Applied & Translational Genomics 1 (2012) 9–20

Contents lists available at SciVerse ScienceDirect

Applied & Translational Genomics

journal homepage: www.elsevier.com/locate/atg

Co-evolution of RNA polymerase with RbpA in the phylum Actinobacteria

Abhinav Dey [a,*], V.R. Adithi [b], Dipankar Chatterji [a]

[a] Molecular Biophysics Unit, Indian Institute of Science, Bangalore-560012, India
[b] Department of Plant Molecular Biology and Biotechnology, Center for Plant Molecular Biology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

ARTICLE INFO

Article history:
Received 28 December 2011
Received in revised form 21 February 2012
Accepted 16 March 2012

Keywords:
Rifampicin
RNA polymerase
RbpA
Antibiotic tolerance
Coevolution
Protein–protein interaction
Transcription
Actinobacteria

ABSTRACT

The role of RbpA in the backdrop of *M. smegmatis* showed that it rescues mycobacterial RNA polymerase from rifampicin-mediated inhibition (Dey et al., 2010; Dey et al., 2011). Paget and co-workers (Paget et al., 2001; Newell et al., 2006) have revealed that RbpA homologs occur exclusively in actinobacteria. Newell et al. (2006) showed that MtbRbpA, when complemented in a Δ*rbpA* mutant of *S. coelicolor*, showed a low recovery of MIC (from 0.75 to 2 μg/ml) as compared to complementation by native RbpA of *S. coelicolor* (MIC increases from 0.75 to 11 μg/ml). Our studies on MsRbpA show that it is a differential marker for *M. smegmatis* RNA polymerase as compared to *E. coli* RNA polymerase at IC_{50} levels of rifampicin. A recent sequence-based analysis by Lane and Darst (2010) has shown that RNA polymerases from Proteobacteria and Actinobacteria have had a divergent evolution. *E. coli* is a representative of Proteobacteria and *M. smegmatis* is an Actinobacterium. RbpA has an exclusive occurrence in Actinobacteria. Since protein–protein interactions might not be conserved across different species, therefore, the probable reason for the indifference of MsRbpA toward *E. coli* RNA polymerase could be the lineage-specific differences between actinobacterial and proteobacterial RNA polymerases. These observations led us to ask the question as to whether the evolution of RbpA in Actinobacteria followed the same route as that of RNA polymerase subunits from actinobacterial species. We show that the exclusivity of RbpA in Actinobacteria and the unique evolution of RNA polymerase in this phylum share a co-evolutionary link. We have addressed this issue by a blending of experimental and bioinformatics based approaches. They comprise of induction of bacterial cultures coupled to rifampicin-tolerance, transcription assays and statistical comparison of phylogenetic trees for different pairs of proteins in actinobacteria.

1. Introduction

The situation inside the crowded environment of a prokaryotic cell is a classical case of organizational complexity. Macromolecular machines like the bacterial flagellum (Pijper, 1948; Berg, 2003), aminoacyl-tRNA synthetase (Delarue, 1995; Norcum et al., 2005), ribosome (Nomura, 1973; Sykes and Williamson, 2009), RNA polymerase (Ishihama, 1969; Ishihama and Ito, 1972; Ishihama et al., 1973; Zhang et al., 1999), DNA polymerase (O'Donnell and Kornberg, 1985; Yao and O'Donnell, 2009), bacteriorhodopsin (Kouyama et al., 1988; Khorana, 1988), secretion systems (Galan and Collmer, 1999), and RNA degradosome (Marcaida et al., 2006; Carpousis, 2007) can be regarded as the hallmarks of self-organization in the bacterial world. These multi-component, macromolecular complexes are ubiquitous in all three domains of life. By the end of the 20th century, a paradigm shift took place when Bruce Alberts (1998) described the cell as a collection of protein machines. This replaced the conventional view of treating bacteria as "bags of second-order chemical reactions" (Alberts, 1984) because these macromolecular machines are now well-known to form crucial, modular units of function in all cells. A macromolecular complex is a product of an ordered assembly of smaller subunits or proteins. This assembly results from physical interactions between its constituent proteins. Inside each protein assembly the intermolecular collisions, are not only restricted to a small set of possibilities, but are also controlled in a cascade (Alberts, 1998, 1984). However, these interactions have serious implications on the coevolution of interacting proteins as they govern major biochemical pathways. That is because, if one partner's binding surface undergoes any divergent changes then it needs to be complemented by the interacting partner at the interface (Goh et al., 2000; Atwell et al., 1997; Jespers et al., 1999; Moyle et al., 1994; Pazos et al., 1997). If co-evolution does not take place, interaction between the proteins will be lost, and consequently so would their function.

