LA-UR-17-24508 (Accepted Manuscript)

Structural and biophysical characterization of the Mycobacterium tuberculosis protein Rv0577, a protein associated with neutral red staining of virulent tuberculosis strains and homolog of the Streptomyces coelicolor protein KbpA.

Buchko, Garry
Terwilliger, Thomas Charles
Kim, Chang Yub
Kim, Heungbok
Echols, Nathanial
Flynn, E. Megan
Ng, Ho-Leung
Stephenson, Sam
Myler, Peter
Alber, Tom

To be published in: Biochemistry

DOI to publisher's version: 10.1021/acs.biochem.7b00511

Permalink to record: http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-17-24508

Structural and biophysical characterization of Rv0577, a protein associated with neutral red staining of virulent strains of *Mycobacterium tuberculosis* and homolog of the *Streptomyces coelicolor* protein KbpA

Journal:	*Biochemistry*
Manuscript ID	Draft
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Buchko, Garry; Pacific Northwest National Laboratory, Earth and Biological Sciences Directorate Echols, Nathanial; Lawrence Berkeley National Laboratory, Physical Biosciences Division Flynn, E.; University of California - Berkeley, Department of Molecular and Cell Biology Ng, Ho; University of California - Berkeley, Department of Molecular and Cell Biology Stephenson, Sam; University of California - Berkeley, Department of Molecular and Cell Biology Kim, Heung-Bok; Los Alamos National Laboratory, Bioscience Division Myler, Peter; Center for Infectious Disease Research; University of Washington School of Medicine, Departments of Global Health and Biomedical Informatics and Medical Education Terwilliger, Thomas; Los Alamos National Laboratory, Life Sciences Division Alber, Tom; University of California - Berkeley, Department of Molecular and Cell Biology Kim, Chang-Yub; Los Alamos National Laboratory, Bioscience Division

Structural and biophysical characterization of the *Mycobacterium tuberculosis* protein Rv0577, a protein associated with neutral red staining of virulent tuberculosis strains and homolog of the *Streptomyces coelicolor* protein KbpA.

Garry W. Buchko[a,b,*], Nathanial Echols[c], E. Megan Flynn[d,±], Ho-Leung Ng[d,⊥], Sam Stephenson[d], Heung-Bok Kim[e], Peter J. Myler[a,f,g], Thomas C. Terwilliger[e], Tom Alber[d], and Chang-Yub Kim[e]

[a]Seattle Structural Genomics Center for Infectious Diseases
[b]Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA.
[c]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94158-2330, USA.
[d]Department of Molecular and Cell Biology, University of California-Berkeley, Berkeley, CA 94158-2330, USA.
[e]Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.
[f]Department of Medical Education and Biomedical Informatics & Department of Global Health, University of Washington, Seattle, WA 98195, USA.
[g]Center for Infectious Disease Research, Seattle, WA 98109-5219, USA.

[±] Current Address: Department of Early Discovery Biochemistry, Genentech Inc., South San Francisco, CA 94080, USA.
[⊥]Current Address: Department of Chemistry, University of Hawaii - Manoa, Honolulu, HI 96822-2275 USA.

Running Title: Crystal structure for *Mycobacterium tuberculosis* protein Rv0577

[*]Correspondence e-mail: garry.buchko@pnnl.gov

ABSTRACT: The 261-residue *Mycobacterium tuberculosis* protein Rv0577 is a prominent antigen in tuberculosis patients, the responsible component for neutral red staining of virulent strains of *M. tuberculosis*, a putative component in a methylglyoxal detoxification pathway, and an agonist of toll-like receptor 2. It also has 36% amino acid sequence identity to *Streptomyces coelicolor* AfsK-binding protein A (KbpA), a component in the complex secondary metabolite pathways in the *Streptomycetes* genus from which many commercial antibiotics are derived. To gain insight into the biological function of Rv0577 and the family of KpbA kinase regulators, the crystal structure for Rv0577 was determined to a resolution of 1.75 Å (3OXH), binding properties with neutral red and deoxyadenosine (Ado) surveyed, backbone dynamics measured, and thermal stability assayed by CD spectroscopy. The protein is composed of four approximate repeats with an βαββββ topology arranged radially in consecutive pairs to form two continuous eight-strand β-sheets capped on both ends with an α-helix. The two β-sheets intersect in the center at roughly a right angle and form an asymmetric deep "saddle" on both sides of the protein, saddle one (P11 to A129) and saddle two (L143 to A258), that may serve to bind ligands. NMR chemical shift perturbation experiments show that neutral red binds to Rv0577, further associating Rv0577 with the neutral red staining of virulent strains of *M. tuberculosis*. Similar experiments show that adenosine also binds to Rv0577, although less tightly, with estimated dissociation constants of 4.1 ± 0.3 mM for saddle one and > 1 M for saddle two. Heteronuclear steady-state $\{^1H\}$-^{15}N NOE, T_1, and T_2 values were generally uniform through-out the sequence with only a few modest pockets of differences suggestive of slightly different motion in loops between β-strands in saddle 1. Circular dichroism spectroscopy characterization of the thermal

stability of Rv0577 indicated irreversible unfolding upon heating with an estimated melting temperature of 56 °C. While it is not known if Rv0577 has a kinase regulatory role similar to its *Streptomyces* homolog KbpA, protein kinase and phosphatase signaling help *M. tuberculosis* adapt to the hostile host environment during infections. Consequently, new anti-tuberculosis drugs targeting Rv0577 may act by interfering with multiple mechanisms; a potential signaling machinery as well as toll-like receptor 2 activation and the methylglyoxal detoxification pathway.

Key Words: tuberculosis; Ser/Thr protein kinases; virulence factors; host-pathogen interactions; toll-like receptor 2 activation.

INTRODUCTION

Approximately two thirds of all natural, commercial antibiotics are secondary metabolites produced by soil-dwelling bacteria from the genus *Streptomycetes*.[1] The secondary metabolic pathways for these antibiotics are complex as are the signal transductions pathways that turn them on. Perhaps the best understood system in *Streptomycetes* is the blue colored polyketide actinorhodin produced by *Streptomyces coelicolor*.[2] As many as 100 genes for actinorhodin biosynthesis are activated by the pathway-specific *act*II-ORF4 transcriptional activators. The *act*II-ORF4 activators are themselves activated by a 63-residue protein, AfsS, after it forms a complex with RNA polymerase and is converted into a transcriptionally active open state with the energy available from ATP hydrolysis by AfsR, a transcriptional activator with ATPase activity that binds to the promotor of *AfsS* in the phosphorylated form.[3] As shown schematically in Figure 1, cytoplasmic AfsR is phosphorylated by AfsK,[4, 5] a serine/threonine kinase loosely bound to the membrane after AfsK auto-phosphorylates itself, likely upon sensing some external or internal stimuli. The genome of *S. coelicolor* contains 36 serine/threonine or tyrosine kinases and at least two of them, PkaG and AfsL, phosphorylate AfsR suggesting that AfsR is an integrator of multiple signals sensed by these kinases. Modulating the kinase activity of AfsK is KbpA (AfsK-binding protein A), a conserved hypothetical protein co-expressed with AfsR that binds to unphosphorylated AfsK and prevent auto-phosphorylation.[6] Six KbpA paralogs are encoded in the *S. coelicolor* genome, suggesting that KbpA and its paralogs regulate the 36 *S. coelicolor* kinases. The

mechanism of kinase recognition and regulation by KbpA or its six paralogs remains to be determined.

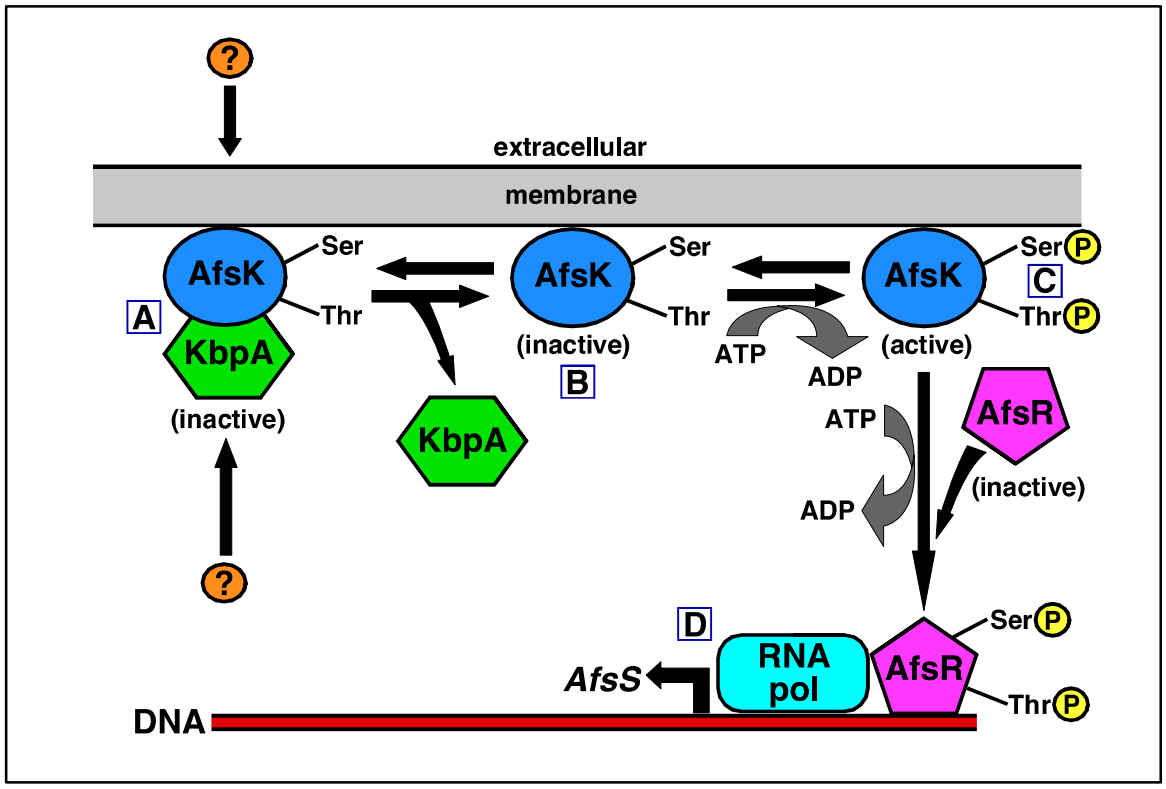


Figure 1. A simplified hypothetical model highlighting the role of KpbA in the regulation of actinorhodin production in *Streptomyces coelicolor*.[6] The AfsK-binding protein A, KbpA, binds to AfsK, a protein loosely associated with the membrane (**A**). An unknown internal or external effector, labeled with a question mark, modulates the release of KbpA from AfsK **(B)** to allow activation of KpbA by autophosphorylation (**C**). Activated AfsK, as well as other kinases, phosphorylates AfsR which then binds to the *AfsS* promotor to activate AfsS transcription (**D**). In turn the 63-residue protein, AfsS, is an *act*II-ORF4 promotor that turns on the transcription of the ~100 genes involved in actinorhodin biosynthesis. AfsK is expressed during all stages of cell growth[4] while KpbA is actively expressed after actinorhodin production has begun.

The *Streptomycetes* genus belongs to the same taxonomical order containing *Mycobacterium tuberculosis*, the etiological agent responsible for tuberculosis (TB), a chronic infectious disease that killed approximately 1.8 million people and infected another 10.4 million in 2015.[7] The *M. tuberculosis* genome encodes eleven Ser/Thr kinases[8] and two KbpA homologs, Rv0577 and Rv0911, whose amino acid sequence are 36% and 33% identical to KbpA, respectively. While the physiological roles of Rv0577 and Rv9011 are unknown, Rv0577 is present in *M. tuberculosis* culture filtrates,{Rosenkrands, 2000 #249} is a prominent antigen in TB patients,[9, 10] and is the responsible component for neutral red staining of virulent strains of *M. tuberculosis.*[11] Rv0577 has also been reported to be an agonist of toll-like receptor 2 (TLR-2), inducing dendrite cell maturation and driving a Th1 immune response.[12] Overall, such correlations suggest that Rv0577 may play a role in virulence, and therefore, represent a potential target to base new strategies to treat and control tuberculosis.[13, 14] To assist our understanding of the biological function of Rv0577 in *M. tuberculosis* infections and the apparently novel inhibitory mechanism of the *S. coelicolor* homolog KbpA, the crystal structure for Rv0577 was determined to 1.75 Å resolution. This structure, together with the NMR amide chemical shift assignments for this 261-residue protein (BMRB ID 17597)[14] allowed us to use chemical-shift perturbation experiments[15] to determine if Rv0577 binds neutral red and to map the potential ligand-binding surface.

MATERIALS AND METHODS

Cloning, Expression, and Purification. The *Rv0577* gene was amplified by PCR from the genomic DNA of *M. tuberculosis* strain H37Rv and employed to generate two clones using the pET28b expression vector (Novagen, Madison, WI). The difference between the two clones was that the expressed gene product for one contained an N-terminal, 20-residue extension (MGSSHHHHHHSSGLVPRGSH-; nRv0577) and the other a C-terminal eight residue extension (-RSHHHHHH; Rv0577c). Both recombinant plasmids were transformed into *Escherichia coli* BL21(DE3) cells (Novagen, Madison, WI) by a heat shock method. The protein containing the N-terminal extension was used for crystallography and was purified as described previously for the sensor domain of PknD[16] using three chromatography steps: immobilized metal ion affinity, HighTrap Q Sepharose ion exchange, and Superdex75 gel-filtration. Pooled fractions from the last chromatography step were dialyzed into the buffer used for crystallization (50 mM NaCl, 20 mM Tris, 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP), 1% glycerol (v/v), pH 7.5) and the protein then concentrated to 20 mg/mL. Selenomethionine-substituted (SeMet-) nRv0577 was expressed following a minimal media growth protocol that inhibited the methionine biosynthesis pathway[17, 18] and purified following the same protocol except for the addition of 1 mM TCEP in all buffers. Protein containing either the N- or C-terminal extension used for the solution studies was ^{15}N-labeled using a standard minimal media based auto-induction approach[19] and purified as previously described[14] using two chromatography steps: immobilized metal ion affinity and Superdex75 gel-filtration. The latter step exchanged Rv0577 into the NMR buffer used for the biophysical studies (100

mM NaCl, 20 mM Tris, 1.0 mM dithiothreitol, pH 7.0). For a few NMR experiments the N-terminal tag was removed from nRv0577 by concentrating the protein to ~ 20 mg/mL following the size exclusion chromatography step and diluting the protein 50% with thrombin digestion buffer (20 mM Tris, 100 mM NaCl, 2 mM $CaCl_2$, pH 8.0). N-terminal cleavage was complete after room temperature overnight incubation with ~5 mg of thrombin (MP Biomedicals, Solon, OH). Thrombin-cleaved nRv0577 was then purified and exchanged into NMR buffer by Superdex75 gel-filtration chromatography. Note that three non-native residues, GSH-, remained at the N-terminal of nRv0577 after thrombin cleavage.

Nuclear Magnetic Resonance (NMR) Spectroscopy. The NMR data were collected at 30 °C on ^{15}N-labeled samples using Varian spectrometers with ^{1}H resonance field strengths of approximately 600, 750, or 800 MHz. All spectrometers were equipped with a triple resonance probe and pulse field gradients with the highest field instrument containing a cyroprobe. Chemical shift perturbations studies with nRv0577 and Rv0577c (250 uL, 0.2 (Ado) and 0.3 (neutral red) mM) were conducted with adenosine (Ado; Research Product International) and the dye 3-amino-7-dimethylamino-2-methylphenazine hydrochloride (neutral red; Sigma Aldrich, St. Louis, MO) in a Shigemi NMR tube (Shigemi Inc., Allison Park, PA) using 50 mM stock solutions of each compound prepared in NMR buffer. An ^{1}H-^{15}N HSQC spectrum was collected after the addition of a two microliter aliquot of each compound at increasing compound:protein molar ratios up to 10 to one. An overall rotational correlation time (τ_c) for nRv0577 and Rv0577c was estimated at 20 °C from backbone amide ^{15}N $T_{1\rho}/T_1$ ratios measured using

a modified ^{1}H-^{15}N HSQC experiment to record an ^{15}N-edited one-dimensional spectrum[20, 21] Nitrogen-15 T_1 and T_2 values were measured (30 °C, 750 MHz ^{1}H resonance field strength) for Rv0577c using previously described experiments [22] with the following time delays (s): 0.0, 0.1, 0.2, 0.3. 0.4, 0.6, 0.8 (T_1); 0.01, 0.03, 0.05, 0.07, 0.09, 0.11, and 0.13 (T_2). Values for T_1 and T_2 and an estimate of their associated errors were obtained using the rate analysis function in Sparky (v3.115)[23] by fitting the measured peak heights to a two-parameter exponential function, $I(t) = I_o\exp(-t*T_{1,2})$. The average values for T_1 and T_2 were used to estimate τ_c at 30 °C using the following equation[20]:

$$\tau_c \cong (1/4\pi\nu_N)(6(T_1/T_2) - 7)^{-1/2} \quad \text{where } \nu_N = {}^{15}\text{N frequency in Hz} \qquad (1)$$

Steady-state $\{^{1}H\}$-^{15}N heteronuclear NOE values (NOE = I_{sat}/I_{unsat}) were measured (30 °C, 800 MHz ^{1}H resonance field strength) in triplicate by taking the ratio of ^{1}H-^{15}N HSQC cross peak volumes in spectra recorded in the presence (I_{sat}) and absence (I_{unsat}) of 3.0 s of proton presaturation prior to the ^{15}N excitation pulse.[22] All NMR data was processed using Felix2007 (MSI, San Diego, CA) software and analyzed with the program Sparky (v3.115) making use of previous amide chemical shift assignments (BMRB ID 17597).[14] Note that in the published ^{1}H-^{15}N HSQC spectrum[14] two assignments were mislabeled, F181 = R181 and A181 = A188.

