Carbohydrate Recognition by RpfB from *Mycobacterium tuberculosis* Unveiled by Crystallographic and Molecular Dynamics Analyses

Flavia Squeglia,[†‡] Maria Romano,[†§] Alessia Ruggiero,[†] Luigi Vitagliano,[†] Alfonso De Simone,[¶]* and Rita Berisio[†]*

[†]Institute of Biostructures and Bioimaging, C.N.R., Naples, Italy; [‡]Department of Chemistry, University of Naples Federico II, Napoli, Italy; [§]Seconda Università di Napoli, Caserta, Italy; and [¶]Division of Molecular Biosciences, Imperial College London, United Kingdom

Submitted February 21, 2013, and accepted for publication April 16, 2013.

*Correspondence: a.de-simon@imperial.ac.uk or rita.berisio@cnr.it

Supplementary Information

Supplementary Figure Legends

Figure S1. Localisation of strongest water sites in the MD simulation. (A) RpfBcat is represented in cartoon and water sites as balls. The water site W1 is indicated by an arrow. (B) RpfBcat is represented in surface representation. Water sites are defined as totally buried (TB), nearly totally buried (NTB) and solvent exposed (SE)

Figure S2. Flexibility of the complex between $RpfB_{cat}$ and NAG6. (A) RMSF values calculated for the backbone atoms of free (black) and NAG6-bound (red) RpfBcat. (B) Stick representation of NAG3 (orange, from x-ray structure) and NAG6 (light blue, from MD) after superposition of RpfBcat. (C) RMSF values computed on C1 atoms of NAG moieties in the subsites from -4 to +2. (D) φ-ψ distribution of glycosidic bonds between subsequent NAG moieties in the NAG6 molecule. φ and ψ dihedral angles were defined by atoms O5-C1-O1-C4 and C1-O1-C4-C5, respectively.

Figure S3.
Distributions of RMSD values from the starting structure, calculated on all atoms of NAG rings at positions from -4 to +2 (A to F, respectively).

Figure S4. (A) Domain organization and (B) sequence alignement. of the five Rpf homologs. The catalytic Glu292 (RpfB numbering) is reported in red, whereas conserved Gly residues are reported in blue.

Figure S5. (A) Consurf sequence alignment showing the conserved residues on the surface of RpfBcat. Residue coloring, reflecting the degree of residue conservation over the entire domain family, ranges from magenta (highly conserved) to cyan (variable). The identified homologues (using an E-value cutoff of 0.001, See Methods) include mainly Resuscitation Promoting Factors from different sources and transglycosydases. Key residues involved in hydrogen bonding with NAG6 in -3;-2 and -1 sites are reported in green, blue and red, respectively. (B) Degree of residue conservation mapped on RpfBcat surface.

Figure S6. W1 is partially covered by NAG3 in the crystal state (A) and becomes totally buried in the simulated complex between RpfBcat and NAG6 (B).

Figure S7. (A) Stick representation of the modeled muropeptide ($NAG_3NAM_3Ala_6Glu_3DAP_3$) mapped on $RpfB_{cat}$ surface. (B) Top view of the complex.

Figure S8. Electrostatic potential surfaces of the five Rpfs (A-E). The sixth panel reports the computed pI values for the five domains and for the entire proteins.

