involved in the binding are affected by the addition of Tri-NAG (for example, Glu292 and Thr315). The main perturbation is observed for the loop  $\alpha 4$ - $\alpha 5$ , with Gln347 being the most affected. The equivalent conserved glutamine in the lysozyme (Gln104) is a key residue in binding oligosaccharide and forms hydrogen bonds to the linking oxygen between the NAGs in positions +2 and +3 and also to the acetyl group of the NAG in position +3. This loop moves the most between apo and holo crystal structures of c-type lysozyme (Supplementary Fig. 2 online), suggesting that the same binding mechanism is present in the c-type lysozyme and the RPFs.

#### Active site glutamate is required for *in vivo* activity

To test the importance of the putative catalytic glutamate of the RPF, we produced a mutant protein, RpfC E80A, and compared its resuscitation activity with that of wild-type recombinant RpfC. Circular dichroism (CD) spectra show that both the wild-type and mutant proteins are folded very similarly (Supplementary Fig. 3 online). The resuscitation activity of the mutant protein was completely suppressed, highlighting the critical role of the conserved glutamate (Fig. 3). However, the RPFs we have expressed and described here do not degrade peptidoglycan to the point that the bacteria were lysed in assay conditions in which lysozyme completely lysed the cells, nor did they have substantial activity in a commercial fluorescence assay for lysozyme (EnzCheck lysozyme assay kit, Molecular Probes; data not shown).

#### DISCUSSION

The structure of the functional RPF domain is therefore a compact hybrid of the SLT and c-type lysozyme folds, both of which cleave peptidoglycan. Although the overall sequence conservation between the RPFs and either of these protein families is low, a few crucial features such as the catalytic glutamate are conserved. We have demonstrated that this residue is critical for the resuscitation activity. The need for the active site glutamate, combined with a clear binding pocket for a large molecule that undergoes rearrangement in presence of a polysaccharide, strongly suggests that the function of RPFs is to cleave oligosaccharide and this is responsible for their role in revival of dormant bacteria. However, the RPFs we studied do not degrade peptidoglycan in the same way as lysozyme to degrade the cell wall entirely. This does not preclude an oligosaccharide cleavage being the mechanism of action. A second enzyme might be required to release the product of the cleavage from peptidoglycan; a second factor has been

shown to act synergistically with RPF in *Corynebacterium glutamicum*<sup>11</sup>. Alternatively, the true substrate may be a modified form of peptidoglycan<sup>12</sup> found only in dormant cells or in a limited region related to cell septation. Our results do not rule out a model in which RPF may simply bind to an oligosaccharide in a lectin-like manner. However, the data suggest that RPF probably acts enzymatically on the cell wall of the bacteria. The identification of the physiological substrates for the RPFs coupled with structural and biochemical analysis will improve our understanding of their functional activity and will be a first step in the development of inhibitors to help understand and prevent the reactivation of pathogenic mycobacteria such as *M. tuberculosis* or *Mycobacterium leprae*.

#### METHODS

**NMR structure determination.** NMR spectra were recorded on samples of 0.5 mM, <sup>15</sup>N- or <sup>15</sup>N, <sup>13</sup>C-labeled RpfBc domain dissolved in 25 mM Na-acetate, pH 4.6, 2 mM  $\beta$ -mercaptoethanol with 5–10% D<sub>2</sub>O for the lock. Some spectra were also collected on samples prepared in D<sub>2</sub>O with the same buffer. NMR experiments were carried out at 20 °C on Bruker AVANCE 500 and 600 spectrometers equipped with 5-mm z-shielded gradient <sup>1</sup>H-<sup>13</sup>C-<sup>15</sup>N triple-resonance cryogenic probes. NMR data processing and analysis were carried out using GIFA<sup>13</sup> and NMRView<sup>14</sup>. Nucleus resonances (<sup>1</sup>H, <sup>15</sup>N and <sup>13</sup>C) were assigned using standard three-dimensional experiments<sup>5</sup>. The backbone dihedral angle restraints were