Circular Dichroism (CD) Spectroscopy. An Aviv Model 410 spectropolarimeter (Lakewood, NJ) was used to collect circular dichroism data on a 400 μL 0.02 mM sample of Rv0577c (100 mM NaCl, 20 mM Tris, 1.0 mM dithiothreitol, pH

7.0) in a quartz cell of 0.1 cm path length. A thermal denaturation curve was obtained by recording the ellipticity at 220 nm in 2.0 °C intervals from 10 to 80 °C. The first derivative of the thermal denaturation curve[24] provides a quantitative estimation of the melting temperature, T_m, for Rv0577c. A steady-state wavelength spectra was recorded in 0.5 nm increments between 200 and 260 nm at 25 °C. The reported steady-state wavelength spectrum is the average of two consecutive scans, collected with a bandwidth of 1.0 nm and a time constant of 1.0 s, processed by subtracting a blank spectrum from the protein spectrum and automatic line smoothing using Aviv software.

Isothermal Titration Calometry (ITC). The interactions between Rv0577c and dAdo were measured on a MicroCal iTC_{200} calorimeter (MicroCal, Northampton, MA) using standard procedures. The same NMR buffer used in the Superdex75 gel-filtration chromatography step to purify Rv0577c was used to prepare Rv0577c dilutions and stock solutions of Ado. Two hundred twenty μL of 0.14 mM Rv0577c was placed in the mixing cell and titrated with 10 mM Ado. Each titration experiment was performed at 37 °C with 2.0 μL injections of 4 s duration in 180 s intervals with a mixing rate of 1000 rpm. The resulting isotherms were deconvoluted and fit to a single-site binding model by nonlinear least-squares regression using the Origin software package supplied with the instrument. A blank reference titration was recorded for Ado to account for the heats of dilution into the sample buffer. The reported disassociation constant (K_d) was determined from the average of two titrations.

Protein Crystallization, X-ray Diffraction, and Model Building. Conditions for the growth of diffraction quality crystals were identified from vapor-diffusion crystallization trials using the hanging drop method at 18 °C by mixing equal volumes of the protein (20 mg/mL) and the precipitant. For native nRv0557, crystals were observed with precipitant containing 1.3 M sodium citrate, 4% butanol (v/v), 50 mM MES, pH 6.5, with 1 mM p-chloromercuricbenzene sulfonic acid (PCMBS) added to solution prior to freezing. For the SeMet-labeled nRv0577, crystals were obtained with precipitant containing 0.9 M sodium citrate, 50 mM MES, 8% butanol, pH 6.5. All crystals were mounted on nylon cryo-loops, immersed briefly in a drop of the well precipitant adjusted to 0.75 M NaCl and 14% xylitol (v/v), flash-frozen in liquid nitrogen, stored under liquid nitrogen, and shipped to the Lawrence Berkeley National Laboratory Advanced Light Source for X-ray data collection.

All X-ray diffraction data was collected at beamline 8.3.1 with an ADSC Quantum 210 detector. Diffraction data were initially collected on a single PCMBD-derivatized native crystal to a resolution of 1.75 Å resolution, however, MAD/SAD phasing was unsuccessful, and because no structure had yet been determined from this family of proteins it was not possible to obtain a solution using molecular replacement. Consequently, a three-wavelength MAD data set was collected on a single SeMet-labeled crystal of nRv0577 to a resolution of 2.3 Å and used to solve the structure. The MAD data were integrated and scaled using HKL2000 software[25] and the program SOLVE[26] used to identify, refine, and obtain phases from seven selenium sites. Phases were improved by density modification with RESOLVE[27] to produce an interpretable electron density map at 2.3 Å resolution. The program SHARP[28] was then used to further refine

the selenium sites and phasing. Approximately 90% of the model was built automatically using ARP/wARP.[29] The native data were integrated and scaled using Elves software.[30] Due to slight non-isomorphism of the native and SeMet derivative crystals, molecular replacement with EPMR[31] was used to place the partial model obtained from the SAD data in the unit cell of the native data. The model was then refined against the native data set using REFMAC[32] cycled with manual building in O.[33] The stereochemical quality of the final model was evaluated using the programs MolProbity[34] and PROCHECK[35] and any conflicts addressed. MolProbity analysis showed that the overall geometry of the final deposited model ranked in the 93rd percentile (MolProbity score = 1.50) with an "all-atoms" clash score of 5.73. PROCHECK analysis indicated that 97.1% of the Φ/Ψ pairs were in most favored region and the remainder in additionally allowed regions. The data collection and structure refinement statistic are listed in Table 1. The native diffraction data and coordinates for the final model have been deposited to the Protein Data Bank (PDB entry 3OXH).

RESULTS AND DISCUSSION

Crystal Structure of nRv0577. Recombinant nRv0577 was expressed in *E. coli* and crystallized by vapor diffusion using a citrate/butanol solution. The crystal structure of this 281-residue protein (including the 20-residue tag) was determined at 1.75 Å resolution using MAD analysis of a selenomethionine derivative (**Table 1**). The refined model (R/R_{free} = 0.176/0.219) contained one molecule each of xylitol and p-chloromercuricbenzene sulfonic acid (PCMBS) used to prepare a heavy-atom derivative,

two chloride ions, and 271 water molecules. No electron density could be mapped to 18 residues of the native Rv0577 sequence: V42-G46, P70-G72, E173-N179, and P259-Q261. The N-terminal hexahistidine tag was partially ordered in both the SeMet and native crystals through degenerate contacts between adjacent molecules that made it possible to model H-13 to S-10 and G-3 to H-1 of the tag into the electron density as well (including the first methionine of the native sequence, labeled M0). Because the overall model was the same in both structures, aside from differences in the linker region of the tag (outside the cleft), only the native structure was completely refined.

Table 1. Summary of the diffraction data collection and refinement statistics for Rv0557.

	SeMet	**Native**
Data Collection		
X-ray source	ALS beamline 8.3.1	ALS beamline 8.3.1
Detector	ADSC Quantum 210	ADSC Quantum 210
X-ray wavelength (Å)	0.892	1.0094
Temperature (K)	100	100
Space group	$P2_12_12$	$P2_12_12$
Unit-cell parameters		
a (Å)	80.65	81.6
b (Å)	83.35	81.98
c (Å)	36.42	36.04
$\alpha = \beta = \gamma$	90º	90º
Matthews coefficient ($Å^3Da^{-1}$)	2.07	2.03
Solvent content (%)	44.6	43.9 %
Resolution range (Å)	41.67-2.25 (2.33–2.25)	20 – 1.75 (1.81 – 1.75)
Mean $I/\sigma(I)$	22.9 (2.0)	16.8 (1.9)
No. of observed reflections	12162 (1162)	25186 (1230)
Completeness (%)	99.9 (98.9)	99.7 (98.9)
Multiplicity	4.0 (4.05)	7.0
R_{merge} (%)[2]	7.6 (28.7)	7.3 (56)
R_{meas} (%)[3]	8.8 (33.0)	8.6
Refinement		
No. of used reflections	-	25186

R_{work} (%)[4]	-	17.6 (21.3)
R_{free} (%)[5]	-	21.2 (23.5)
Mean B factor (Å^2)	-	22.9
RMSD bonds (Å)	-	0.01
RMSD angles (°)	-	1.35
Model Validation		
MolProbity Ramachandran analysis		
Most favored (%)	-	97.1
Additionally allowed (%)	-	2.9
MolProbity[6]		
Clash score, all atoms [percentile]	-	5.73 [93th]
MolProbity score [percentile]	-	1.50 [93th]
PDB identifier	-	3oxh

[1]Value in parenthesis are statistics for the highest resolution shell (1.81 - 1.75 Å).
[2]$R_{merge} = \Sigma_h\Sigma_i|I_i(h) - <I(h)>|/\Sigma_h\Sigma_i I_i(h)$.
[3]$R_{meas} = \{\Sigma_h[n_h/(n_h-1)]^{1/2}\Sigma_i^{nk}|I_h-I_{h,i}|\}|\Sigma_h\Sigma_i^{nh}I_{h,n}$, where $I_h = (1/n_h)\Sigma_i^{nk}I_{h,i}$.
[4]$R_{work} = \Sigma||F_{obs}| - F_{calc}||/\Sigma|F_{obs}|$ where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.
[5]$R_{free} = \Sigma_h||F_{obs}| - F_{calc}||/\Sigma_h|F_{obs}|$. The free R factor was calculated using 5% of the reflections omitted from the refinement.
[6]Calculated adding hydrogens and making the suggested flips.

As illustrated in Figures 2, the four approximate repeats in the Rv0577 sequence adopt a similar structure with a βαβββ topology (↑↑↓↑), comprising a mixed, four-stranded β-sheet with the helix packed parallel to the β-strands. This fold is commonly found in glyoxylases and microbial antibiotic sequestration proteins.[36] The repeat motifs are arranged radially in consecutive pairs with the first β-strand in each pair anti-parallel (β1:β5 and β9:β13). The consequence is two continuous eight-strand β-sheets (β2:β3:β4:β1:β5:β8:β7:β6 and β10:β11:β12:β9:β13:β16:β15:β14) capped on both ends with an α-helix. The two β-sheets intersect in the center at roughly a right angle and form a deep saddle on both sides of the protein with an inter-saddle loop from β8 to β9 running underneath one side of this center. The cores of the two saddles (residues P11 to

A129 (saddle 1) and L143 to A258 (saddle 2)) align well, with an overall Cα root-mean-square deviation (RMSD) of 2.1 Å.

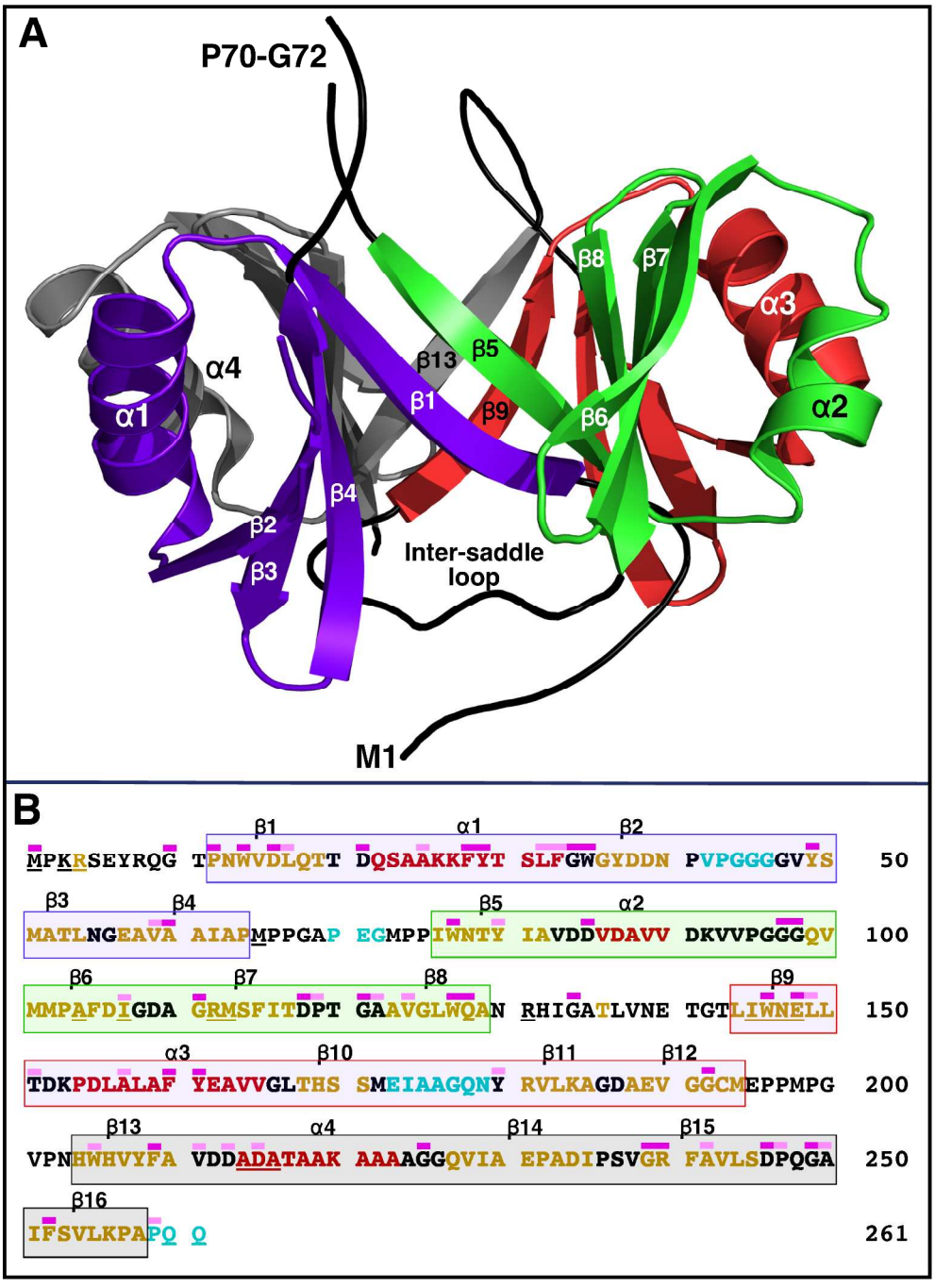


Figure 2. A) Crystal structure of nRv0577 (3OXH) with the four approximate βαβββ repeats colored purple, green, red, and grey sequentially starting from the N-terminus. Shown is the face of the first of two saddles composed of two sequential repeats that form a continuous eight-strand β-sheet (β2:β3:β4:β1:β5:β8:β7:β6) capped on both ends with α1 and α2. The two saddles, connected by an inter-saddle loop between β8 and β9, intersect near the center at roughly a right angle and form a deep cleft on both sides of the protein. **B**) The primary amino acid sequence of Rv0577 with the four approximate repeats shaded in purple, green, red, and grey boxes and the α-helices and β-strands colored red and gold, respectively. The repeats sequentially form two saddles of two repeats each: residues P11 - A129 (saddle 1) and L143 - A258 (saddle 2). Arginine-4 and T136 form a single hydrogen-bond β-bridge. Residues with either missing or unassigned amide resonances in the ^{1}H-^{15}N HSQC spectrum[14] are underlined and residues missing electron density in the crystal structure are colored cyan. The most conserved residues identified by the ConSurf server (http://consurf.tau.ac.il/2016/)[37] are identified with a magenta (9) and pink (8) rectangle above the residue.

No configuration of residues appropriate for binding a metal ion was found in either of the two saddles. Glyoxylases and dioxygenases metalloenzymes in this fold family typically have up to four residues involved in metal coordination.[38] In the *E. coli* glyoxalase, this site is formed by two glutamates and two histidines, which together with two water molecules coordinate a Ni^{2+} (or similar) ion with octahedral geometry[39]. The human glyoxalase I binds zinc in a similar arrangement but with one histidine replaced by glutamine.[40] Although several of these residue types are present in both clefts, their locations do not appear to permit chelation of a metal ion, and several are instead involved in a conserved hydrogen-bonding network (described below). The absence of metal bound to the protein was corroborated by NMR spectrometry where no change was observed in the ^{1}H-^{15}N HSQC spectrum of Rv0577c upon making the solution 5 mM in

EDTA or upon titrating Rv0577c with excess divalent cations (Zn^{2+}, Co^{2+}, and Mg^{2+}; data not shown).

Microbial antibiotic sequestration proteins in this fold family typically bind ligand in one, or both, of the fold saddles without performing any catalytic function.[36] As illustrated in Figure 3, one of the two saddles, saddle 1, was partially occupied by a segment of the hexahistidine tag (yellow) of a crystallographically related molecule in the crystals of both the native and selenomethionine-labeled protein. Only three histidines and a serine from the tag could be distinguished in the density. However, aside from these residues, there was no significant buried surface due to intermolecular association in the crystal and the PDBe PISA server (http://www.ebi.ac.uk/pdbe/pisa/) predicts the biological unit of nRv0577 is a monomer. This conclusion was supported by solution observations. nRv0577 eluted off a Superdex75 26/60 size exclusion column with a retention times (110 min) characteristic of a monomeric protein and a rotational correlation time estimated (20 °C) from a modified ^{1}H-^{15}N HSQC experiment (18.7 ± 1.1 ns) was in the range expected for a monomeric ~28 kDa protein.