Supplementary Figures

Figure S1

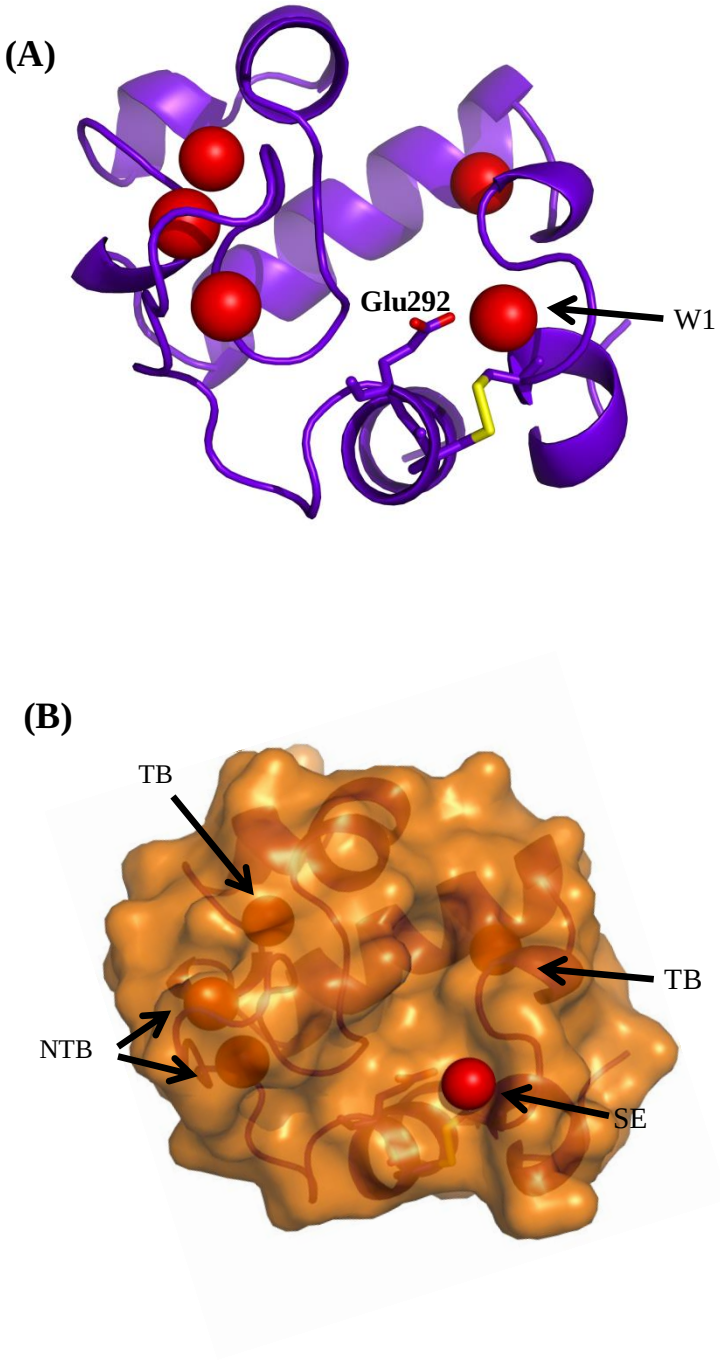
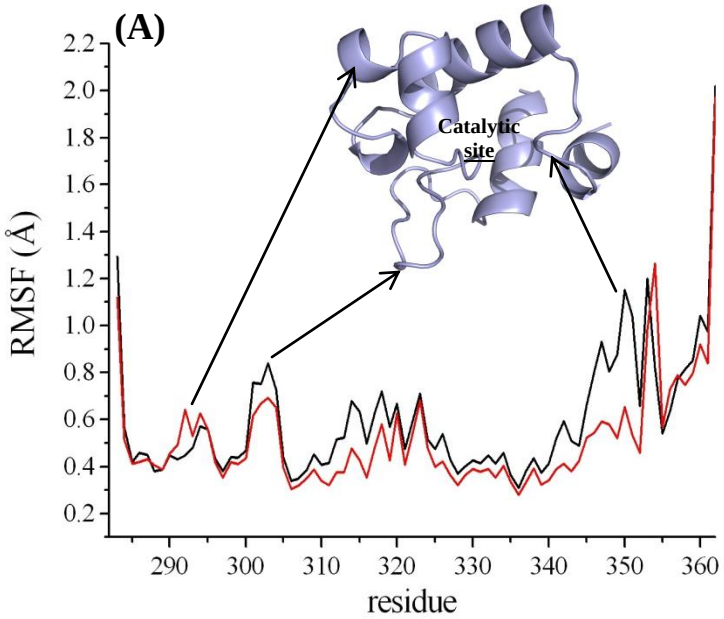
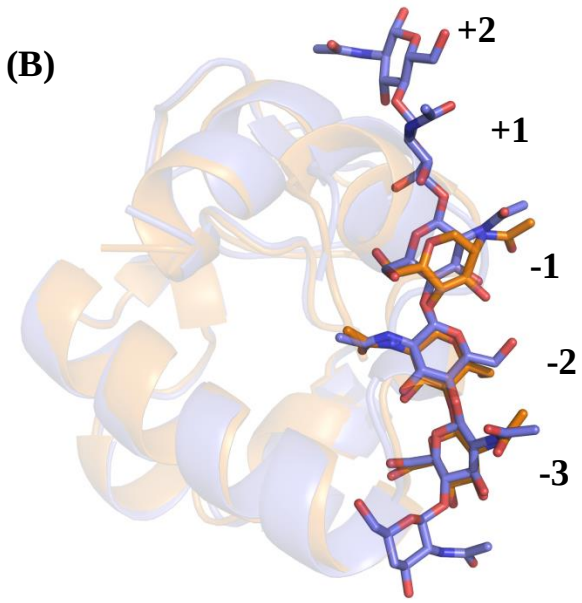
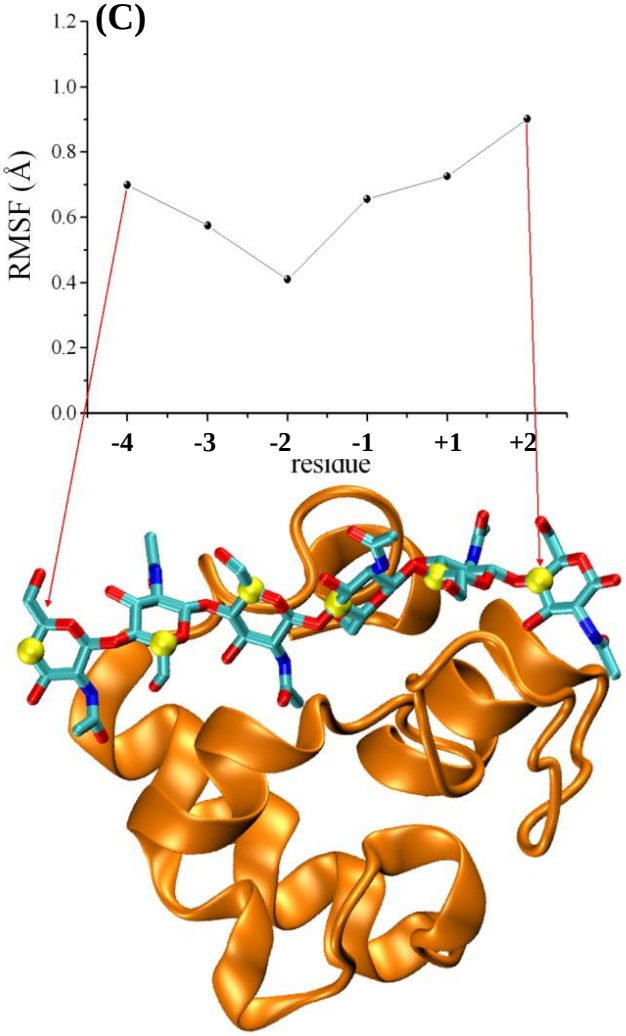