**Table 1** NMR and refinement statistics

	RpfBc
<b>NMR distance and dihedral constraints</b>	
Distance constraints	
Total NOE	1,613
Intra-residue	629
Inter-residue	
Sequential ( $ i - j  = 1$ )	432
Medium-range ( $ i - j  < 4$ )	276
Long-range ( $ i - j  > 5$ )	276
Intermolecular	–
Hydrogen bonds	10
Total dihedral angle restraints	
$\phi$	62
$\psi$	62
<b>Structure statistics</b>	
Violations (mean and s.d.)	
Distance constraints (Å)	0.11 ± 0.09
Dihedral angle constraints (°)	1.21 ± 0.17
Max. dihedral angle violation (°)	3.75 ± 0.44
Max. distance constraint violation (Å)	0.36 ± 0.09
Deviations from idealized geometry	
Bond lengths (Å)	0.01 ± 0.00
Bond angles (°)	0.68 ± 0.02
Impropers (°)	1.78 ± 0.10
Average pairwise r.m.s. deviation (Å) <sup>a</sup>	
Heavy	1.07 ± 0.14
Backbone	0.57 ± 0.11

<sup>a</sup>Pairwise r.m.s. deviation was calculated among 30 refined structures for residues 25–105.

obtained from chemical shift index analysis using TALOS<sup>15</sup>. The two cysteine residues have upfield-shifted <sup>13</sup>Cβ resonances (38.33 and 38.16 p.p.m., respectively) consistent with a disulfide bond formation<sup>16</sup>, which was enforced for structure calculations. Hydrogen bond restraints were incorporated if they were protected from H<sub>2</sub>O-D<sub>2</sub>O exchange as revealed from an HSQC spectrum recorded on a freshly lyophilized sample dissolved in a D<sub>2</sub>O buffer, and were supported by NOE analysis. NMR-derived restraints are shown in **Table 1**. Structure calculations were done with ARIA 1.2 (ref. 17) implemented in CNS 1.1 (ref. 18).

The 30 structures with the lowest energy were further refined by molecular dynamics calculations in explicit solvent to remove artifacts and analyzed with PROCHECK-NMR<sup>19</sup>. Analysis of the Ramachandran plot showed that all modeled residues were in allowed regions with 90.2% in the most favored region. The r.m.s. deviations and visual display were carried out with MolMol<sup>20</sup>.

**Chemical shift mapping.** RpfBc <sup>15</sup>N and <sup>1</sup>H<sub>N</sub> backbone chemical shift perturbations (Δδ) were measured from HSQC experiments upon titration with Tri-NAG (Sigma). In these experiments, the concentration of <sup>15</sup>N-RpfBc was kept constant (50 μM) and the concentration of unlabeled Tri-NAG was varied from 0.5 to 20 mM. An additional reference spectrum was taken on an RpfBc sample (50 μM) without Tri-NAG. All <sup>1</sup>H-<sup>15</sup>N HSQC spectra were recorded using a time domain data size of 64 *t*<sub>1</sub> × 512 *t*<sub>2</sub> complex points, and 32 transients per *t*<sub>1</sub> increment. For analysis, <sup>15</sup>N and <sup>1</sup>H<sub>N</sub> chemical shift changes were combined using the equation

$$\Delta\delta = [(\Delta\delta_{\text{H}})^2 + (\Delta\delta_{\text{N}} \times (\gamma_{\text{N}} / \gamma_{\text{H}}))^2]^{0.5}$$

where values of Δδ > 0.85 p.p.m. have been defined as significant perturbations.

**Assay of RPF activity.** *Micrococcus luteus* NCIMB 13267 was passaged on Nutrient Agar (Oxoid) plates. Overnight cultures of *M. luteus* in lactate minimal media (LMM) were grown to an A<sub>600</sub> of 1.8–2.2. The cells were removed by centrifugation and the supernatant was filter-sterilized, stored at 4 °C and used within 2 h of preparation as a source of endogenous *M. luteus* Rpf. For assay of RPF activity, an overnight culture of *M. luteus* in Nutrient Broth-2 (Oxoid) was diluted 1:10<sup>3</sup> in LMM and added to 30 ml universal containers in 4-ml aliquots. Suitable dilutions of recombinant RPF proteins (40 μl) were added to the universals and bacteria were cultured overnight on a shaking incubator at 37 °C. The growth of the cultures was monitored by the A<sub>600</sub> and compared with a negative control of *M. luteus* on its own and a positive control of *M. luteus* with sterile culture supernatant containing the endogenous *M. luteus* Rpf. The data presented in **Figure 3** are the growth after 4 d (average of duplicate cultures) with error bars of twice the s.e. The results were confirmed on a second separate growth assay.

**CD spectra.** CD spectra of unmutated and E80A recombinant RpfC were recorded on an AVIV 202 spectrometer at 0.25 mg ml<sup>-1</sup> in 20 mM sodium phosphate, pH 7.0. Protein concentrations were determined using calculated extinction coefficients. All data are normalized to mean residue ellipticity (MRE) using CDTool<sup>21</sup>.