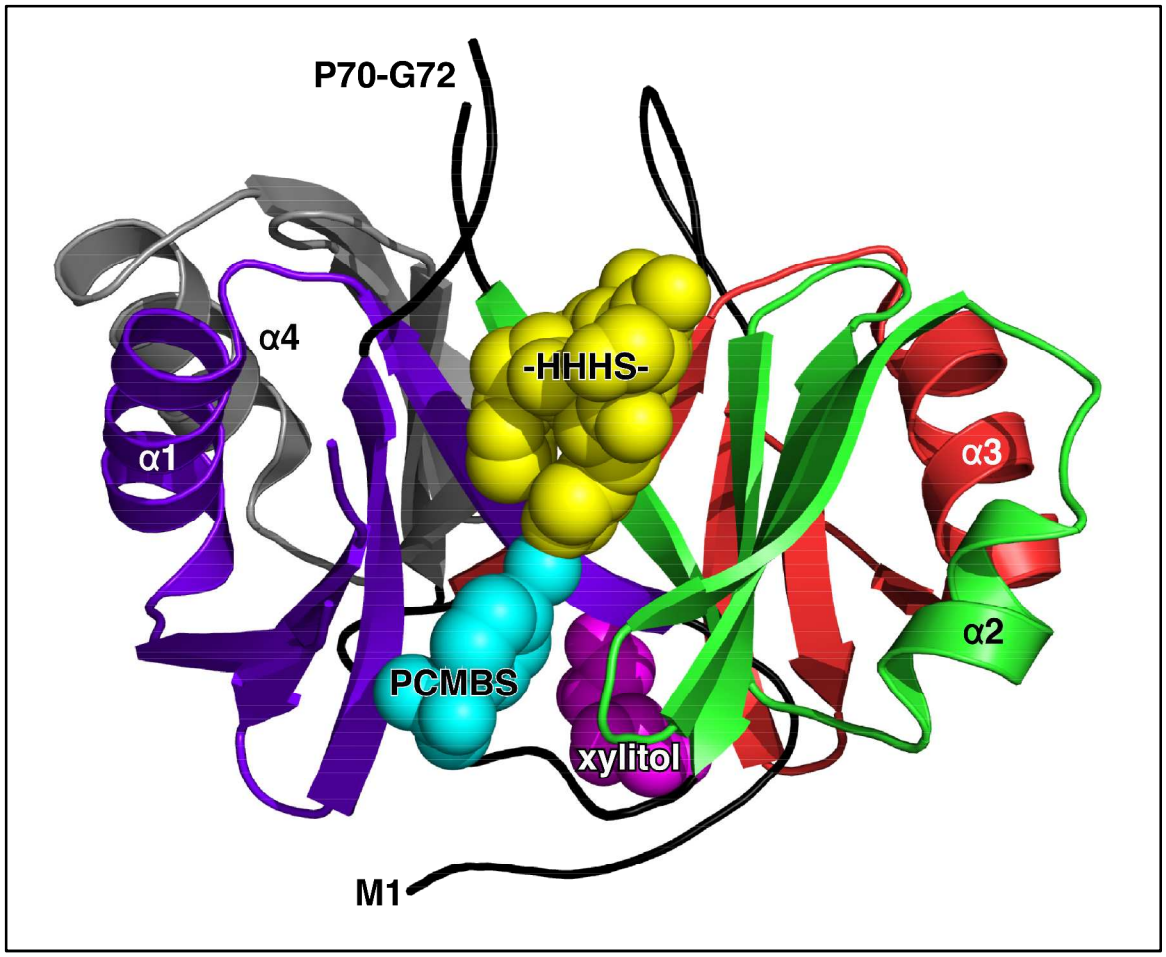


Figure 3. Cartoon representation of the crystal structure of nRv0577 (3OXH) with the ligands observed associated with the structure shown as spherical representations: xylitol (magenta), p-chloromercuricbenzene sulfonic (PCMBS, cyan), and H-13 – S-10 (-HHHS-, yellow). Residues P70-G72 were not observed in the crystal structure.

In addition to a segment of the hexahistidine tag, electron density was also modeled for a molecule of the mercury compound PCMBS (cyan) and the sugar alcohol xylitol (magenta) associated with the protein as shown in Figure 3. The PCMBS is bound to one of the histidines of the tag, and together the PCMBS and tag occupy a large part of the surface area defined by saddle 1, hinting that a binding function for this site may be conserved. Several residues lining this cleft (D16, Q18, D78, and Y80) form hydrogen

bonds with nitrogen atoms in the histidine side chains or the backbone of the tag; however, van der Waals and/or hydrophobic interactions dominate. On the other hand, the xylitol is found in the second saddle, covering a smaller surface area than the PCMBS-tag, and appears to be held by a single hydrogen bond to the side chain of N139.

Rv0577 and the βαββββ-fold Family. Searching for similar folds using the NCBI VAST[41] and DALI[42] servers largely retrieved proteins annotated as dioxygenases, glyoxylases, or glyoxalase/bleomycin resistance proteins. These three families of proteins all contain four βαββββ repeats arrayed either as four tandem repeats in a single polypeptide chain or a dimer of two tandem repeats.[36] Over the complete Rv0577 sequence, only six structures were identified using VAST that contained four tandem βαββββ repeats in a single sequence. The folds of these six proteins aligned poorly and showed limited homology with nRv0577, with backbone RMSD alignments between 3.89 to 4.11 Å and sequence identity ranging between 7 to 15%. Because of this lack of sequence identity, none of these structural homologs (or any of the other proteins in this structural class) were detected by a BLAST search using the primary sequence of Rv0577 or KbpA. All six VAST identified proteins contain a metal binding site, a feature not observed in Rv0577, with the sequence identity confined mostly to hydrophobic residues in the core of the structure. While these six proteins display the same general arrangement of four βαββββ repeats organized into two saddles as observed in nRv0577,[43-47] the absence of a metal binding site suggests Rv0577 does not perform the catalytic functions attributed to dioxygenases and glyoxylases, but instead, recruited the scaffold from a different metabolic pathway to bind ligands.[48]

To assist identification of possible functional sites in Rv0577, we performed multiple sequence alignment of Rv0577 using the ConSurf server (http://consurf.tau.ac.il/2016/)[37] to identify the most conserved residues. These ConSurf identified residues are shown with pink (8-score) and magenta (9-score) bars over the primary amino acid sequence of Rv0577 in Figure 2B. Overall, more conserved residues are observed in saddle 1 (29) over saddle 2 (23) suggesting both clefts are not symmetrical. Such asymmetry is also observed in the dioxygenases where, for example, the metal binding site is present only on one face.[44] Many of the conserved patches are motifs characteristic of the βαββββ proteins in general[36] and include a DPXG sequence at the turn between β7 and β8 in the first saddle and β15 and β16 in the second saddle (Figure 2B). As shown in Figure 4, both saddles contain a conserved quartet of residues, WYD/EY, whose side chains line the surface in the deepest part of both saddles and are candidates for ligand binding. While not among the saddle 1 residues more perturbed by the addition of Ado in the chemical shift perturbation study (Figure 5, discussed in following section), three of the residues in the quartet, W14, D16, and Y49, are adjacent to two of the more perturbed residues, V15 and V48, supporting a ligand binding role for at least part of this conserved quartet surface. Unlike saddle 2, saddle 1 contains a conserved tryptophan (W127) and isoleucine (L107) whose side chains sit near the WYDY quartet, potentially assisting in ligand binding (Q128 is one of the more perturbed residues in the Ado chemical shift perturbation study). The largely hydrophobic nature of these conserved quartet motifs suggest a ligand-binding mechanism similar to the hydrophobic "sandwich" formed in the *Streptomyces* mitomycin-C resistance protein[48] and bleomycin resistance protein[49] where the quinone

moiety of 1,2-*cis*-hydroxy-2,7-diaminomitosene and bithiazole moiety of bleomycin, respectively, is sandwiched between a tryptophan and a histidine, and two tryptophan side chains. With Rv0577, the planar three-ring moiety of neutral red or the base of Ado may be sandwiched between aromatic side chains in the conserved quartet. Indeed, the phenyl ring of PCMBS contacts the edge of this pocket in the crystal structure presented here for nRv0577, sitting half-way over the top of Y49 (Figure 4). To further explore ligand binding to one or both saddles, solution state chemical shift perturbation experiments[15] were performed that made use of the previous amide chemical shift assignments for Rv0577c[14] and the crystal structure presented here for nRv0577.

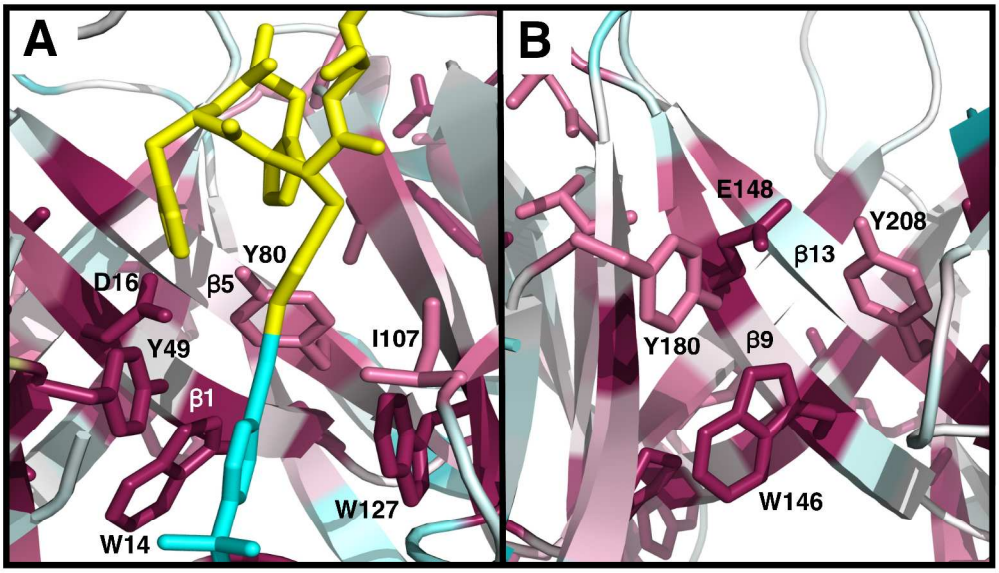


Figure 4. Cartoon representation of the deepest part of the saddles on opposite faces of the crystal structure of nRv0577 highlighting the side chains of conserved residues identified by the ConSurf server (http://consurf.tau.ac.il/2016/).[37] Both saddles feature a highly conserved WYD/EY quartet. **A**) Saddle 1 with the bound PCMBS and -HHHS- shown and colored cyan and yellow, respectively. **B**) Saddle 2 with the xylitol observed

in the crystal structure of nRv0577 not shown because it sits just outside the deep part of the saddle below W146.

Chemical Shift Perturbation Studies. One of the properties of Rv0577 is its association with neutral red staining of virulent strains of *M. tuberculosis.*[11] Multiple attempts to generate a co-crystal of Rv0577c with neutral red were unsuccessful. Consequently, to determine if neutral red bound to Rv0577c and to map the ligand binding surface if it did bind, a chemical shift perturbation study[15, 50] with neutral red was performed by titrating neutral red into ^{15}N-labeled Rv0577c and collecting ^{1}H-^{15}N HSQC spectra at each titration point at dye:Rv0577c molar ratios of up to six to one. As summarized in Figure 5, the major observation was the disappearance of amide cross peaks in the ^{1}H-^{15}N HSQC with the addition of dye. At a dye:Rv0577c molar ratio of 4:1, 34 amide cross peaks (full-lines plus half-line with asterisk) disappeared and the chemical shifts of 11 amide cross peaks (half-lines) shifted slightly. Except for a couple residues, most of the cross peaks in the latter group of 11 had disappeared at a dye:Rv0577c molar ratio of 2:1 although cross peaks were not observed to simply shift until higher dye:Rv0577c molar ratios, typically 4:1 (data not shown). At dye:Rv0577c molar ratios of 5:1 and 6:1, cross peaks reappeared for 5 residues identified with blue asterisks in Figure 4, although at slightly different chemical shifts. Overall, the disappearance of a select set of amide cross peaks in the ^{1}H-^{15}N HSQC spectra of Rv0577c upon addition of neutral red indicates that the dye is binding to the protein. While the majority of the perturbations are scattered through-out the first saddle, there are two pockets of perturbation in the second saddle, suggesting neutral red binds to the surface of both clefts. In summary, neutral red appears to be binding to multiple sites on

Rv0577 with saddle 1 more covered then saddle 2 and suggests that neutral red staining of virulent strains of *Mycobacterium* may be due to the dye physically binding to Rv0577.

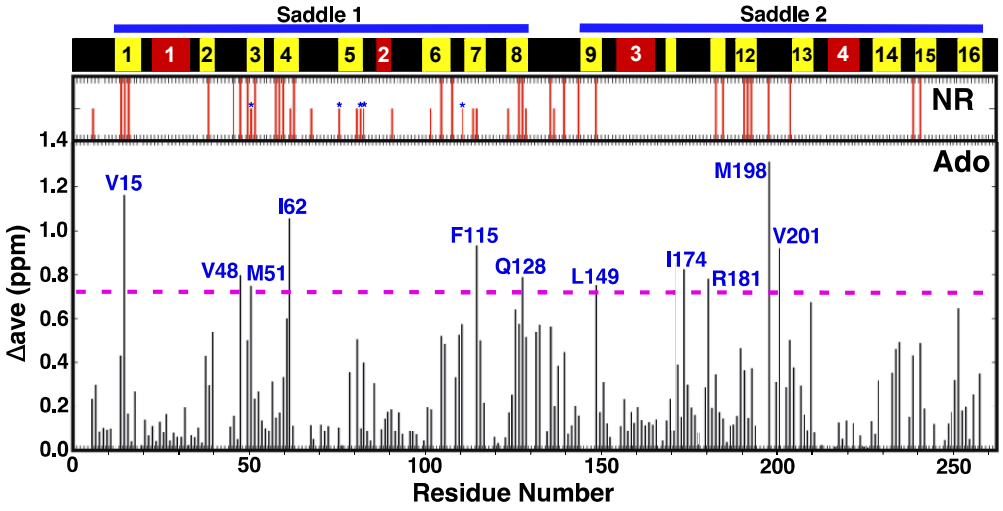


Figure 5. Summary of the Rv0577c titration studies with deoxyadenosine (Ado) and neutral red (NR) at 8:1 and 4:1 molar ratios, respectively. **NR**) The most significant perturbations to the amide chemical shifts of nRv0577 upon titration with neutral red were their disappearance in ^{1}H-^{15}N HSQC spectra. The chemical shifts of a smaller number of amides also shifted with some reappearing after first disappearing at lower neutral red concentrations. Therefore, the plot for neutral red categorizes amides into three groups: missing, shifted, and reappeared with a change in chemical shift following disappearance. A full line represents an amide cross peak that disappeared, a half-line represents an amide cross peak that shifted, and a half-line with a blue asterisk on top represents an amide cross peak that disappeared in early titration points and re-appeared at a different position at higher neutral red molar ratios. **Ado**) Combined average chemical shift change in the amide $^{1}H^{N}$ and ^{15}N resonances of Rv0577c after the addition of an eight-fold molar excess of Ado. The average chemical shift change = Δ_{ave} (ppm) = $[((\Delta^{1}N^{H})^{2} + (\Delta^{15}N/5)^{2})/2]^{1/2}$. The dashed red line represents a Δ_{ave} of 0.07 ppm. On top of the graphs is a schematic representation of the elements of α-helical (red) and β-strand

(yellow) secondary structure observed in the crystal structure of nRv0577 with β-strands 10 and 11 not numbered for clarity.

Using a dye-based affinity chromatography assay for nucleosides and nucleotides, it was shown that Rv0577 bound to deoxyadenosine (Ado) and a couple adenosine base-methylated derivatives (2-methyl adenosine and 6N-methyl adenosine).[51] While this assay indicated Ado bound to Rv0577 less tightly then the methylated derivatives, Ado was used in our chemical shift perturbation studies because it is readily available. As summarized in Figure 5, titration of Ado into Rv0577c resulted in the perturbation of a subset of chemical shifts with increasing Ado concentrations suggesting that Ado was binding to the protein. Amide cross peak in the ^{1}H-^{15}N HSQC spectra were observed to move, and not split into two, during the titration indicating that we were observing the average of two states due to "fast exchange", a feature attributed to low affinity ligand binding ((K_d > 100 μM).[52] Eleven residues were above the arbitrary average chemical shift perturbation threshold of 0.7 ppm and these were almost equally located in saddle 1 and 2. Out of the six most perturbed amide chemical shifts in saddle 1, only the side chains of two residues, M51 and F115, were located on the surface of saddle 1 suggested that perhaps the other perturbations were due to internal structural changes upon ligand binding. Out of the five most perturbed amide chemical shifts in saddle 2, none are side chain exposed on the surface of saddle 2 with four of the side chains in loops between β-strands. Hence, in terms of direct perturbations to residues with side chains on the surface of each saddle, the surface of saddle 1 was more affected than the surface of saddle 2.