Figure S2


(A)
Catalytic site
RMSF (Å)
residue


(B)
+2
+1
-1
-2
-3


(C)
RMSF (Å)
-4
-3
-2
-1
+1
+2
residue

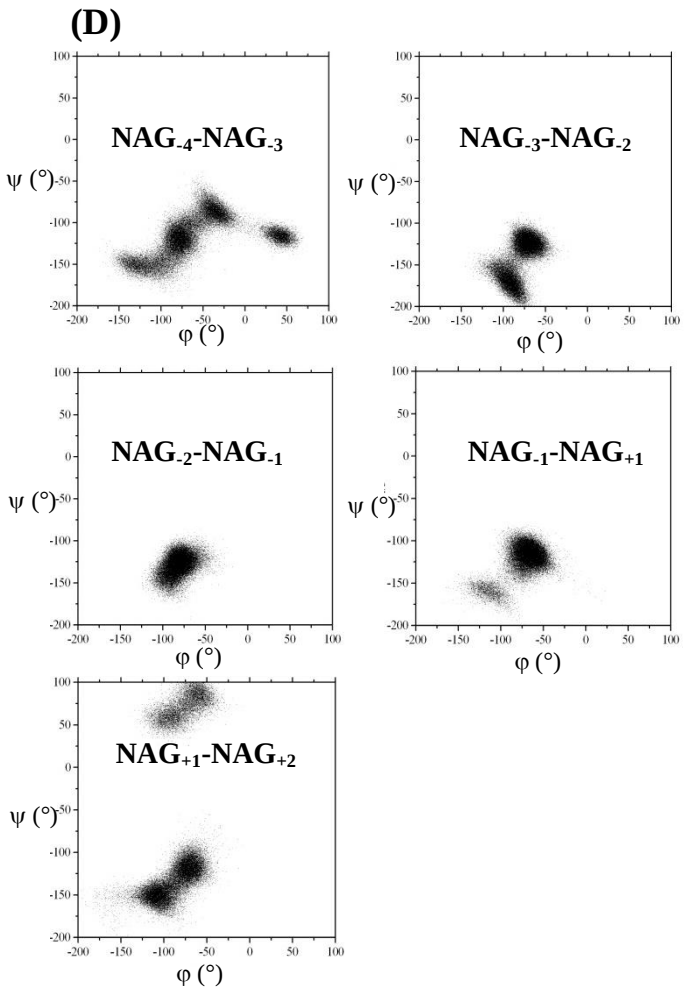

(D)
NAG-4-NAG-3
NAG-3-NAG-2
NAG-2-NAG-1
NAG-1-NAG+1
NAG+1-NAG+2
ψ (°)
φ (°)

