Supplementary Material

Supplementary Table I
Supplementary References
Supplementary Figure 1
Supplementary Figure 2

Molecular insights into the binding of coenzyme F_{420} to the conserved protein Rv1155 from *Mycobacterium tuberculosis*

Ellene H. Mashalidis[1,2], Apostolos G. Gittis[3], Aurelie Tomczak[4], Chris Abell[2], Clifton E. Barry III[1] & David N. Garboczi[3]*

[1]Tuberculosis Research Section, National Institute of Allergy and Infectious Diseases, 9000 Rockville Pike, Bethesda, MD, USA 20892

[2]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW

[3]Structural Biology Section, Research Technologies Branch, National Institute of Allergy and Infectious Diseases, Rockville, MD, USA 20852

[4]Molecular Signaling Section, Laboratory of Molecular Immunology, National Institute of Allergy and Infectious Diseases, Rockville, MD, USA 20852

Running title: *Rv1155 is a novel F_{420}-binding protein in M. tuberculosis*

*Correspondence to: David N. Garboczi, Structural Biology Section/RTB, NIAID, Twinbrook 2 Rm 110, 12441 Parklawn Drive, Rockville, MD, USA 20852. Telephone: 301-496-4773. Fax: 301-402-0284. E-mail: dgarboczi@niaid.nih.gov.

Supplementary Table I. Genes and proteins in *Mtb* H37Rv that are predicted to be associated functionally with Rv1155 based on co-occurrence. The H37Rv *Mtb* genes listed have orthologs that co-occur with Rv1155 orthologs in multiple genomes. A confidence score (CS) based on co-occurrence was calculated using the STRING v9.1 server[1] (http://string-db.org/). A CS of 0.7 or higher indicates a high confidence association. The cell location data were sourced from the TubercuList database[2] (http://tuberculist.epfl.ch/index.html). Signal peptides, transmembrane regions, and amphipathic membrane anchors were predicted respectively with SignalP[3], TMHMM[4], and AmphipaSeeK[5]. Of the 1133 genomes that were examined with the STRING v9.1 server, 63 genomes contained at least one ortholog of Rv1155 and also contained at least six of the proteins listed here.

Protein/ gene	**CS**	**Location in cell**	**Description of protein**
Rv2061c	0.753	membrane	PNPOx-like unknown protein; putative F_{420}-binding protein
Rv0495c	0.751	not known	Conserved hypothetical protein of unknown function
Rv3916c	0.748	membrane	Conserved hypothetical protein of unknown function
Rv3547	0.746	membrane	Ddn, deazaflavin-dependent nitroreductase
Rv1558	0.744	membrane	Ddn-like putative nitroreductase
Rv3718c	0.727	membrane	Conserved hypothetical protein; contains a polyketide cyclase/dehydrase and lipid transport domain (Polyketide_cyc2 / PF10604) often involved in binding of lipid molecules[15,49]
Rv2694c	0.724	membrane	Conserved hypothetical protein of unknown function
Rv1727	0.709	whole cell lysate	Conserved hypothetical protein of unknown function
Rv2074	0.65	membrane	PNPOx-like unknown protein; putative F_{420}-binding protein
Rv0260c	0.71	cell wall	Possible transcription regulatory protein

The 63 genomes were from species in genera that are listed below. Shown in parentheses are the number of genomes (species) in each genus that contain at least six of the proteins listed in the table and an ortholog of Rv1155: *Mycobacterium* (21), *Micrococcineae* (17), *Frankineae* (5), *Nocardiaceae* (4), *Pseudonocardineae* (3), *Streptosporanineae* (3), *Streptomyces* (3), *Propionibacterineae* (2), *Salinispora* (2), *Catenulispora* (1), *Gordonia* (1), and *Kineococcus* (1).

Supplementary References

1. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41:D808–15.