Dissociation Constant (K_d) Measurement. Often a crude estimation of K_d may be obtained from the cross peak pattern observed in ^{1}H-^{15}N HSQC spectra collected during titration with a ligand due to the NMR chemical shift timescale. Ligands with high affinity for the protein ($K_d < 10^{-7}$ M) usually are in "slow exchange" and display two resonances representing the bound and unbound state. On the other hand, ligands with low affinity for the protein ($K_d > 10^{-5}$ M) are usually in "fast exchange" and display a single resonance that is the average between the bound and unbound state. In between these two extremes is a region of intermediate affinity (K_d approximately $10^{-5} – 10^{-7}$ M) in "intermediate exchange" where the chemical shifts broaden out beyond detection. Neutral red binding to Rv0577c, characterized largely by the disappearance of amide cross peaks in ^{1}H-^{15}N HSQC spectra, is a feature attributed to intermediate affinity (K_d = 0.5 – 100 mM). On the other hand, Ado binding to Rv0577c, characterized by a shifting, single amide cross peak in ^{1}H-^{15}N HSQC spectra, is a feature attributed to low affinity ligand binding (($K_d > 100$ μM). Together these two observations indicate that Rv0577c had a greater affinity for neutral red than Ado.

To more accurately define K_d the chemical shifts perturbations for the Ado titration were analyzed using a 1:1 binding model and the online webserver www.supramolecular.org. Using the proton chemical shift of V15, V48, M51, I62 and Q128, residues located in the first saddle with the largest chemical shift perturbation, the best fit was obtained using a single binding site to obtain a K_d of 4.2 ± 0.3 mM. On the other hand, it was not possible to obtain a good fit for the proton chemical shifts for any of the five major resonances in the second saddle. In the latter case, the data was best fit with a straight line with a small slope, suggesting weak binding in the M range.

Together, these analyses suggested that saddle 1 was the primary binding site for Ado with weaker binding in saddle 2 (efforts to fit the data to two sites failed likely because binding to the second site was weak). A similar analysis was not possible for neutral red because the disappearance of resonances indicating that binding was in the intermediate exchange region.[52] Moreover, a subset of amide cross peaks only changed chemical shift or disappeared and then re-appeared at higher neutral red concentrations at a slightly different chemical shift, suggest that neutral red binding to Rv0577c was complicated and likely involved multiple events.

To corroborate the K_d determined from the chemical shift perturbation studies, a value was obtained for Ado using isothermal titration calometry.[53] A representative ITC isotherm for the titration of Ado into Rv0577c is shown in Figure 6. Using a binding stiocheometry of 1:1, a K_d of 1.11 ± 0.06 mM was obtained at 37 ºC which is of the same magnitude (4.2 ± 0.3 mM) as obtained from the analysis of the chemical shift perturbation data. Although the chemical shift perturbations suggested Ado bound to both saddles, the ITC data could not be fit to a two-site model. On the other hand, while the ITC experiments with neutral red suggested the ligand was binding to Rv0577c, as suggested by the observations in the chemical shift perturbation study, it was not possible to analyze the data to generate a K_d.

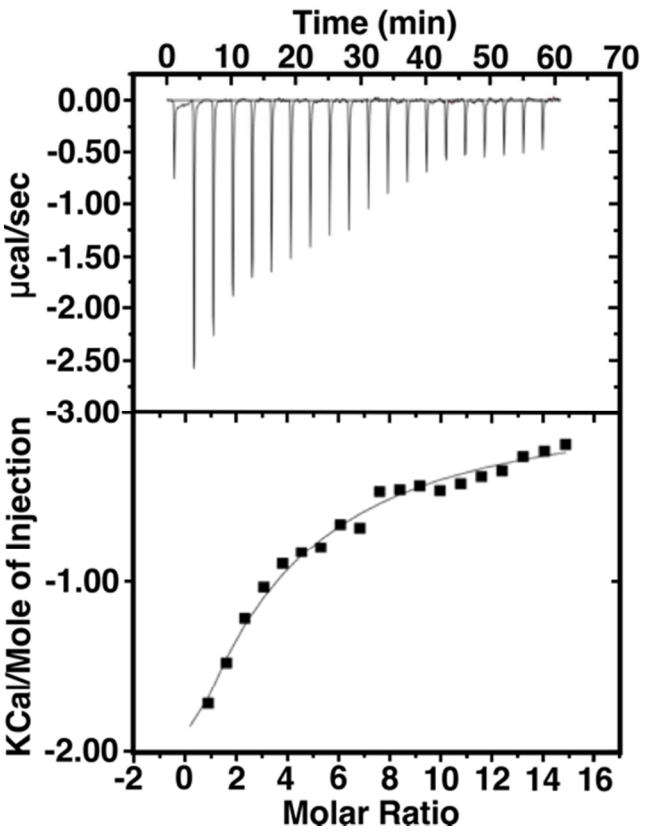


Figure 6. Raw and fit ITC isotherms for Ado titrated into Rv0577c collected at 37 ºC.

Backbone Dynamics. The chemical shift perturbation studies indicate that Ado and neutral red bind to Rv0577c and identify residues that may play a role in ligand binding. However, upon ligand binding it is possible for the chemical shifts of amides distant from the binding site to also be perturbed due to propagated structural changes manifested by ligand binding. An alternative method to identify regions of a protein responsible for ligand binding may be obtained by measuring the relaxation properties of the amide resonances (heteronuclear steady-state $\{^1H\}$-^{15}N NOE, T_2, and T_1) to probe the dynamics along the protein backbone. This is because localized protein dynamics can play a role in the binding and catalysis properties at, and around, active sites. Figure 7 is

a plot of the backbone $\{^1H\}$-^{15}N heteronuclear NOE, T_1, and T_2 values measured for Rv0577c at 30 ºC. Small, or negative, heteronuclear $\{^1H\}$-^{15}N NOE values identify regions of the protein with fast time scale (picosecond) motion.[54] Such fast time scale motion may also effect the longitudinal (T_1) and transverse (T_2) ^{15}N relaxation times at or near these regions. As illustrated in Figure 7A, the majority of the heteronuclear $\{^1H\}$-^{15}N NOE values were above 0.8 with an average value of 0.85 ± 0.13. There were four pockets of slightly lower values clustered in loops between β-strands 2 and 3, 4 and 5, 10 and 11, and 12 and 13. The T_1 and T_2 values, summarized in Figure 7B and 7C, respectively, were also relatively uniform through-out the sequence of Rv0577c with many of the small changes from the average in the same regions containing the smallest heteronuclear $\{^1H\}$-^{15}N NOE values. Overall, the relaxation properties of amide resonances were relatively uniform throughout the backbone with mild differences observed in four loops. Out of these four loops with mild differences, all except the loop between β-strands 4 and 5 contained some of the most significant chemical shift perturbations upon titration with Ado (Figure 5). The loop between β-strands 4 and 5 is adjacent to the site of PCMBS binding in the crystal structure of nRv0577. In summary, the relaxation data do not unambiguously flag any specific regions of the protein with unusual dynamics that may signal a region with a role in ligand binding or catalysis.

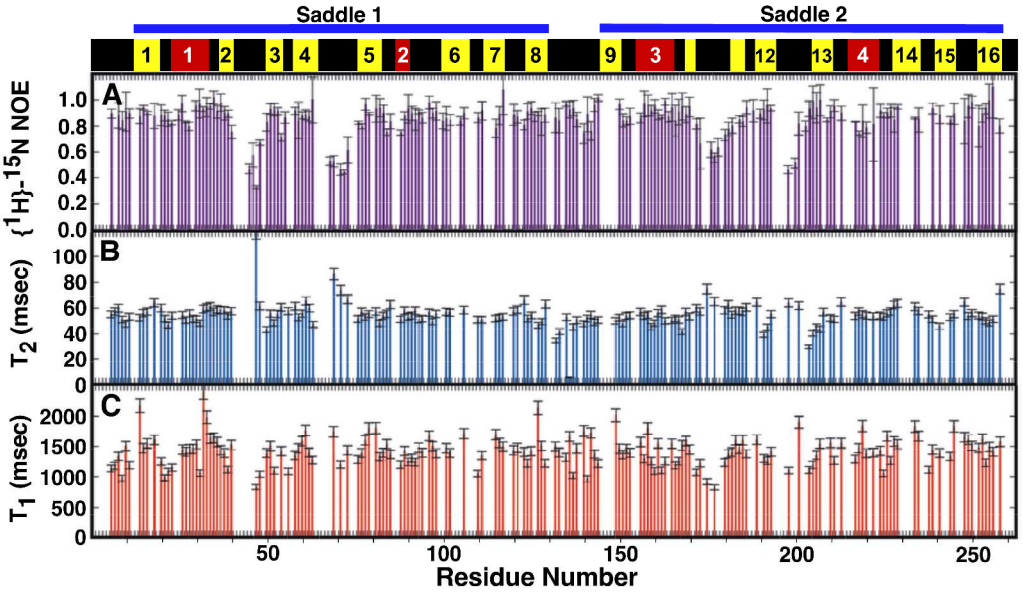


Figure 7. Backbone amide (**A**) $\{^1H\}$-^{15}N heteronuclear NOE (purple), (**B**) T_2 (blue), and (**C**) T_1 (red) values for Rv0577c at 30 ºC. On top of the graphs is a schematic representation of the elements of α-helical (red) and β-strand (yellow) secondary structure observed in the crystal structure of nRv0577 with β-strands 10 and 11 not numbered for clarity.

Note that the T_1 and T_2 values appear consistent with estimations of τ_c for Rv0577c. The average T_1 and T_2 values were 1427 ± 248 and 55 ± 9 ms, respectively, and consistent with a previous estimation of τ_c for Rv0577c at 20 ºC. This consistency conclusion was reached by inserting the average T_1 and T_2 values at 30 ºC into equation (1) to generate a τ_c of 12.7 ns that is near the 12.8 ns value for τ_c calculated using the Stokes equation and the 14.2 ± 0.5 ns value determined for Rv0577c at 20 ºC with ^{15}N-edited one-dimensional experiments.[14]

Transient Protein-Protein Association. As reported earlier[14] and shown in Figure 8, the ^{1}H-^{15}N HSQC spectrum of Rv0577c (magenta) contains amide cross peaks with good chemical shifts dispersion in both the ^{1}H and ^{15}N dimensions, features characteristic of a structured protein.[55] On the other hand, while the ^{1}H-^{15}N HSQC spectrum of nRv0577 (Figure 8 black) also contained amide cross peaks with good chemical shift dispersion in both dimensions, relative to the ^{1}H-^{15}N HSQC spectrum of Rv0577c these amide cross peaks were generally broader and of non-uniform intensity. Moreover, relative to the ^{1}H-^{15}N HSQC spectrum for Rv0577c, 41 cross peaks were missing (circled) and a smaller subset of cross peaks were slightly shifted in the ^{1}H-^{15}N HSQC spectrum for nRv0577. These spectral features are suggestive of intermolecular association[56] and are corroborated by a comparison of the size exclusion chromatography behavior and τ_c estimations for nRv0577 and Rv0577c. nRv0577 eluted off a Superdex75 26/60 size exclusion column with a slightly shorter retention time than Rv0577c, 110 versus 112 min, respectively, suggesting that nRv0577 was behaving as a slightly larger species. The rotational correlation time estimated (20 °C) for nRv0577 was larger than for Rv0577c, 18.7 ± 1.1 versus 14.2 ± 0.5 ns,[14] respectively, also suggesting that nRv0577 was behaving as a larger species. Collectively, these values are not characteristic of dimer or higher order aggregates, and hence, the poor NMR spectral properties of nRv0577 relative to Rv0577c is likely be due to transient association and not stable oligomer formation.[56]

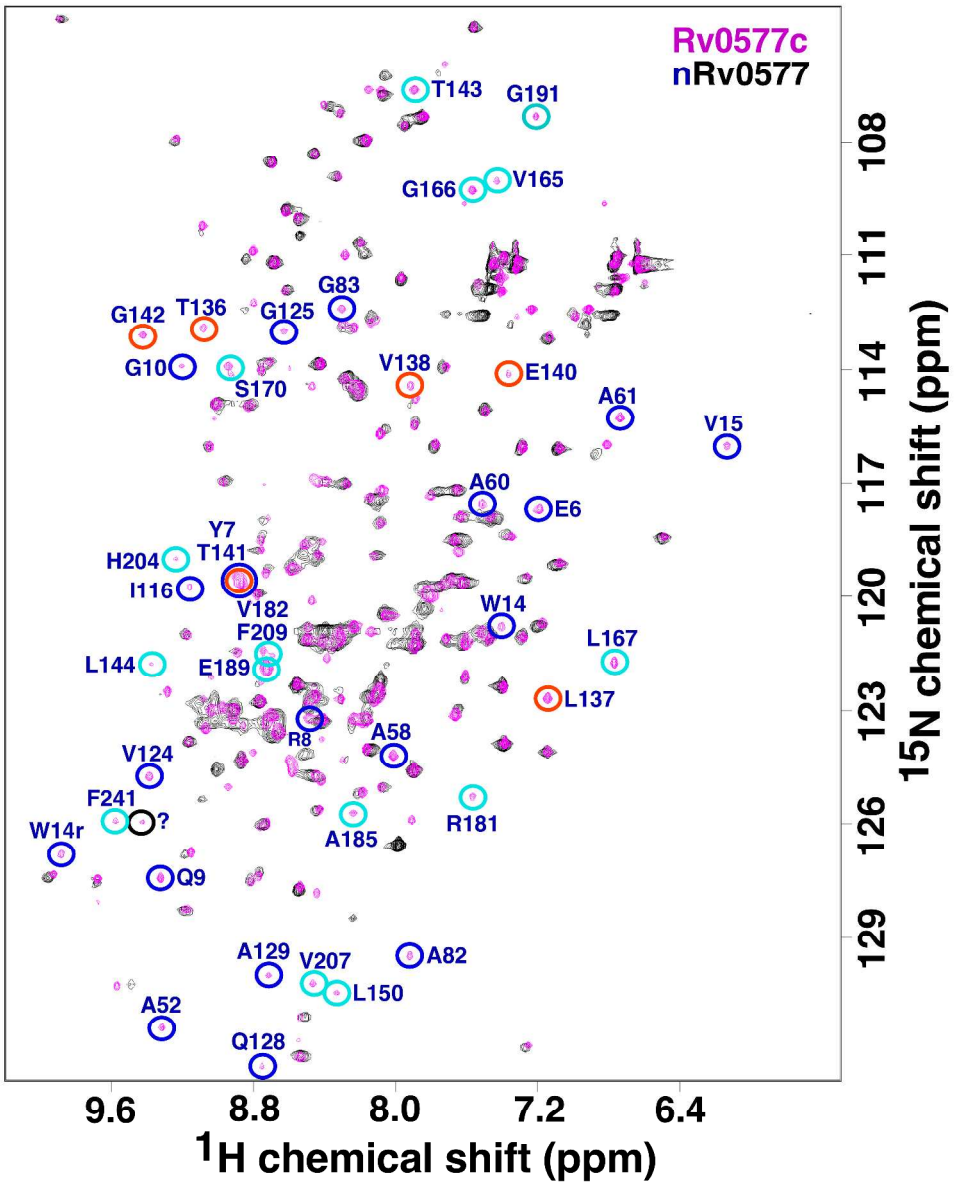


Figure 8. Overlay of ^{1}H-^{15}N HSQC spectra for Rv0577c (magenta) and nRv0577 (black) collected at a ^{1}H resonance frequency of 800 MHz, 30 ºC. The amide cross peaks observed in Rv0577c but missing in nRv0577 are circled, labeled, and colored blue (saddle 1), cyan (saddle 2), or orange (inter-saddle linker).

In the crystal structure reported here for nRv0577, part of the N-terminal polyhistidine tag in one nRv0577 molecule is observed bound in saddle 1 of a second nRv0577 molecule (Figure 3), providing strong support for the role of the N-terminal tag

in initiating intermolecular interactions between nRv0577 molecules. The distribution of the 41 missing resonances in the ^{1}H-^{15}N HSQC spectrum for nRv0577 relative to Rv0577c suggests this interaction is intermolecular and not intramolecular. In Figure 8, nRv0577 amide resonances missing for residues in saddle 1, saddle 2, or the inter-saddle linker are colored blue, cyan, and orange, respectively. Eighteen backbone amide resonances are missing in saddle 1, 15 in saddle 2, and six in the inter-saddle linker region. If the N-terminal tag folded into saddle 1 intramolecularly, amide resonances should primarily be missing in saddle 1 due to intermediate exchange or heterogeneous interactions. Because a similar number of amide resonances are missing in both saddles suggests the N-terminal tag folds over saddle 2 intramolecularly and binds to saddle 1 intermolecularly with intermediate exchange or heterogeneous binding surface contributing to the missing amide resonances at the dimer interface.[57] Because the major difference between the two constructs of Rv0577 was the tag at either termini of the protein, it was hypothesized that the transient interactions observed in solution were manifest by the 20-residue N-terminal tag. Three experiments were conducted to test this hypothesis with mixed results.