Figure S3

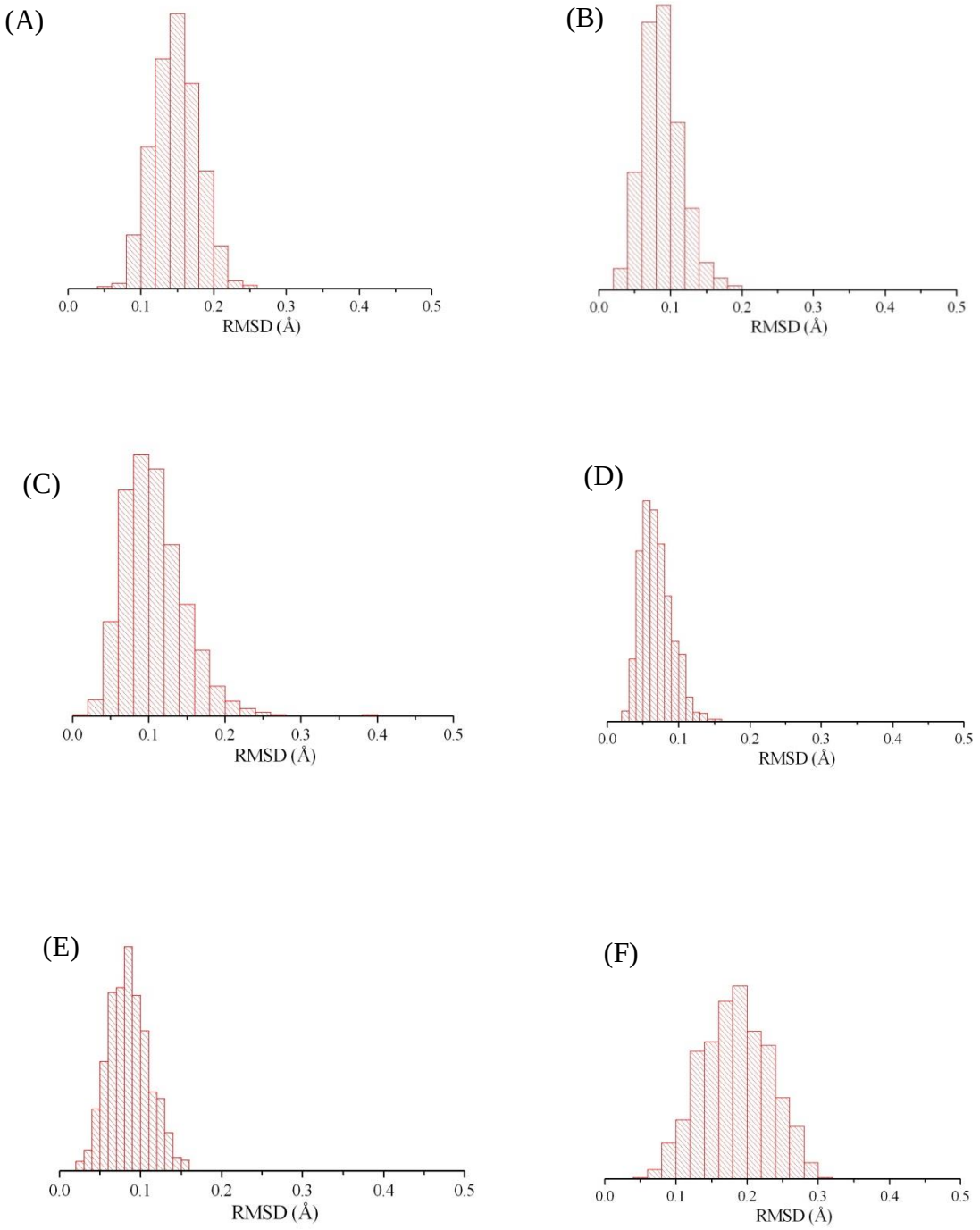
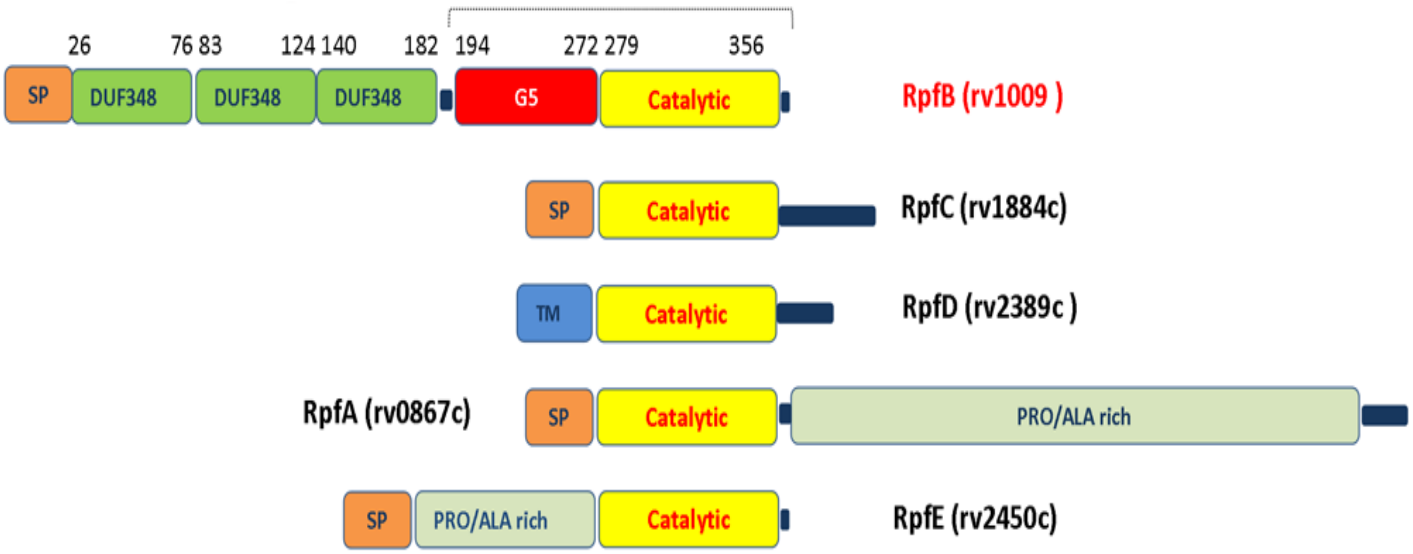

(A)
0.0 0.1 0.2 0.3 0.4 0.5
RMSD (Å)
(B)
0.0 0.1 0.2 0.3 0.4 0.5
RMSD (Å)
(C)
0.0 0.1 0.2 0.3 0.4 0.5
RMSD (Å)
(D)
0.0 0.1 0.2 0.3 0.4 0.5
RMSD (Å)
(E)
0.0 0.1 0.2 0.3 0.4 0.5
RMSD (Å)
(F)
0.0 0.1 0.2 0.3 0.4 0.5
RMSD (Å)