**Coordinates.** The chemical shifts of the RpfBc domain and the atomic coordinates have been deposited, respectively, in the BioMagResBank (accession code BMRB-6221) and in the Protein Data Bank (accession code 1XSF).

*Note: Supplementary information is available on the Nature Structural & Molecular Biology website.*

## ACKNOWLEDGMENTS

This work was supported by European Union grant QLK2-2001-02018 (M.C.G. and N.H.K.). NMR experiments were recorded and analyzed using the facilities of the structural biology platform RIO (Centre de Biochimie Structurale, Montpellier, France). We thank B. Sarra for help with CD.

## COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 3 December 2004; accepted 28 January 2005

Published online at <http://www.nature.com/nsmb/>

1. Tufariello, J.M., Chan, J. & Flynn, J.L. Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection. *Lancet Infect. Dis.* **3**, 578–590 (2003).
2. Mukamolova, G.V., Kaprelyants, A.S., Young, D.I., Young, M. & Kell, D.B. A bacterial cytokine. *Proc. Natl. Acad. Sci. USA* **95**, 8916–8921 (1998).
3. Mukamolova, G.V. *et al.* A family of autocrine growth factors in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **46**, 623–635 (2002).
4. Cohen-Gonsaud, M. *et al.* Resuscitation-promoting factors possess a lysozyme-like domain. *Trends Biochem. Sci.* **29**, 7–10 (2004).
5. Cohen-Gonsaud, M. *et al.* <sup>1</sup>H, <sup>15</sup>N and <sup>13</sup>C chemical shift assignments of the Resuscitation Promoting Factor domain of Rv1009 from *Mycobacterium tuberculosis*. *J. Biomol. NMR* **30**, 373–374 (2004).
6. Mukamolova, G.V. *et al.* The *rpf* gene of *Micrococcus luteus* encodes an essential secreted growth factor. *Mol. Microbiol.* **46**, 611–621 (2002).
7. Holm, L. & Sander, C. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **26**, 316–319 (1998).
8. Song, H., Inaka, K., Maenaka, K. & Matsushima, M. Structural changes of active site cleft and different saccharide binding modes in human lysozyme co-crystallized with hexa-*N*-acetyl-chitohexaose at pH 4.0. *J. Mol. Biol.* **244**, 522–540 (1994).
9. van Asselt, E.J. *et al.* Crystal structure of *Escherichia coli* lytic transglycosylase Slt35 reveals a lysozyme-like catalytic domain with an EF-hand. *Structure Fold. Des.* **7**, 1167–1180 (1999).
10. van Asselt, E.J., Thunnissen, A.M. & Dijkstra, B.W. High resolution crystal structures of the *Escherichia coli* lytic transglycosylase Slt70 and its complex with a peptidoglycan fragment. *J. Mol. Biol.* **291**, 877–898 (1999).
11. Hartmann, M. *et al.* The glycosylated cell surface protein Rpf2, containing a resuscitation-promoting factor motif, is involved in intercellular communication of *Corynebacterium glutamicum*. *Arch. Microbiol.* **182**, 299–312 (2004).
12. Raymond J.B., Mahapatra, S., Crick, D.C. & Pavelka, M.S. Jr. Identification of the *namH* gene, encoding the hydroxylase responsible for the *N*-glycolylation of the mycobacterial peptidoglycan. *J. Biol. Chem.* **280**, 326–333 (2005).
13. Pons, J.L., Malliavin, T.E. & Delsuc, M.A. Gifa V 4: a complete package for NMR data set processing. *J. Biomol. NMR* **8**, 445–452 (1996).
14. Johnson, B.A. & R.A., B. NMRView: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603–614 (1994).
15. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
16. Sharma, D. & Rajarathnam, K. <sup>13</sup>C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR* **18**, 165–171 (2000).
17. Linge, J.P., O'Donoghue, S.I. & Nilges, M. Automated assignment of ambiguous nuclear overhauser effects with ARIA. *Methods Enzymol.* **339**, 71–90 (2001).
18. Brunger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination *Acta Crystallogr. D* **54**, 905–921 (1998).
19. Laskowski, R.A., MacArthur, M.W., Hutchinson, E.G. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
20. Koradi, R., Billeter, M. & Wuthrich, K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55 (1996).
21. Lees, J.G., Smith, B., Wien, F., Miles, A., and Wallace, B.A. CDTool: an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal. Biochem.* **332**, 285–289 (2004).