First, the chemical shift binding experiments presented here show that Ado binds to Rv0577c with greater affinity for saddle 1. Because saddle 1 is the same face where part of the polyhistidine tag is observed in the crystal structure of nRv0577, it was postulated that it may be possible to reverse transient association by adding Ado to nRv0577 to displace the polyhistidine tag. This was observed upon the addition of approximately eight-fold molar excess dA to nRv0577; all the amide resonances missing in the ^{1}H-^{15}N HSQC spectrum of nRv0577 returned and the overall cross peak line shape

became more uniform (data not shown), suggesting that Ado and the polyhistidine tag shared the same binding surface on nRv0577 with the protein having greater affinity for Ado than the N-terminal tag.

Second, a synthetic peptide corresponding to the N-terminal of nRv0577 (MGSSHHHHHHSSGLVPR; Genscript, Piscataway, NJ) was titrated into Rv0577c. Even at a peptide:Rv0577c molar ratio of 10:1 there was no evidence of perturbations in the ^{1}H-^{15}N HSQC spectrum of Rv0577c to suggest this peptide bound to the protein. This experiment indicates that if the intermolecular association of nRv0577 was initiated by the N-terminal tag, it was very weak (as may be expected by its transient nature). It is also possible that while the N-terminal tag may be responsible for initiating transient association, interactions between other regions on the protein surface also contribute to the intermolecular association.

Third, if the N-terminal tag of nRv0577 was solely responsible for the transient association, removing the tag should eliminate transient association and generate an ^{1}H-^{15}N HSQC spectrum similar to the one observed for Rv0577c. Treatment of nRv0577 with thrombin removed most of the N-terminal tag to leave only three non-native residues at the N-terminal. Aside from the appearance of four additional resonances, the ^{1}H-^{15}N HSQC spectrum of cleaved nRv0577 was essentially identical to the ^{1}H-^{15}N HSQC spectrum for uncut nRv0577 indicating that transient association was still occurring with cleaved nRv0577. Consequently, either the three additional N-terminal residues, GSH-, was sufficient to effect transient association, or, the C-terminal tag, -RSHHHHHH, sterically prevented transient association with saddle 2 (T11 is the first residue in an element of secondary structure in the crystal structure, hence, there is a

significant, 10-residue N-terminal available for intermolecular interactions). Regardless of the mechanism responsible for transient association in nRv077 and not Rv0577c, the transient association with nRv0577 appeared to assist crystallization as crystal screens with Rv0577c or cleaved nRv0577 failed to generate crystals. This is consistent with the conclusion that a good ^{1}H-^{15}N HSQC spectrum is not predictive of crystallization success[58] and an example where leaving the purification tag on the native protein assisted crystal growth.

Thermal Stability. Circular dichroism, a valuable technique to characterize protein structure in solution and follow gross structural changes under different conditions,[24, 59] was used to characterize the structure of Rv0577c and the stability of this structure over a range of temperatures. Figure 9A shows the steady-state wavelength spectrum for Rv0577c in NMR buffer at 25 ºC. The profile is highlighted by a single minimum at 218 nm with a very faint suggestion of a shoulder around 210 nm and an extrapolated maximum under 200 nm. Such a profile is characteristic of a folded protein dominated by β-strand secondary structure[60, 61] as expected given the crystal structure (43% β-strand, 15% α-helical). To assess the thermal stability of Rv0577c, the ellipticity at 218 nm was measured at 2.0 degree intervals between 10 and 80 ºC. Typically, during the denaturation of a structured protein with increasing temperature the transition may be followed by the concomitant increase in ellipticity of the negative band.[62] The raw temperature data for Rv0577c is plotted in Figure 9B and shows the protein is relatively stable to ~ 50 ºC and then a very distinct inflection point characteristic of a transition to a denatured state is observed. Insoluble precipitate was visible when the temperature of the

sample was returned to 25 ºC, indicating the transition was irreversible and preventing a thermodynamic analysis of the CD data.[62] However, a quantitative estimation of the T_m for this transition can be obtained by assuming a two-state model and taking a first derivative of the curve in Figure 9B.[24] The maximum of this first derivative, shown in Figure 9C, is 56 ºC.

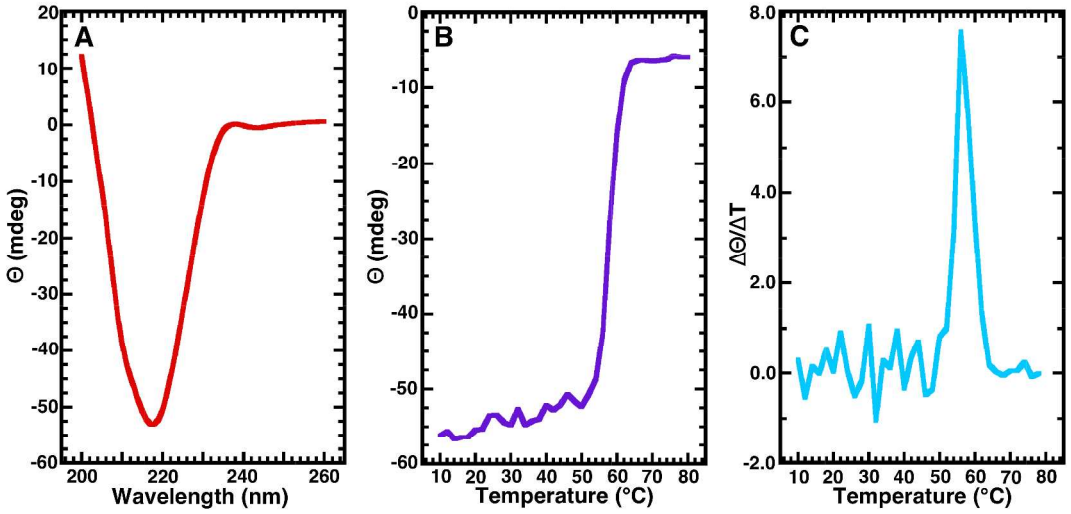


Figure 9. A) Circular dichroism steady-state wavelength spectrum for Rv0577c (0.02 mM) at 25 ºC obtained in buffer containing 100 mM NaCl, 20 mM Tris, 1.0 mM dithiothreitol , pH 7.0. **B**) The CD thermal melt for Rv0577c obtained by measuring the ellipticity at 220 nm in 2.0 ºC intervals between 10 and 80 ºC. **C**) First derivative of the 220 nm thermal melt curve indicates the protein has a melting temperature of 56 ºC.

The KpbA Family of Kinase Regulators. Rv0577 has 36% sequence identity with the *S. coelicolor* kinase regulator KbpA and the structure reported here for Rv0577 represent the first from the KbpA family of proteins. In *S. coelicolor* the protein KbpA and its six known paralogs regulate AfsK and the 36 known *S. coelicolor* kinases by binding to AfsK to prevent activation by autophosphorylation (Figure 1).[6] While the mechanism of kinase recognition and regulation by KpbA remains to be determined,

protein-protein interactions between AfsK and KpbA are one component in the proposed mechanism. The transient association observed in the ^{1}H-^{15}N HSQC spectrum of nRv0577 (Figure 8) and the N-terminal polyhistidine tag observed bound to the protein in the crystal structure (Figure 3) imply that Rv0577 has a disposition for protein-protein interactions. The asymmetry of the two saddles in Rv0577, including a wider conserved hydrophobic binding surface in one saddle (Figure 4), offers potential mechanisms of kinase regulation for KpbA. If these clefts are functionally distinct, one face of KbpA may bind an AfsK monomer preventing autophosphorylation. Autophosphorylation prevention may be accomplished by KpbA sterically blocking phosphorylation sites or by preventing kinase dimer formation which is necessary for the *M. tuberculosis* Ser/Thr kinase PknB.[63] In turn we speculate that KpbA may release ApsK upon binding an internal ligand (Figure 1A) on the opposite open saddle by effecting structural changes in the protein binding saddle that causes the protein to be released, or, the internal target ligand may simply displace the bound protein. In mild support of the latter hypothesis, the titration studies reported here showed that the addition of Ado to nRv0577 prevented transient protein-protein association.

Biological Role of Rv0577 in *M. tuberculosis* Infections. The first evidence that Rv0577 was secreted by *M. tuberculosis* was its identification in cell culture filtrates.[64] Shortly thereafter it was observed that clinical isolates from TB patients expressed elevated levels of Rv0577 relative to laboratory-adapted strains of *M. tuberculosis*.[9] Moreover, in part of the same study Rv0577 was identified in the sputum of TB patients with the suggestion it may be a good diagnostic marker for the disease.[9] Using an *M.*

tuberculosis cosmid library in *Mycobacterium smegmatis* (a closely related, non-pathogenic *Mycobacterium* species), Rv0577 was identified as the responsible component for neutral red staining of virulent strains of *M. tuberculosis,*[11] a well-established assay to distinguish virulent strains of *Mycobacterium*.[65] The chemical shift perturbations studies reported here corroborate this last finding, showing that purified Rv0577 binds to neutral red. Rv0577 was also identified as a highly antigenic protein in exosomes released by *Mycobacterium* infected macrophages.[10] These observations are important because they suggest that Rv0577 may be an *M. tuberculosis* virulence factor and its secretion part of the bacteria's defense against host mycobacteriocidal immune mechanisms.[9]

The host immune response consists of an orchestrated balance of pro- and anti-inflammatory events to remove, but not over-zealously, threat agents. Secretion of Rv0577 during *M. tuberculosis* infection may serve to tilt the delicate balance in its favor.[66] For example, interleukin-10 is a transforming growth factor that is part of the anti-inflammatory cascade and elevated levels of interleukin-10 have been correlated with *M. tuberculosis* infections.[9] Rv0577 has also been reported to be an agonist of toll-like receptor 2 (TLR-2), inducing dendrite cell maturation and driving a Th1 immune response,[12] all pro-inflammatory events to combat infection.[67] However, there is evidence suggesting that TLR-2 signaling by *M. tuberculosis* may be part of a complex strategy used by *M. tuberculosis* to escape the immune response as TLR-2 stimulation has a greater effect on the levels of anti-inflammatory modulators, such an interleukin-10, than pro-inflammatory modulators.[68] Protein-protein interactions between Rv0577 and TLR-2 have been shown[12] and are required for TLR-2 stimulation. Protein-protein interactions are also required for the protein homologous to Rv0577 in *S. coelicolor,*

KbpA, to inhibit AfsK autophosphorylation. With regards to TB, perhaps *M. tuberculosis* recruited the scaffold from *S. coelicolor*, bacteria in the same taxonomical order, to bind to TLR-2.

In addition to its potential role as a virulence factor, Rv0577 has also been associated with a putative methylglyoxal detoxification pathway involving the metabolism of glycerol.[69] The association was made with the identification of a group of novel pyrimidine-imidazole compounds potent to *M. tuberculosis* cells cultured *in vitro* with media containing glycerol but not potent to *M. tuberculosis* cells grown *in vivo* where glycerol was absent. Rv0577 was identified by pulling it out of a mycobacterial cell lysate passed over sepharose beads containing a covalently linked pyrimidine-imidazole. The hypothesis was that the pyrimidine-imidazoles bind tightly to Rv0577, preventing methylglyoxal removal/neutralization and the accumulation of toxic sugar phosphates. While the mechanism of putative methylglyoxal detoxification by Rv0577 is unknown and its role in methylglyoxal detoxification may be surreptitious, like other similarly structured four βαβββ repeat proteins, such as the mitomycin-C and bleomycin resistance proteins, one of Rv0577's biological roles may involve self-protection from and/or exportation of harmful compounds generated internally.

While the crystal structure reported here for Rv0577 along with previous NMR chemical shift assignments[14] provide blueprints for understanding the function of Rv0577, further research is necessary to tease out the potential biological role(s) of Rv0577 in *M. tuberculosis* infections. Additional clues may be revealed by extending our efforts to better understanding the protein Rv0576. Transcriptional analysis of the *M. tuberculosis* genome suggests that the *Rv0577* gene is part of an operon containing the

gene *Rv0576* upstream that codes for a 434-residue protein.[11] While Rv0576 has not been biophysically characterized, a BLAST search indicates that the N-terminal contains a region similar to transcriptional regulators and the C-terminal contains a region similar to conserved hypothetical proteins. It may also be useful to explore if Rv0577 has a kinase regulatory role, similar to the one reported for its homolog KbpA in *Streptomyces*. Protein kinase and phosphatase signaling play a significant role in helping *M. tuberculosis* adapt to the hostile host environment during infection.[70] Interfering with proteins in this signaling machinery may be a promising strategy for the next generation of anti-tuberculosis drugs.

ACKNOWLEDGEMENTS

We thank James Holton, George Meigs, and Jane Tanamachi for support at ALS beamline 8.3.1. Funding for this research was provided by the National Institute of Allergy and Infectious Diseases (NAIAD), National Institute of Health (NIH), Department of Health and Human Services, under Federal Contract numbers HHSN272201200025C and HHSN272200700057C, by a NIH grant awarded to Tom Albers (R01 GM070962), and by the National Institute of General Medical Sciences, NIH, under Federal Contract number 5U01 GM094568. The SSGCID internal ID for Rv0577c is MytuD.17269.a. Part of this research was performed at the W.R. Wiley Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility located at Pacific Northwest National Laboratory (PNNL) and sponsored by U.S.

Department of Energy's Office of Biological and Environmental Research (BER) program. Battelle operates PNNL for the U.S. Department of Energy.

REFERENCES

[1] Bentley, S. D., Chater, K. F., Cerdeno-Tarraga, A.-M., Challis, G. L., Thomson, N. R., James, K. D., Harris, D. E., Quall, M. A., Kleser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C. W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Klesser, T., Larke, L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wletzorrek, A., Woodward, J., Barrell, B. G., Parkhill, J., and Hopwood, D. A. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2), *Nature 417*, 141-147.

[2] Wright, L. F. (1976) Actinorhodin is chromosomally-determined antibiotic in *Streptomyes coelicolor* A3(2), *J. Gen. Microbiol. 96*, 289-297.

[3] Lee, P. C., Umeyama, T., and Horinouchi, S. (2002) afsS is a target of AfsR, a transcriptional factor with ATPase activity that globally controls secondary metabolism in *Streptomyces coelicolor* A3(2), *Mol. Microbiol. 43*, 1413-1430.

[4] Matsumoto, A., Hong, S. K., Ishizuka, H., Horinouchi, S., and Beppu, T. (1994) Phosphorylation of the AfsR protein involved in secondary metabolism in *Streptomyces* species by a eukaryotic-type protein kinase, *Gene 146*, 47-56.

[5] Umeyama, T., Lee, P.-C., and Horinouchi, S. (2002) Protein serine/threonine kinases in signal transduction for secondary metabolism and morphogenesis in *Streptomyces*, *Appl. Microbiol. Biotechnol. 59*, 419-425.

[6] Umeyama, T., and Horinouchi, S. (2001) Autophosphorylation of a bacterial serine/threonine kinase, AfsK, is inhibited by KbpA, an AfsK-binding protein, *J. Bacteriol. 183*, 5506-5512.

[7] WHO. (2014) Global tuberculosis report 2014, United Nations, Geneva.

[8] Av-Gay, Y., and Everett, M. (2000) The eukaryotic-like Ser/Thr protein kinases of *Mycobacterium tuberculosis*, *Trends Microbiol. 8*, 238-244.

[9] Hourd, R. C., Chitale, S., Leung, M., Lazzarini, L. C., Zhu, H., Shashkina, E., Laal, S., Conde, M. B., Kritski, A. L., Belisle, J. T., Kreiswirth, B. N., de Silvia, J. R. L., and Ho, J. L. (2003) The *Mycobacterium tuberculosis* complex-restricted gene CFP32

encodes an expressed protein that is detectable in tuberculosis patients and is positively correlated with pulmonary interleukin-10, *Infect. Immun. 71*, 6871-6883.

[10] Giri, P. K., Kruh, N. A., Dobos, K. M., and Schorey, J. S. (2010) Proteomic analysis identifies highly antigenic proteins in exosomes from M. tuberculosis-infected and culture filtrate protein-treated macrophages, *Proteomics 10*, 3190-3202.

[11] Andreu, N., Soto, C. Y., Roca, I., Martin, C., and Gibert, I. (2004) *Mycobacterium smegmatis* displays the *Mycobacterium tuberculosis* virulence-related neutral red character when expressing the Rv0577 gene, *FEMS Microbiol. 231*, 283-289.