Figure S4

(A)



(B)

	Sequences of catalytic domains
RpfA **RpfB** **RpfC** **RpfD** **RpfE**	GEWDQVARCESGGNWSINTGNGYLGGLQFTQSTWAAHGGGEFAPSAQLASREQQIAVGERVLATQGRGAWPVCGR-GL SIWDAIAGCEAGGNWAINTGNGYYGGVQFDQGTWEANGGLRYAPRADLATREEQIAVAEVTRLRQGWGAWPVCAA-RA PNWDAVAQCESGGNWAANTGNGKYGGLQFKPATWAAFGGVGN---PAAASREQQIAVANRVLAEQGLDAWPTCGAA-S IDWDAIAQCESGGNWAANTGNGLYGGLQISQATWDSNGGVGS---PAAASPQQQIEVADNIMKTQGPGAWPKCSSCSQ VNWDAIAQCESGGNWSINTGNGYYGGLRFTAGTWRANGGSGS---AANASREEQIRVAENVLRSQGIRAWPVCGR-RG 292 302 312 322 332 342 352

Figure S5

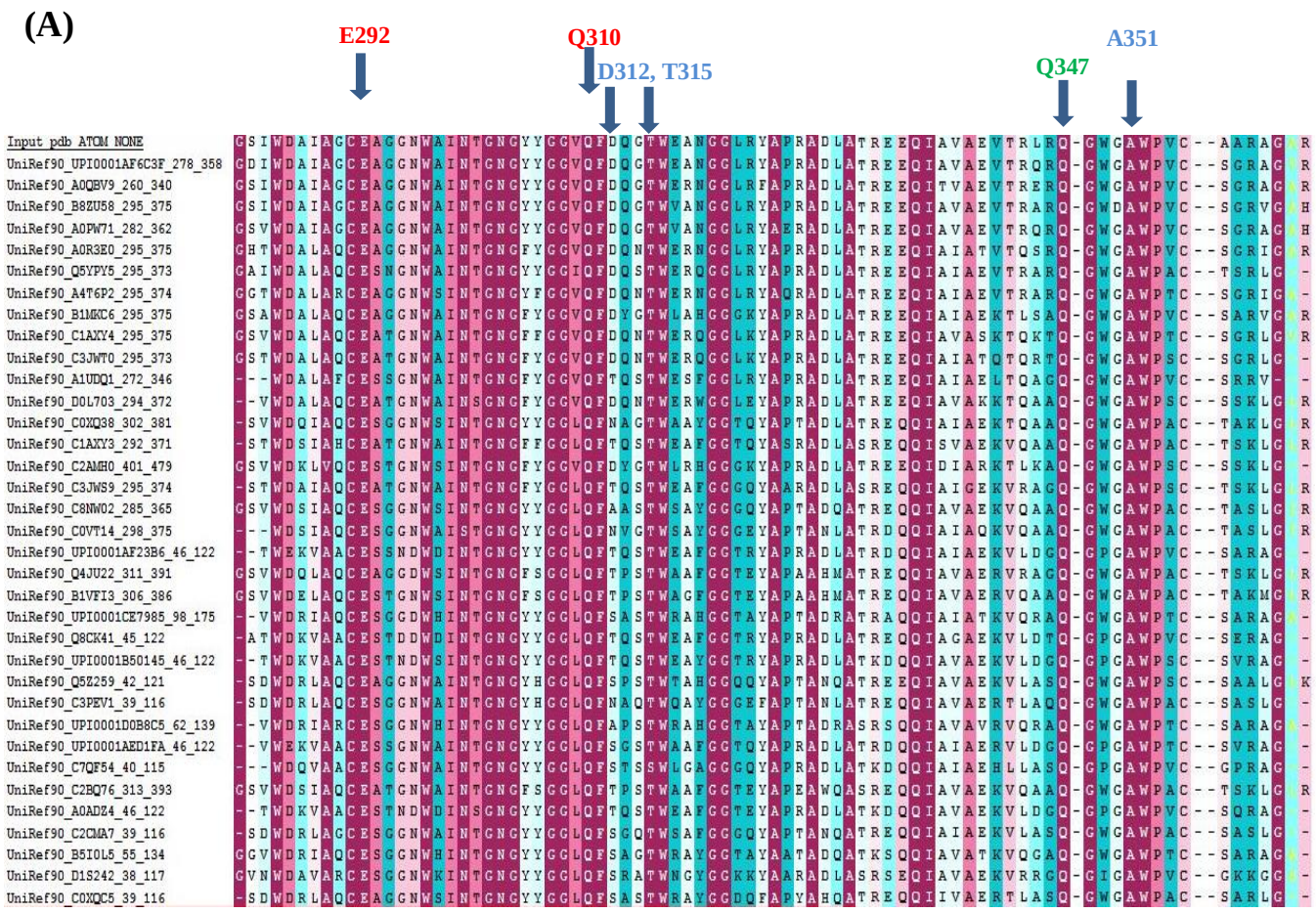
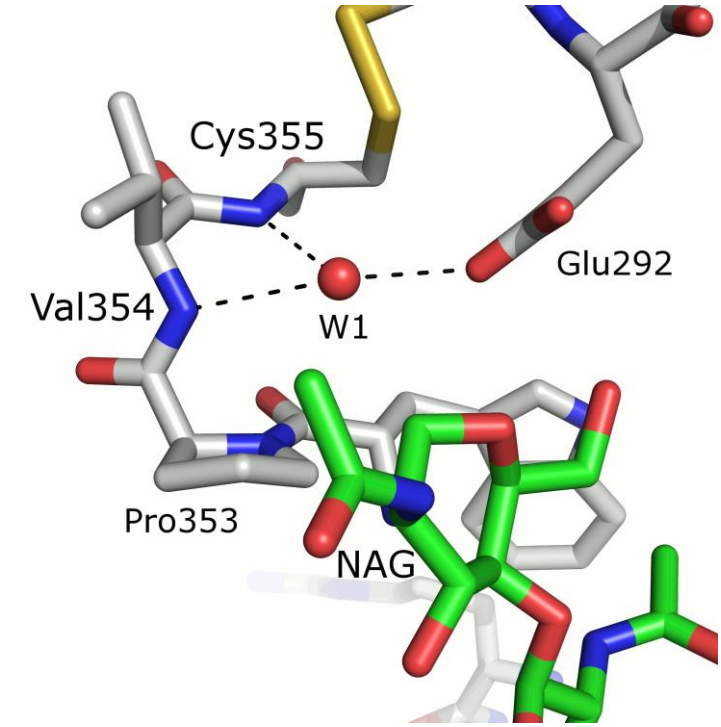