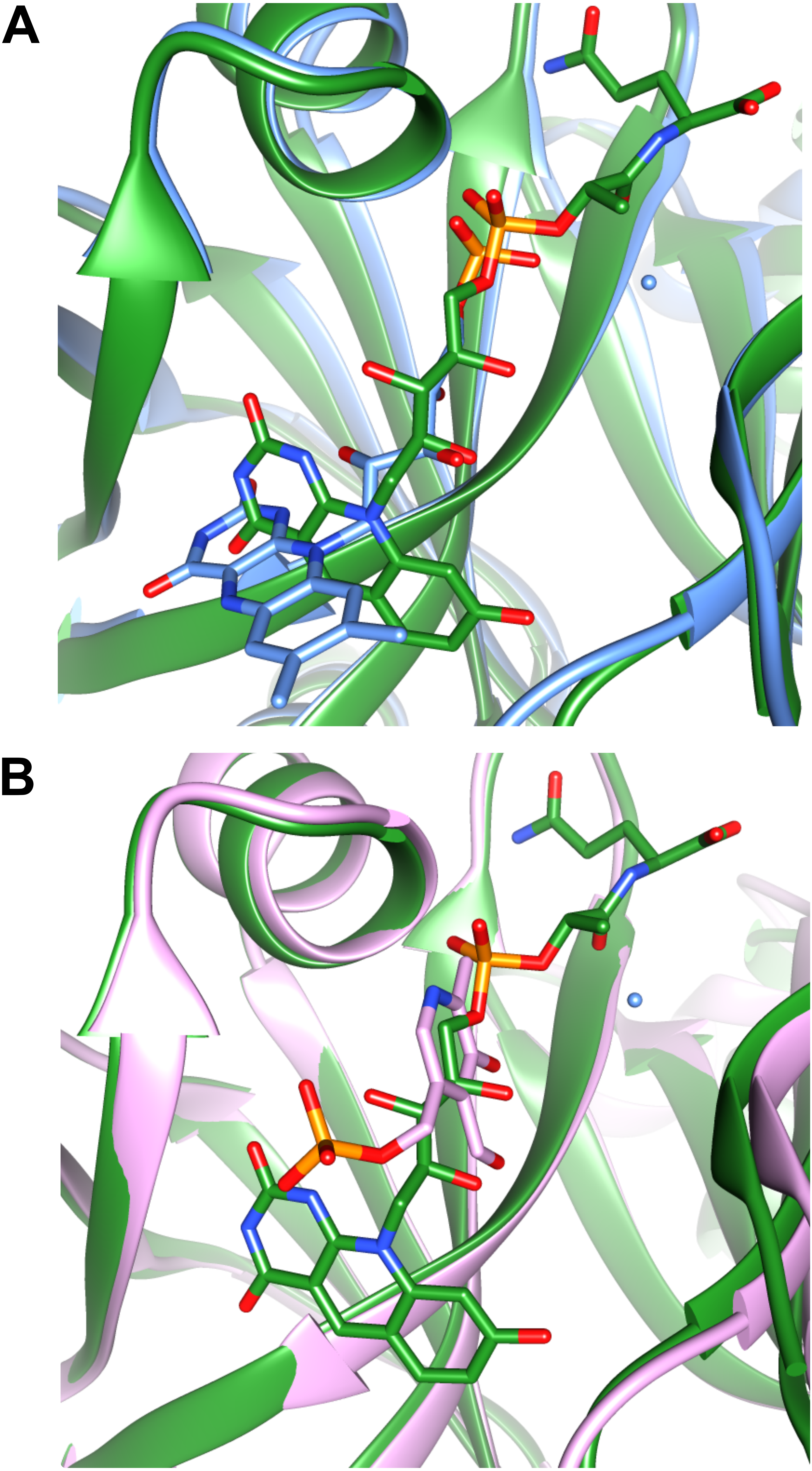
2. Lew JM, Kapopoulou A, Jones LM, Cole ST (2011) TubercuList--10 years after. Tuberculosis (Edinb) 91:1–7.

3. Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8(10):785–786.