[12] Byun, E.-H., Kim, W. S., Kim, J.-S., Jung, I. D., Park, Y.-M., Kim, H.-J., Cho, S.-N., and Shin, S. J. (2012) *Mycobacterium tuberculosis* Rv0577, a novel TLR2 agonist, induces maturation of dendritic cells and drives Th1 immune response, *FASEB J. 26*, 2695-2711.

[13] Stacy, R., Begley, D. W., Phan, I., Staker, B. L., Van Voorhis, W. C., Varani, G., Buchko, G. W., Stewart, L. J., and Myler, P. J. (2011) Structural genomics of infectious disease drug targets: the SSGCID, *Acta Cryst. F67*, 979-984.

[14] Buchko, G. W., Kim, H., Myler, P. J., Terwilliger, T. C., and Kim, C.-Y. (2012) Chemical shift assignments for Rv0577, a putative glyoxylase associated with virulence from *Mycobacterium tuberculosis*, *Biomol. NMR Assign. 6*, 43-46.

[15] Zuiderweg, E. R. P. (2002) Mapping protein-protein interactions in solution by NMR spectroscopy., *Biochemistry 41*, 1-7.

[16] Good, M. C., Greenstein, A. E., Young, T. A., Ng, H. L., and Alber, T. (2004) Sensor domain of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinase, PknD, forms a highly symmetric beta propeller, *J. Mol. Biol. 360*, 398-408.

[17] van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Chardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin, *J. Mol. Biol. 229*, 105-124.

[18] Doublie, S. (1997) Preparation of selenomethionyl protiens for phase determination, *Meth. Enzy. 276*, 523-530.

[19] Studier, W. F. (2005) Production of auto-induction in high-density shaking cultures, *Protein Expr. Purif. 41*, 207-234.

[20] Rossi, P., Swapna, G. V. T., Huang, Y. P. J., Aramini, J. M., Anklin, C., Conover, K., Hamilton, K., Xiao, R., Acton, T. B., Ertekin, A., Everett, J. K., and Montelione, G. T. (2010) A microscale protein NMR sample screening pipeline, *J. Biomol. NMR 46*, 11-22.

[21] Buchko, G. W., Tarasevich, B. J., Bekhazi, J., Snead, M. L., and Shaw, W. J. (2008) A solution NMR investigation into early events of amelogenin nanosphere self-

assembly initiated with sodium chloride or calcium chloride, *Biochemistry 47*, 13215-13222.

[22] Farrow, N. A., Muhandiram, D. R., Singer, A. U., Pascal, S. M., Kay, L. E., Gish, G., Shoelson, S. E., Pawson, T., and Forman-Kay, J. D. (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homlogy 2 domain studied by ^{15}N NMR relaxation, *Biochemistry 33*, 5984-6003.

[23] Goddard, T. D., and Kneller, D. G. Sparky 3, University of California, San Francisco.

[24] Greenfield, N. J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions, *Nat. Prot. 6*, 2527-2535.

[25] Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, *Meth. Enzy. 276*, 307-326.

[26] Terwilliger, T. C., and Berendzen, J. (1999) Automated MAD and MIR structure solution, *Acta Cryst. D55*, 849-861.

[27] Terwilliger, T. C. (2000) Maximum-likelihood density modification, *Acta Cryst. D56*, 965-972.

[28] de La Fortelle, E., and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods, *Meth. Enzy. 276*, 472-494.

[29] Perrakis, A., Morris, R. M., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement, *Nat. Struct. Biol. 6*, 458-463.

[30] Holton, J., and Alber, T. (2004) Automated protein crystal structure determination using elves, *Proc. Natl. Acad. Sci. USA 101*, 1537-1542.

[31] Kissinger, C. R., Gehlhaar, D. K., and Fogel, D. B. (1999) Rapid automated molecular replacement by evolutionary search, *Acta Cryst. D53*, 240-255.

[32] Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, *Acta Cryst. D Biol. Cryst. 53*, 240-255.

[33] Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Cryst. A47*, 110-119.

[34] Lovell, S. C., Davis, I. W., Arendall III, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. L. (2003) Structure validation by C-alpha geometry: phi, psi, and C-beta deviation., *Proteins 50*, 437-450.

[35] Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the sterochemical quality of protein structures, *J. Appl. Cryst. 26*, 283-291.

[36] Bergdoll, M., Eltis, L. D., Cameron, A. D., Dumas, P., and Bolin, J. T. (1998) All in the family: structural and evolutionary relationships among three modular proteins with diverse functions and variable assembly, *Protein Sci. 7*, 1661-1670.

[37] Glaser, F., Pupko, T., Pas, I., Bell, R. E., Bechor, D., Martz, E., and Ben-Tal, N. (2003) Consurf: identification of functional regions in proteins by surface-mapping of phylogenetic information, *Bioinformatics 19*, 163-164.

[38] Armstrong, R. N. (2000) Mechanistic diversity in a metalloenzyme superfamily, *Biochemistry 39*, 13625-13632.

[39] He, M. M., Clugston, S. L., Honek, J. F., and Matthews, B. W. (2000) Determination of the structure of *Escherichia coli* glyoxalase I suggests a structural basis for differential metal activation, *Biochemistry 39*, 8719-8727.

[40] Cameron, A. D., Olin, B., Ridderstrom, M., Mannervik, B., and Jones, T. A. (1997) Crystal structure of human glyoxalase I--evidence for gene duplication and 3D domain swapping, *EMBO J. 16*, 3386-3395.

[41] Madej, T., Gibrat, J. F., and Bryant, S. H. (1995) Threading a database of protein cores, *Proteins 23*, 356-369.

[42] Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res. 38*, W545-549.

[43] Fritze, I. M., Linden, L., Freigang, J., Auerbach, G., Huber, R., and Steinbacher, S. (2004) The crystal structures of Zea mays and Arabidopsis 4-hydroxyphenylpyruvate dioxygenase, *Plant Physiol. 134*, 1388-1400.

[44] Senda, T., Sugiyama, K., Narita, H., Yamamoto, T., Kimbara, K., Fukuda, M., Sato, M., Yano, K., and Mitsui, Y. (1996) Three-dimensional structures of free form and two substrate complexes of an extradiol ring-cleavage type dioxygenase, the BphC enzyme from *Pseudomonas* sp. strain KKS102, *J. Mol. Biol. 255*, 735-752.

[45] Brownlee, J., He, P., Moran, G. R., and Harrison, D. H. (2008) Two roads diverged: the structure of hydroxymandelate synthase from *Amycolatopsis orientalis* in complex with 4-hydroxymandelate, *Biochemistry 47*, 2002-2013.

[46] Uragami, Y., Senda, T., Sugimoto, K., Sato, N., Nagarajan, V., Masai, E., Fukuda, M., and Mitsu, Y. (2001) Crystal structures of substrate free and complex forms of reactivated BphC, an extradiol type ring-cleavage dioxygenase, *J. Inorg. Biochem. 83*, 269-279.

[47] Pratter, S. M., Konstantinovics, C., Di Giuro, C. M., Leitner, E., Kumar, D., de Visser, S. P., Grogan, G., and Straganz, G. D. (2013) Inversion of enantioselectivity of a mononuclear non-heme iron(II)-dependent hydroxylase by tuning the interplay of metal-center geometry and protein structure, *Angew. Chem. Int. Ed. Engl. 52*, 9677-9681.

[48] Martin, T. W., Dauter, Z., Devedjiev, Y., Sheffield, P., Jelen, F., He, M., Sherman, D. H., Otlewski, J., Derewenda, Z. S., and Derewenda, U. (2002) Molecular basis of mitomycin C resistance in streptomyces: structure and function of the MRD protein, *Structure 10*, 933-942.

[49] Maruyama, M., Kumagai, T., Matoba, Y., Hayashida, M., Fujii, T., Hata, Y., and Sugiyama, M. (2001) Crystal structures of the transposon Tn5-carried bleomycin resistance determinant uncomplexed and complexed with bleomycin, *J. Biol. Chem. 276*, 9992-9999.

[50] Buchko, G. W., Daughdrill, G. W., de Lorimier, R., Rao, B. K., Isern, N. G., Lingbeck, J. M., Taylor, J. S., Wold, M. S., Gochin, M., Spicer, L. D., Lowry, D. F., and Kennedy, M. A. (1999) Interactions of human nucleotide excision repair protein XPA with DNA and RPA70ΔC327: Chemical shift mapping and ^{15}N NMR relaxation studies., *Biochemistry 38*, 15116-15128.

[51] Roberts, J. K. M., Webster, C., Terwilliger, T. C., and Kim, C.-Y. (2008) High-throughput analysis of nucleoside- and nucleotide-binding by proteins, In *Systems Chemistry Symposium* (Hicks, M. G., and Kettner, C., Eds.), pp 207-219, Beilstein-Institut, Bozen, Italy.

[52] Teilum, K., Kunze, M. B., Erlendsson, S., and Kragelund, B. B. (2017) (S)Pinning down protein interactions by NMR, *Protein Sci. 26*, 436-451.

[53] Freyer, M. W., and Lewis, E. A. (2008) Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions, *Methods Cell Biol. 84*, 79-113.

[54] Mandel, A. M., Akke, M., and Palmer, A. G., 3rd. (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme, *J. Mol. Biol. 246*, 144-163.

[55] Lee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C.-H., Cort, J. R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L. P., Gehring, K., Kennedy, M. A., Edwards, A. M., and Arrowsmith, C. H. (2002) An NMR approach to structural proteomics., *Proc. Natl. Acad. Sci. USA 99*, 1825-1930.

[56] Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B., Ramelot, T., Lee, G. M., Bhattacharyya, S., Gutierrez, P., Denisov, A., Lee, C.-H., Cort, J. R., Kozlov, G., Liao, J., Finak, G., Chen, L., Wishart, D., Lee, W., McIntosh, L. P.,

Gehring, K., Kennedy, M. A., Edwards, A. M., and Arrowsmith, C. H. (2002) An NMR approach to strutural proteomics, *Proc. Natl. Acad. Sci. USA 99*, 1825-1830.

[57] Buchko, G. W., Edwards, T. E., Hewitt, S. N., Phan, I. Q., Van Voorhis, W. C., Miller, S. I., and Myler, P. J. (2015) Backbone chemical shift assignments for the sensor domain of the *Burkholderia pseudomallei* histidine kinase RisS: "missing" resonances at the dimer interface, *Biomol. NMR Assign. 9*, 381-385.

[58] Yee, A. A., Savchenko, A., Ignachenko, A., Lukin, J., Xu, X., Skarina, T., Evdokimova, E., Liu, C. S., Semesi, A., Guido, V., Edwards, A. M., and Arrowsmith, C. H. (2005) NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins, *J. Am. Chem. Soc. 127*, 16512-16517.

[59] Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism, *Biochim. Biophys. Acta 1751*, 119-139.

[60] Holzwarth, G. M., and Doty, P. (1965) The ultraviolet circular dichroism of polypeptides, *J. Amer. Chem. Soc. 87*, 218-228.

[61] Woody, R. W. (1974) *Studies of theoretical circular dichroism of polypeptides: Contributions of β-turns*, John Wiley & Sons, New York.

[62] Karantzeni, I., Ruiz, C., Liu, C.-C., and LiCata, V. J. (2003) Comparative thermal denaturation of *Thermus aquaticus* and *Escherichia coli* type 1 DNA polymerases, *Biochem. J. 374*, 785-792.

[63] Lombana, T. N., Echols, N., Good, M. C., Thomsen, N. D., Ng, H. L., Greenstein, A. E., Falick, A. M., King, D. S., and Alber, T. (2010) Allosteric activation mechanism of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinase, PknB, *Structure 18*, 1667-1677.

[64] Rosenkrands, I., Weldingh, K., Jacobsen, S., Hansen, C. V., Florio, W., Gianetri, I., and Andersen, P. (2000) Mapping and identification of *Mycobacterium tuberculosis* proteins by two-dimensional gel electrophoresis, microsequencing and immunodetection, *Electrophoresis 21*, 935-948.

[65] Dubos, R. J., and Middlebrook, G. (1948) Cytochemical reaction of virulent tubercle bacilli, *Am. Rev. Tuberc. 58*, 698.

[66] Netea, M. G., Van der Meer, J. W., and Kullberg, B. J. (2004) Toll-like receptors as an escape mechanism from the host defense, *Trends Microbiol. 12*, 484-488.

[67] Basu, J., Shin, D. M., and Jo, E. K. (2012) Mycobacterial signaling through toll-like receptors, *Front. Cell Infect. Microbiol. 2*, 145.

[68] Yoshida, A., Inagawa, H., Kohchi, C., Nishizawa, T., and Soma, G. (2009) The role of toll-like receptor 2 in survival strategies of *Mycobacterium tuberculosis* in macrophage phagosomes, *Anticancer Res. 29*, 907-910.

[69] Pethe, K., Sequeira, P. C., Agarwalla, S., Rhee, K., Kuhen, K., Phong, W. Y., Patel, V., Beer, D., Walker, J. R., Duraiswamy, J., Jiricek, J., Keller, T. H., Chatterjee, A., Tan, M. P., Ujjini, M., Rao, S. P., Camacho, L., Bifani, P., Mak, P. A., Ma, I., Barnes, S. W., Chen, Z., Plouffe, D., Thayalan, P., Ng, S. H., Au, M., Lee, B. H., Tan, B. H., Ravindran, S., Nanjundappa, M., Lin, X., Goh, A., Lakshminarayana, S. B., Shoen, C., Cynamon, M., Kreiswirth, B., Dartois, V., Peters, E. C., Glynne, R., Brenner, S., and Dick, T. (2010) A chemical genetic screen in *Mycobacterium tuberculosis* identifies carbon-source-dependent growth inhibitors devoid of in vivo efficacy, *Nat. Commun. 1*, 1-8.

[70] Chao, J., Wong, D., Zheng, X., Poirier, V., Bach, H., Hmama, Z., and Av-Gay, Y. (2010) Protein kinase and phosphatase signaling in *Mycobacterium tuberculosis* physiology and pathogenesis, *Biochim. Biophys. Acta 1804*, 620-627.

FIGURES

Figure 1. A simplified hypothetical model highlighting the role of KpbA in the regulation of actinorhodin production in *Streptomyces coelicolor*.[6] The AfsK-binding protein A, KbpA, binds to AfsK, a protein loosely associated with the membrane (**A**). An unknown internal or external effector, labeled with a question mark, modulates the release of KbpA from AfsK **(B)** to allow activation of KpbA by autophosphorylation (**C**). Activated AfsK, as well as other kinases, phosphorylates AfsR which then binds to the *AfsS* promotor to activate AfsS transcription (**D**). In turn the 63-residue protein, AfsS, is an *act*II-ORF4 promotor that turns on the transcription of the ~100 genes involved in actinorhodin biosynthesis. AfsK is expressed during all stages of cell growth[4] while KpbA is actively expressed after actinorhodin production has begun.

Figure 2. A) Crystal structure of nRv0577 (3OXH) with the four approximate βαβββ repeats colored purple, green, red, and grey sequentially starting from the N-terminus. Shown is the face of the first of two saddles composed of two sequential repeats that form

a continuous eight-strand β-sheet (β2:β3:β4:β1:β5:β8:β7:β6) capped on both ends with α1 and α2. The two saddles, connected by an inter-saddle loop between β8 and β9, intersect near the center at roughly a right angle and form a deep cleft on both sides of the protein. **B**) The primary amino acid sequence of Rv0577 with the four approximate repeats shaded in purple, green, red, and grey boxes and the α-helices and β-strands colored red and gold, respectively. The repeats sequentially form two saddles of two repeats each: residues P11 - A129 (saddle 1) and L143 - A258 (saddle 2). Arginine-4 and T136 form a single hydrogen-bond β-bridge. Residues with either missing or unassigned amide resonances in the ^{1}H-^{15}N HSQC spectrum[14] are underlined and residues missing electron density in the crystal structure are colored cyan. The most conserved residues identified by the ConSurf server (http://consurf.tau.ac.il/2016/)[37] are identified with a magenta (9) and pink (8) rectangle above the residue.

Figure 3. Cartoon representation of the crystal structure of nRv0577 (30XH) with the ligands observed associated with the structure shown as spherical representations: xylitol (magenta), p-chloromercuricbenzene sulfonic (PCMBS, cyan), and H-13 – S-10 (-HHHS-, yellow). Residues P70-G72 were not observed in the crystal structure.

Figure 4. Cartoon representation of the deepest part of the saddles on opposite faces of the crystal structure of nRv0577 highlighting the side chains of conserved residues identified by the ConSurf server (http://consurf.tau.ac.il/2016/).[37] Both saddles feature a highly conserved WYD/EY quartet. **A**) Saddle 1 with the bound PCMBS and -HHHS- shown and colored cyan and yellow, respectively. **B**) Saddle 2 with the xylitol observed

in the crystal structure of nRv0577 not shown because it sits just outside the deep part of the saddle below W146.