(A)
E292
Q310
D312, T315
Q347
A351
Input pdb ATOM NONE
UniRef90_UPI0001AF6C3F_278_358
UniRef90_A0QBV9_260_340
UniRef90_B8ZU58_295_375
UniRef90_A0PW71_282_362
UniRef90_A0R3E0_295_375
UniRef90_Q5YPY5_295_373
UniRef90_A4T6P2_295_374
UniRef90_B1MKC6_295_375
UniRef90_C1AXY4_295_375
UniRef90_C3JWT0_295_373
UniRef90_A1UDQ1_272_346
UniRef90_D0L703_294_372
UniRef90_C0XQ38_302_381
UniRef90_C1AXY3_292_371
UniRef90_C2AMH0_401_479
UniRef90_C3JWS9_295_374
UniRef90_C8NW02_285_365
UniRef90_C0VT14_298_375
UniRef90_UPI0001AF23B6_46_122
UniRef90_Q4JU22_311_391
UniRef90_B1VFI3_306_386
UniRef90_UPI0001CE7985_98_175
UniRef90_Q8CK41_45_122
UniRef90_UPI0001B50145_46_122
UniRef90_Q5Z259_42_121
UniRef90_C3PEV1_39_116
UniRef90_UPI0001D0B8C5_62_139
UniRef90_UPI0001AED1FA_46_122
UniRef90_C7QF54_40_115
UniRef90_C2BQ76_313_393
UniRef90_A0ADZ4_46_122
UniRef90_C2CMA7_39_116
UniRef90_B5I0L5_55_134
UniRef90_D1S242_38_117
UniRef90_C0XQC5_39_116
(B)
+2
+1
-1
-2
-3
-4

Figure S6

(A)



(B)