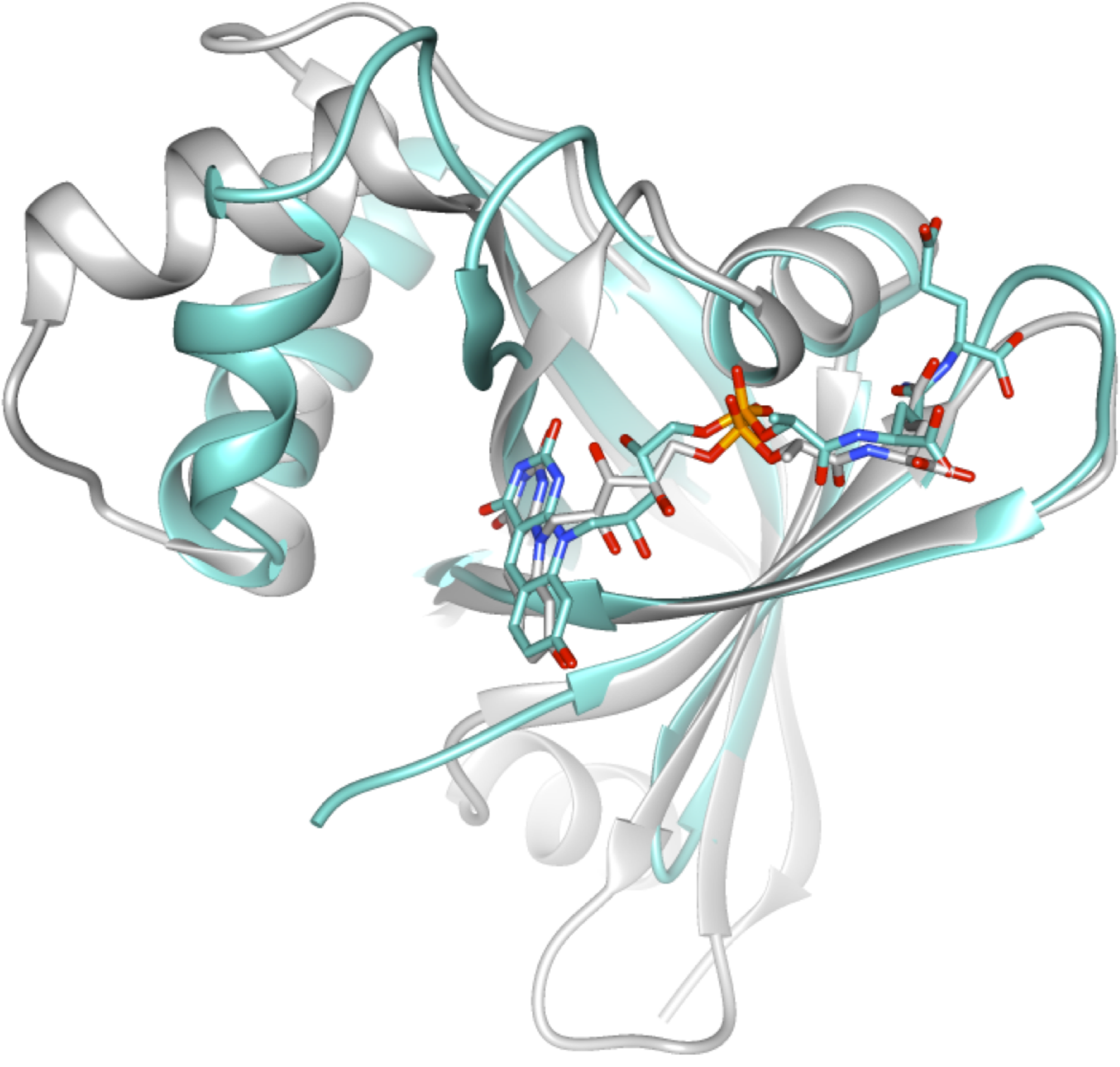
4. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305:567–580.

5. Sapay N, Guermeur Y, Deleage G (2006) Prediction of amphipathic in-plane membrane anchors in monotopic proteins using a SVM classifier. BMC Bioinformatics 7:255.

6. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.


A
B

Supplementary Figure 1. Structural comparison of Rv1155-F_{420}, Rv1155-FMN, and Rv1155-PLP. (A) Structural superimposition of Rv1155-F_{420} (green) and Rv1155-FMN (blue). RMSD = 0.57; 130 paired atoms. (B) Structural superimposition of Rv1155-F_{420} (green) and Rv1155-PLP (pink). RMSD = 0.58; 140 paired atoms. Images were generated using Chimera.



Supplementary Figure 2. Structural superimposition of Ddn-F_{420} and Rv1155-F_{420} monomers. The alignment of Rv1155-F_{420} (grey) and Ddn-F_{420} (light green) has an overall RMSD of 2.1 Å (C-alpha carbons for 87 residue pairs aligned; Rv1155: 140 residues, Ddn: 109 residues). The structures were aligned using Chimera[6].