Figure 5. Summary of the Rv0577c titration studies with deoxyadenosine (Ado) and neutral red (NR) at 8:1 and 4:1 molar ratios, respectively. **NR**) The most significant perturbations to the amide chemical shifts of nRv0577 upon titration with neutral red were their disappearance in ^{1}H-^{15}N HSQC spectra. The chemical shifts of a smaller number of amides also shifted with some reappearing after first disappearing at lower neutral red concentrations. Therefore, the plot for neutral red categorizes amides into three groups: missing, shifted, and reappeared with a change in chemical shift following disappearance. A full line represents an amide cross peak that disappeared, a half-line represents an amide cross peak that shifted, and a half-line with a blue asterisk on top represents an amide cross peak that disappeared in early titration points and re-appeared at a different position at higher neutral red molar ratios. **Ado**) Combined average chemical shift change in the amide $^{1}H^{N}$ and ^{15}N resonances of Rv0577c after the addition of an eight-fold molar excess of Ado. The average chemical shift change = Δ_{ave} (ppm) = $[((\Delta^{1}N^{H})^{2} + (\Delta^{15}N/5)^{2})/2]^{1/2}$. The dashed red line represents a Δ_{ave} of 0.07 ppm. On top of the graphs is a schematic representation of the elements of α-helical (red) and β-strand (yellow) secondary structure observed in the crystal structure of nRv0577 with β-strands 10 and 11 not numbered for clarity.

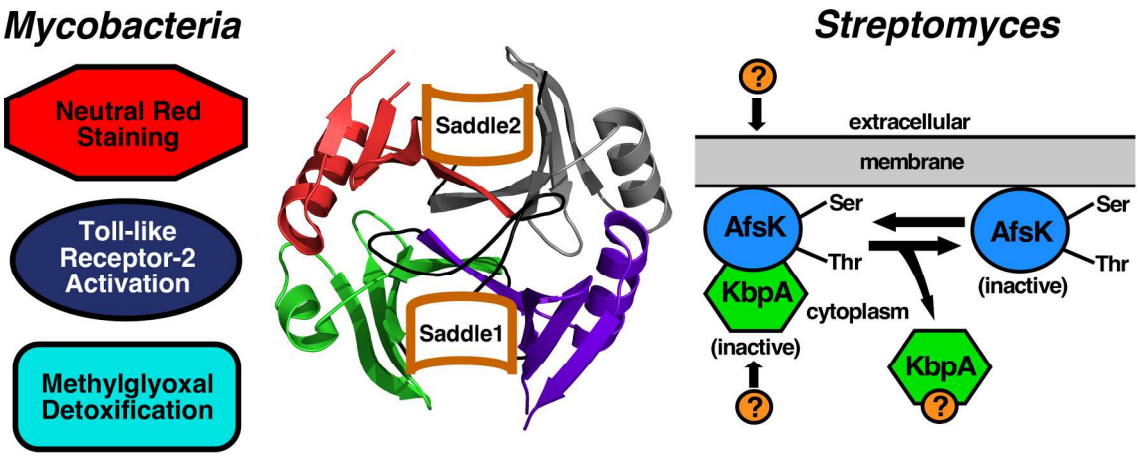
Figure 6. Raw and fit ITC isotherms for Ado titrated into Rv0577c collected at 37 ºC.

Figure 7. Backbone amide (**A**) $\{^1H\}$-^{15}N heteronuclear NOE (purple), (**B**) T_2 (blue), and (**C**) T_1 (red) values for Rv0577c at 30 ºC. On top of the graphs is a schematic representation of the elements of α-helical (red) and β-strand (yellow) secondary structure observed in the crystal structure of nRv0577 with β-strands 10 and 11 not numbered for clarity.

Figure 8. Overlay of ^1H-^{15}N HSQC spectra for Rv0577c (magenta) and nRv0577 (black) collected at a ^1H resonance frequency of 800 MHz, 30 ºC. The amide cross peaks observed in Rv0577c but missing in nRv0577 are circled, labeled, and colored blue (saddle 1), cyan (saddle 2), or orange (inter-saddle linker).

Figure 9. A) Circular dichroism steady-state wavelength spectrum for Rv0577c (0.02 mM) at 25 ºC obtained in buffer containing 100 mM NaCl, 20 mM Tris, 1.0 mM dithiothreitol , pH 7.0. **B**) The CD thermal melt for Rv0577c obtained by measuring the ellipticity at 220 nm in 2.0 ºC intervals between 10 and 80 ºC. **C**) First derivative of the 220 nm thermal melt curve indicates the protein has a melting temperature of 56 ºC.

Table of Contents Graphic



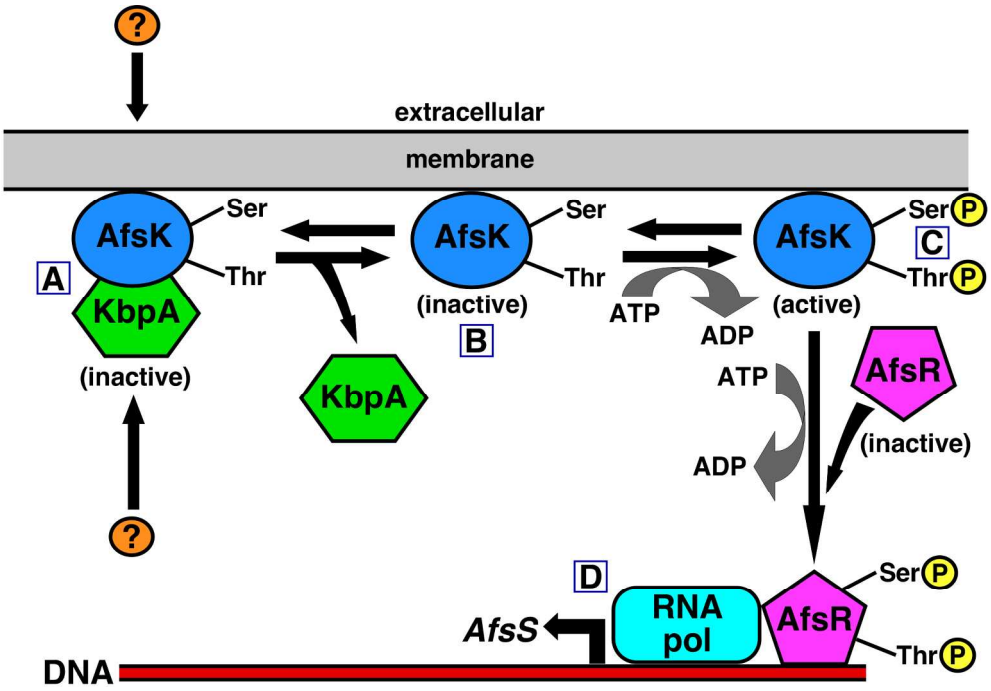


Figure 1. A simplified hypothetical model highlighting the role of KpbA in the regulation of actinorhodin production in *Streptomyces coelicolor*.[6] The AfsK-binding protein A, KbpA, binds to AfsK, a protein loosely associated with the membrane (A). An unknown internal or external effector, labeled with a question mark, modulates the release of KbpA from AfsK (B) to allow activation of KpbA by autophosphorylation (C). Activated AfsK, as well as other kinases, phosphorylates AfsR which then binds to the *AfsS* promotor to activate AfsS transcription (D). In turn the 63-residue protein, AfsS, is an actII-ORF4 promotor that turns on the transcription of the ~100 genes involved in actinorhodin biosynthesis. AfsK is expressed during all stages of cell growth4 while KpbA is actively expressed after actinorhodin production has begun.†

192x130mm (300 x 300 DPI)

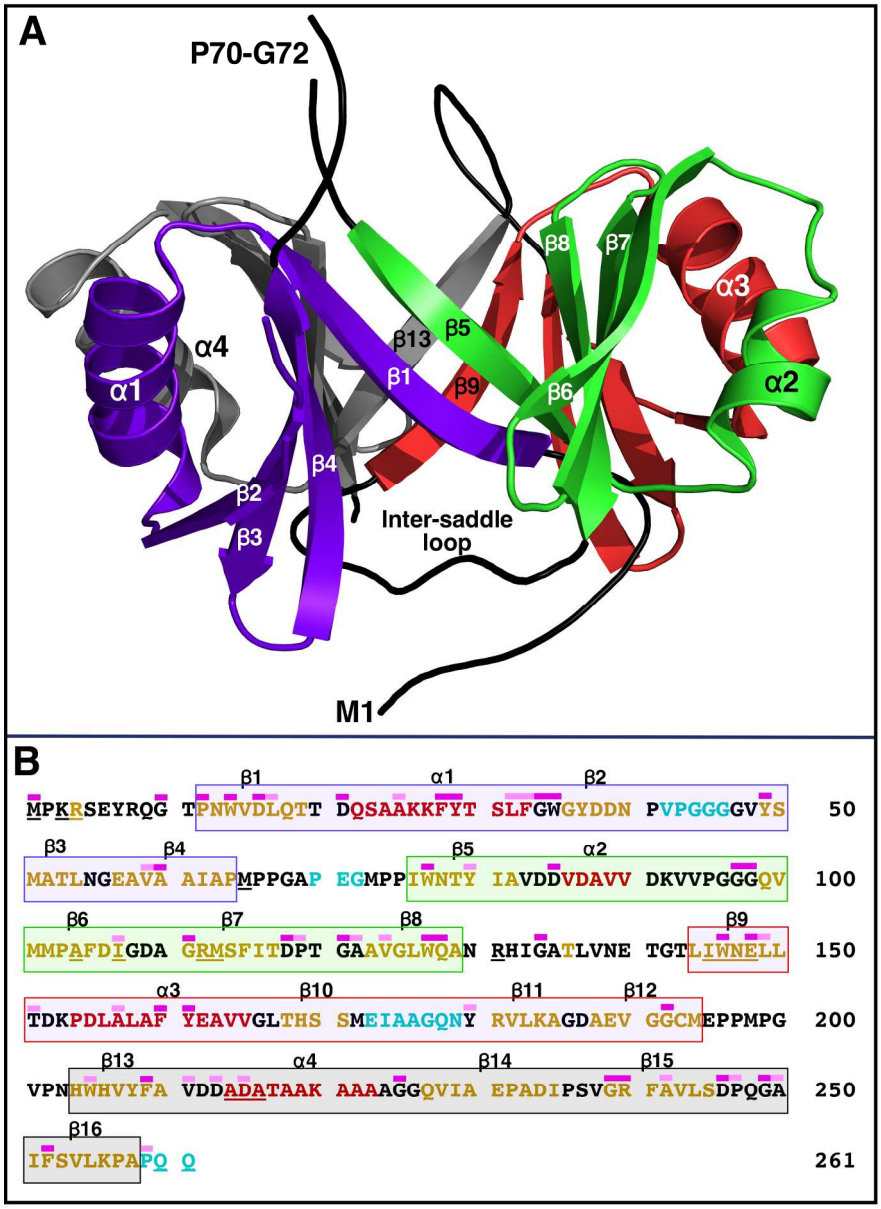


Figure 2. A) Crystal structure of nRv0577 (3OXH) with the four approximate βαββ repeats colored purple, green, red, and grey sequentially starting from the N-terminus. Shown is the face of the first of two saddles composed of two sequential repeats that form a continuous eight-strand β-sheet (β2:β3:β4:β1:β5:β8:β7:β6) capped on both ends with α1 and α2. The two saddles, connected by an inter-saddle loop between β8 and β9, intersect near the center at roughly a right angle and form a deep cleft on both sides of the protein. **B)** The primary amino acid sequence of Rv0577 with the four approximate repeats shaded in purple, green, red, and grey boxes and the α-helices and β-strands colored red and gold, respectively. The repeats sequentially form two saddles of two repeats each: residues P11 - A129 (saddle 1) and L143 - A258 (saddle 2). Arginine-4 and T136 form a single hydrogen-bond β-bridge. Residues with either missing or unassigned amide resonances in the ^{1}H-^{15}N HSQC spectrum[14] are underlined and residues missing electron density in the crystal structure are colored cyan. The most conserved residues identified by the ConSurf server (http://consurf.tau.ac.il/2016/)[37] are identified with a magenta (9) and pink (8) rectangle above the residue.

184x253mm (300 x 300 DPI)

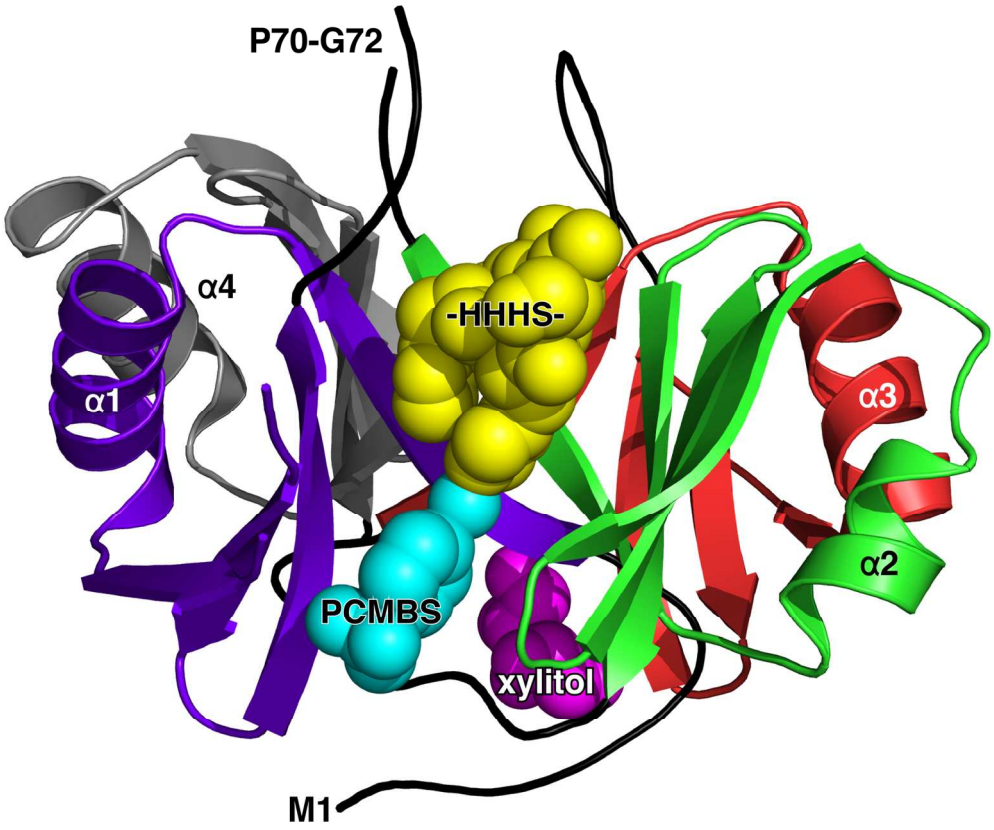


Figure 3. Cartoon representation of the crystal structure of nRv0577 (30XH) with the ligands observed associated with the structure shown as spherical representations: xylitol (magenta), p-chloromercuricbenzene sulfonic (PCMBS, cyan), and H-13 – S-10 (-HHHS-, yellow). Residues P70-G72 were not observed in the crystal structure. ‼ ┼ Residues P70-G72 were not observed in the crystal structure.

153x126mm (300 x 300 DPI)

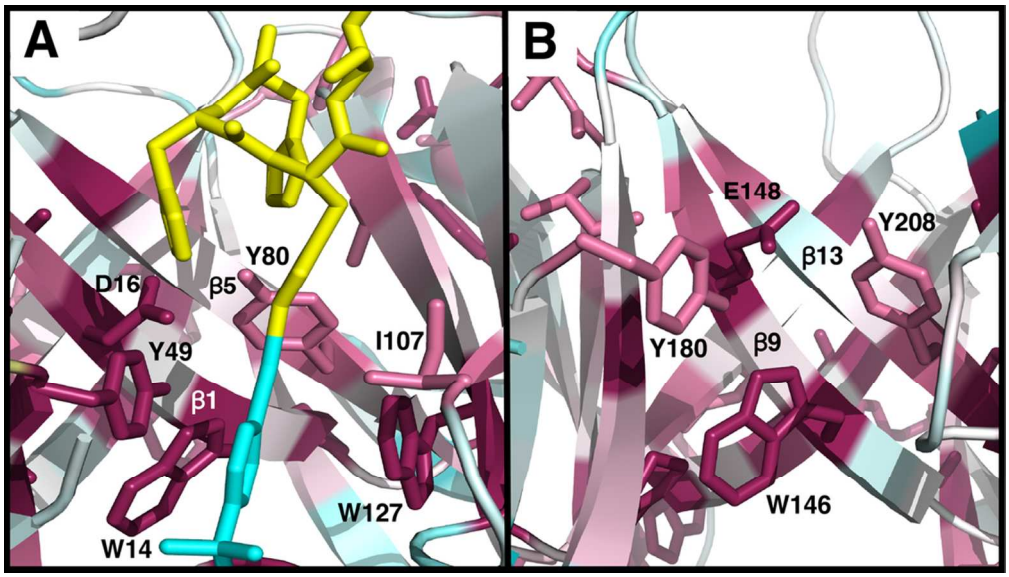


Figure 4. Cartoon representation of the deepest part of the saddles on opposite faces of the crystal structure of nRv0577 highlighting the side chains of conserved residues identified by the ConSurf server (http://consurf.tau.ac.il/2016/).[37] Both saddles feature a highly conserved WYD/EY quartet. A) Saddle 1 with the bound PCMBS and -HHHS- shown and colored cyan and yellow, respectively. B) Saddle 2 with the xylitol observed in the crystal structure of nRv0577 not shown because it sits just outside the deep part of the saddle below W146.