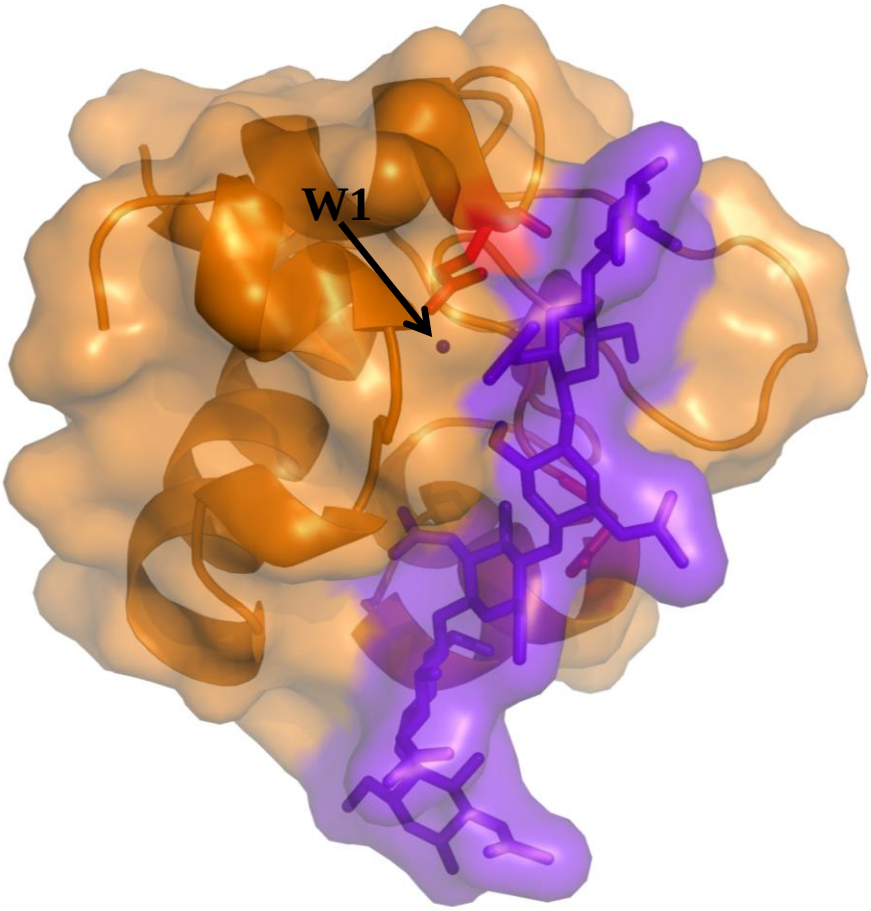


Figure S7

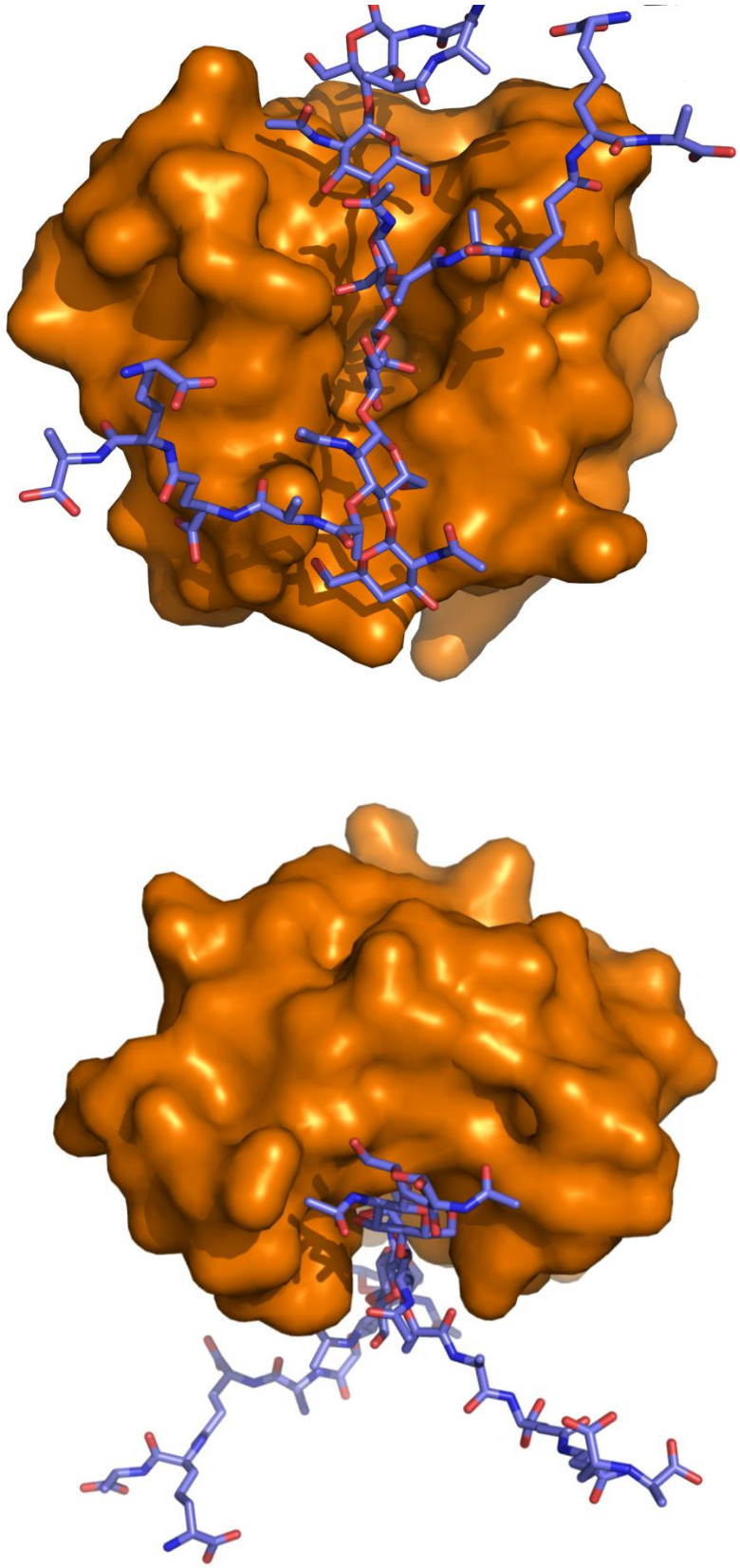
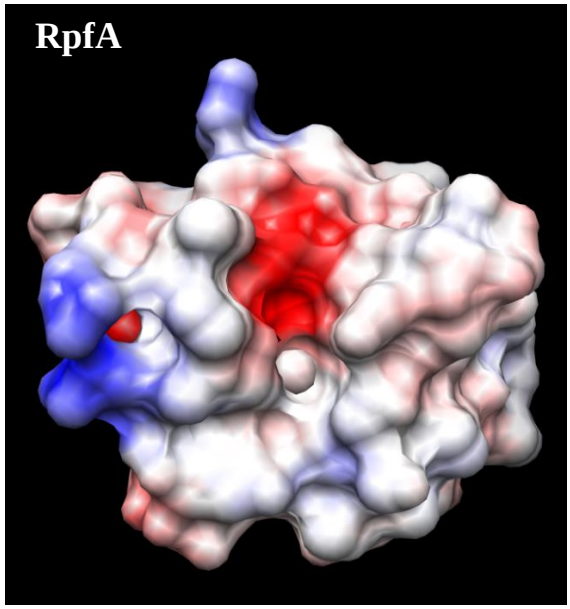
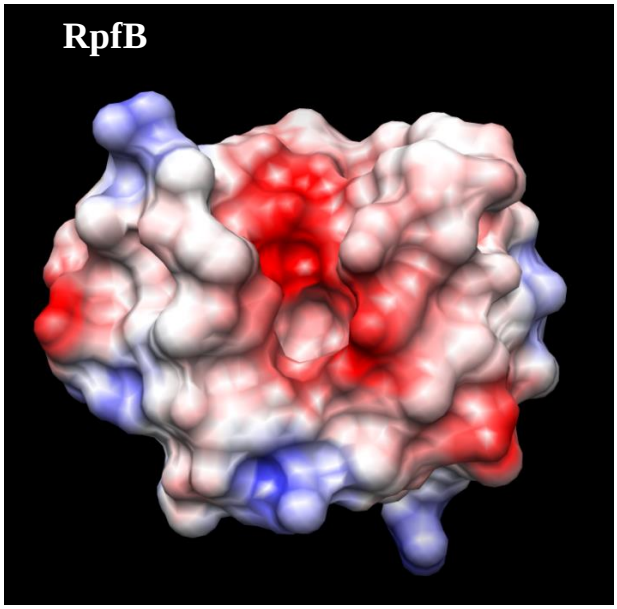
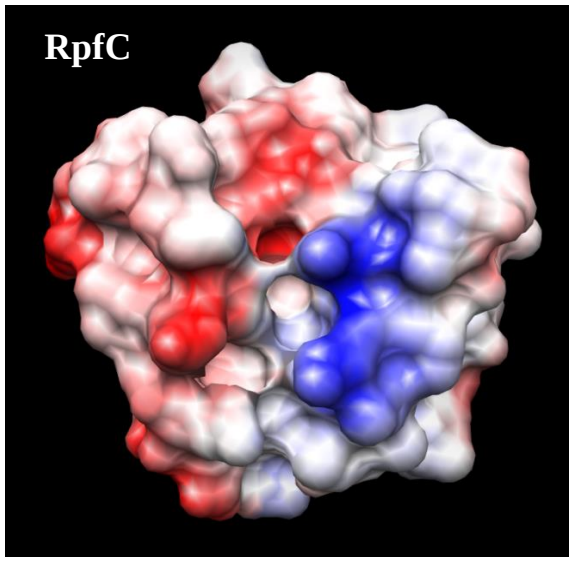