95x54mm (300 x 300 DPI)

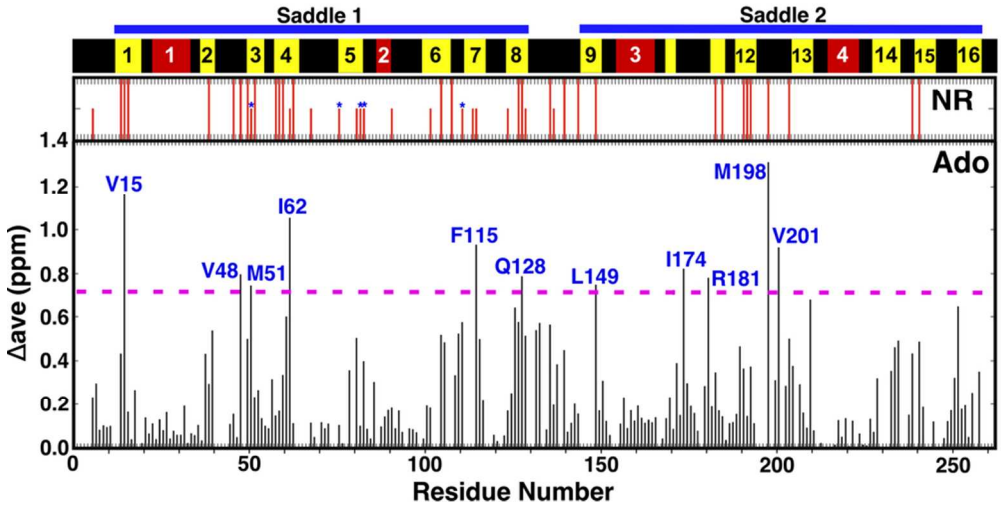


Figure 5. Summary of the Rv0577c titration studies with deoxyadenosine (Ado) and neutral red (NR) at 8:1 and 4:1 molar ratios, respectively. NR) The most significant perturbations to the amide chemical shifts of nRv0577 upon titration with neutral red were their disappearance in ^{1}H-^{15}N HSQC spectra. The chemical shifts of a smaller number of amides also shifted with some reappearing after first disappearing at lower neutral red concentrations. Therefore, the plot for neutral red categorizes amides into three groups: missing, shifted, and reappeared with a change in chemical shift following disappearance. A full line represents an amide cross peak that disappeared, a half-line represents an amide cross peak that shifted, and a half-line with a blue asterisk on top represents an amide cross peak that disappeared in early titration points and re-appeared at a different position at higher neutral red molar ratios. Ado) Combined average chemical shift change in the amide $^{1}H^{N}$ and ^{15}N resonances of Rv0577c after the addition of an eight-fold molar excess of Ado. The average chemical shift change = Δ_{ave} (ppm) = $[((\Delta^{1}N^{N})^{2} + (\Delta^{15}N/5)^{2})/2]^{1/2}$. The dashed red line represents a Δ_{ave} of 0.07 ppm. On top of the graphs is a schematic representation of the elements of α-helical (red) and β-strand (yellow) secondary structure observed in the crystal structure of nRv0577 with β-strands 10 and 11 not numbered for clarity. ‼ ┼

84x43mm (300 x 300 DPI)

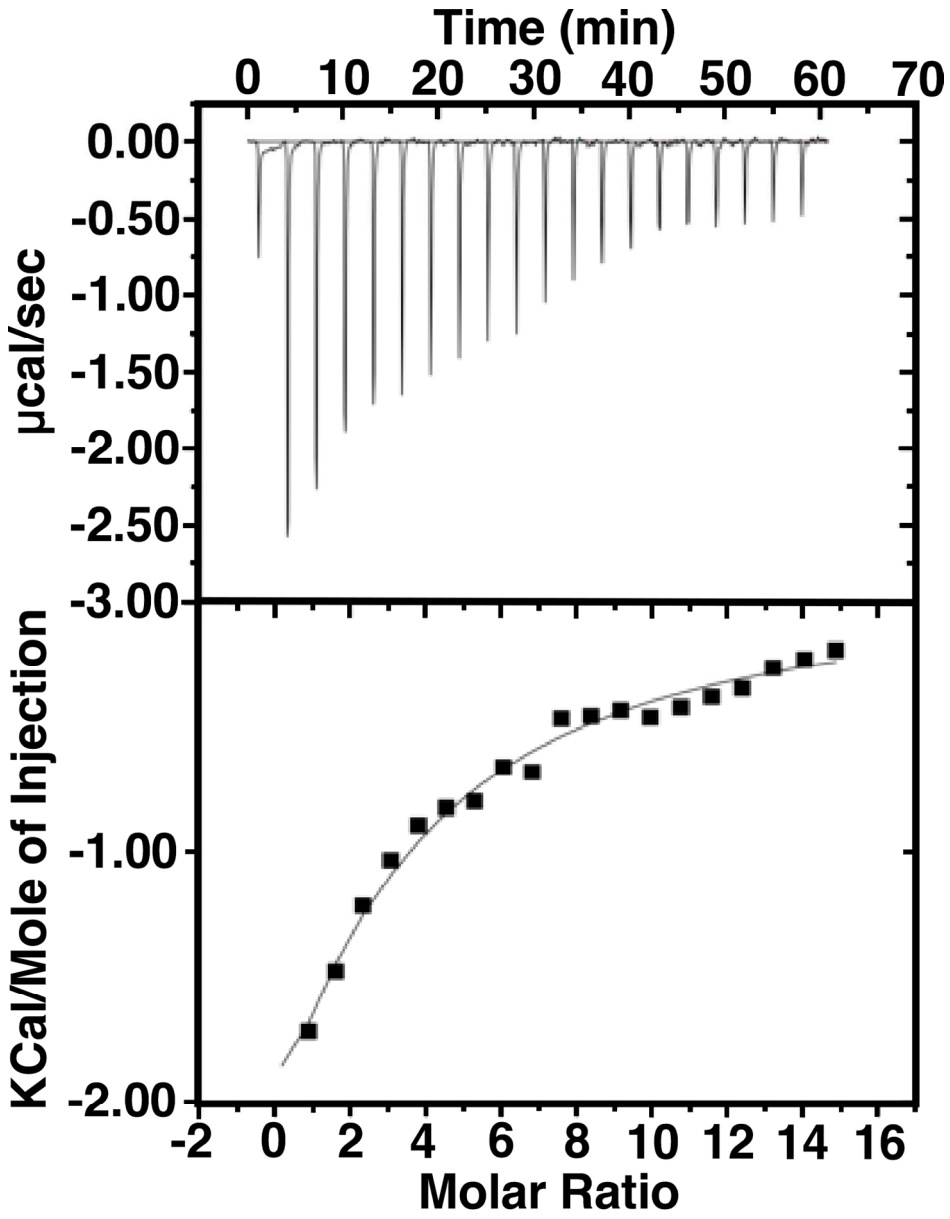


Figure 6. Raw and fit ITC isotherms for Ado titrated into Rv0577c collected at 37 ºC.

135x174mm (300 x 300 DPI)

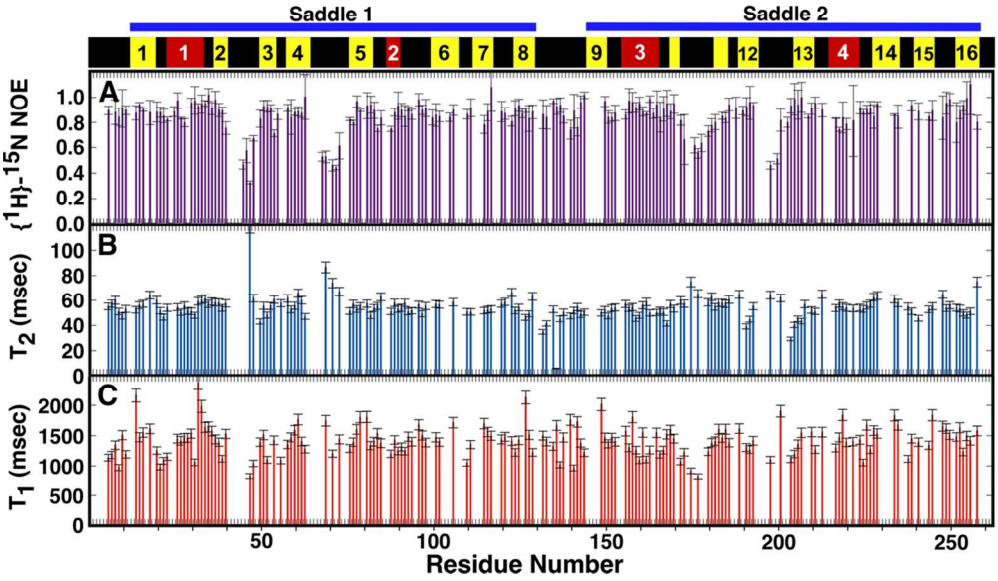


Figure 7. Backbone amide (A) $\{^1H\}$-^{15}N heteronuclear NOE (purple), (B) T_2 (blue), and (C) T_1 (red) values for Rv0577c at 30 ºC. On top of the graphs is a schematic representation of the elements of α-helical (red) and β-strand (yellow) secondary structure observed in the crystal structure of nRv0577 with β-strands 10 and 11 not numbered for clarity.

109x71mm (300 x 300 DPI)

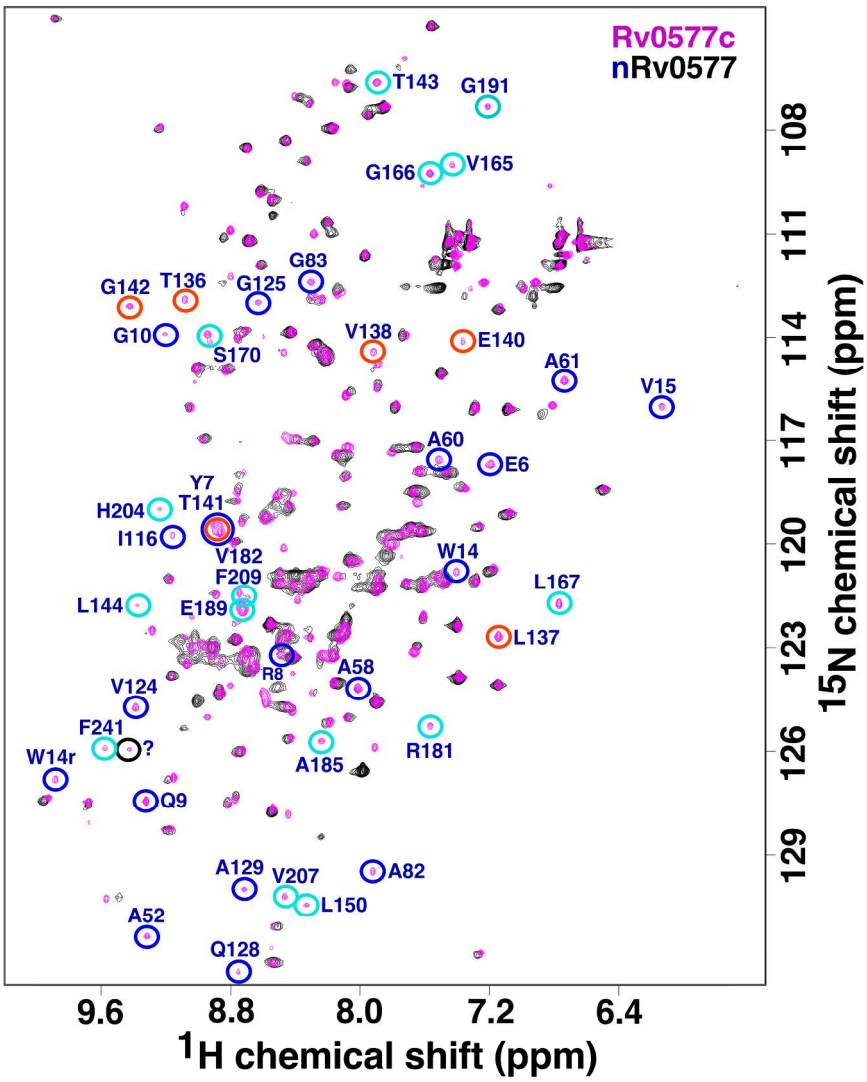


Figure 8. Overlay of ^{1}H-^{15}N HSQC spectra for Rv0577c (magenta) and nRv0577 (black) collected at a ^{1}H resonance frequency of 800 MHz, 30 ºC. The amide cross peaks observed in Rv0577c but missing in nRv0577 are circled, labeled, and colored blue (saddle 1), cyan (saddle 2), or orange (inter-saddle linker).

246x310mm (300 x 300 DPI)

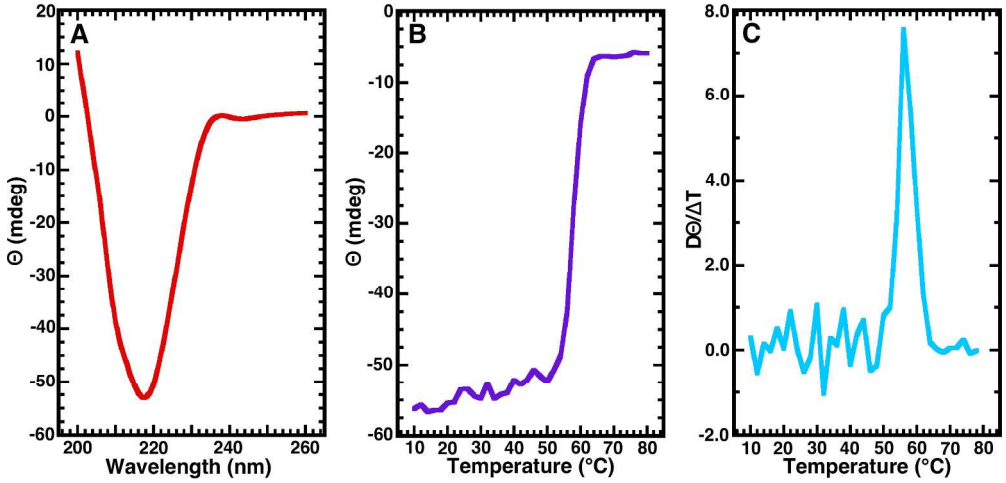
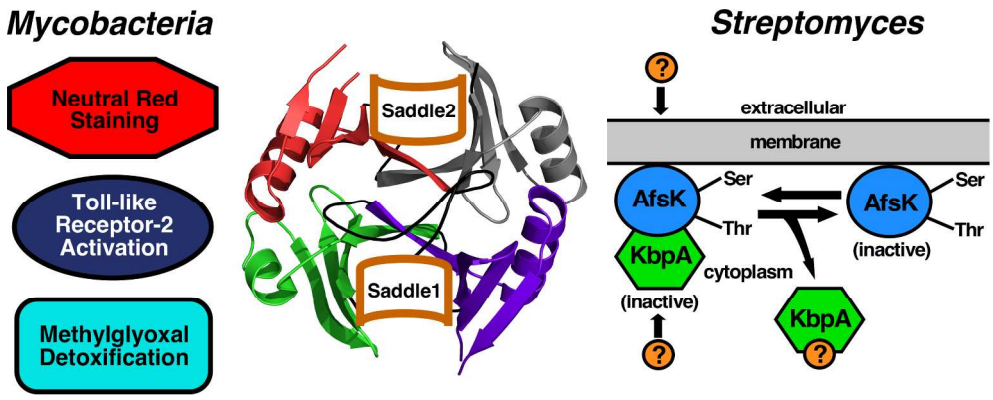


Figure 9. A) Circular dichroism steady-state wavelength spectrum for Rv0577c (0.02 mM) at 25 ºC obtained in buffer containing 100 mM NaCl, 20 mM Tris, 1.0 mM dithiothreitol , pH 7.0. **B)** The CD thermal melt for Rv0577c obtained by measuring the ellipticity at 220 nm in 2.0 ºC intervals between 10 and 80 ºC. **C)** First derivative of the 220 nm thermal melt curve indicates the protein has a melting temperature of 56 ºC.

239x113mm (300 x 300 DPI)



216x83mm (300 x 300 DPI)