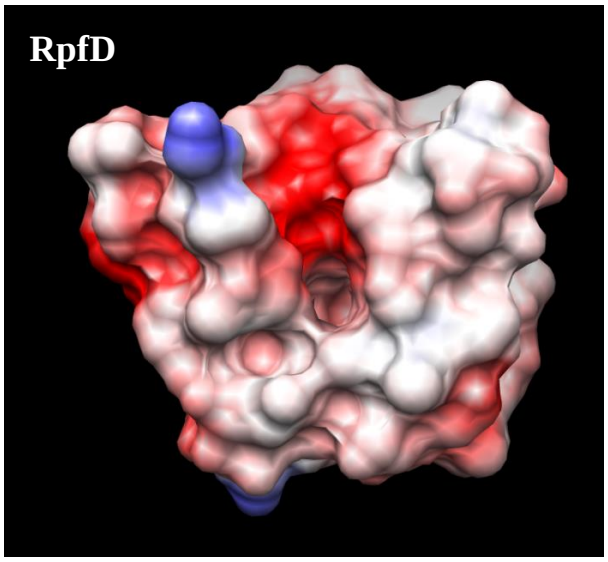


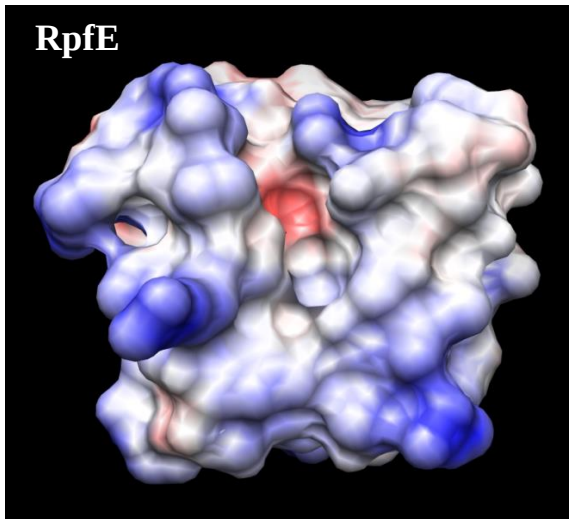
Figure S8











	Theoretical pI	Catalytic domain	Whole protein
RpfA	5.6	3.87	
RpfB	4.8	5.4	
RpfC	4.9	9.5	
RpfD	4.0	4.4	
RpfE	9.5	4.3	