The conventional standard for judging the co-evolution of interacting proteins is by comparison of their phylogenetic trees. Initial observations of qualitative similarities between phylogenetic trees have been made in the case of interacting protein families of insulin and its receptors (Fryxell, 1996), dockerins/cohexins (Pages et al., 1997) and vasopressin/vasopressin receptors (van Kesteren et al., 1996). Such qualitative assessment was substantiated later by a quantification of the relationship between phylogenetic trees and protein interactions in large data sets (Goh et al., 2000; Pazos and Valencia, 2008).

* Corresponding author. Tel.: +91 8022932836; fax: +91 8023600535.
E-mail address: abhinavdey@gmail.com (A. Dey).

2212-0661
doi:10.1016/j.atg.2012.03.001

Our work on the interaction between MsRbpA and mycobacterial RNA polymerase has shown that one of the functional implications of this interaction, is the rescue of rifampicin-mediated inhibition of RNA polymerase activity (Dey et al., 2010, 2011). Subsequently, we probed for a similar effect of MsRbpA on *E. coli* RNA polymerase. The results showed an indifferent behavior of MsRbpA towards a heterologous system of *E. coli* when evaluated in the backdrop of rifampicin tolerance.

Thus, loss of function of actinobacterial MsRbpA on proteobacterial RNA polymerase (from *E. coli*) hint towards a lack of interaction across these species. This led us to a hypothesis on the existence of a co-evolutionary link between exclusivity of RbpA in Actinobacteria and the unique evolution of RNA polymerase in this phylum. In the work presented here, we intend to validate this hypothesis by employing a combination of experimental and bioinformatics based approaches.

2. Materials and methods

2.1. Bacterial strains and plasmids

M. smegmatis mc^2155 is the wild type strain. SM07 is a recombinant strain derived from mc^2155, harboring a chromosomal hexa-histidine tag on the *rpoC* gene (Mukherjee and Chatterji, 2008). RNA polymerase from *E. coli* was purified from strain RL916 (gift from Prof. Robert Landick, University of Wisconsin) (Brar et al., 2005). Jmc^2155 and $JRmc^2155$, are the plasmid transformed versions of mc^2155, carrying the plasmids pJAM2 (Triccas et al., 1998) and pJAM2MsRbpA (Dey et al., 2010), respectively. All *M. smegmatis* strains were grown on MB7H9 media (supplemented with bacto-agar, whenever required) along with 2% glucose and 0.05% Tween 80 (in case of liquid cultures). Jmc^2155 and $JRmc^2155$ were grown with 25 μg/ml Kanamycin. The concentrations of rifampicin varied from 2.5 μg/ml to 400 μg/ml. *Escherichia coli* strain BL21 (DE3) was used for gene expression experiments. The protein expression and purification experiments were carried out using *E. coli* strain BL21 (DE3). *M. smegmatis* strains Jmc^2155 and $JRmc^2155$ were induced by adding 2% acetamide to cultures. *E. coli* BL21 (DE3) cells were transformed with pETMsRbpA and induced with 1 mM IPTG in the presence of 100 μg/ml ampicillin.

2.2. Protein purification

RNA polymerase was purified from *M. smegmatis* strain SM07 and *E. coli* strain RL916 using protocols mentioned previously (Mukherjee and Chatterji, 2008; Brar et al., 2005). MsRbpA was purified using Ni-NTA affinity chromatography (Dey et al., 2010).

2.3. Transcription assays

IC_{50} was determined for RNA polymerase purified from *M. smegmatis* strain SM07 using multiple-round transcription assay described previously (Lowe et al., 1979). For judging the effect of MsRbpA on the RNA polymerase at the IC_{50} concentration of rifampicin, the assay buffer comprised of 40 mM Tris HCl pH 7.8, 200 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 14 mM β-mercaptoethanol, 200 μM each of ATP, GTP, CTP, 50 μM of UTP, and 2 μCi of ^{3}H-UTP (Perkin Elmer). The concentration of DMSO was maintained at 5% in all transcription reactions as a solvent for rifampicin. The assay mix comprising of the template DNA, enzymes and assay buffer, with/without rifampicin and with/without MsRbpA was incubated at 37 °C for 30 min and spotted on DE81 paper (Whatman) presoaked in 5 mM EDTA to stop the reaction. The DE81 papers were dried and washed with 5% Na_2HPO_4 (twice, 15 min each) followed by autoclaved double distilled water (thrice, 10 min each), and eventually with absolute ethanol and dried. Subsequently, the filters were placed into scintillation vials containing toluene-based scintillation fluid and counted by scintillation spectrometry. The concentration of rifampicin was varied depending upon the IC_{50} values, keeping the concentration of DMSO constant at 5%. The concentration of MsRbpA in the transcription assays was varied from 2-fold to 8-fold the concentration of RNA polymerase.

The purified MsRbpA was added to *E. coli* RNA polymerase in varying molar ratios at the IC_{50} of rifampicin (0.10 μg/ml) (Fujii et al., 1995). The assay was carried out by the same method as mentioned previously in case of *M. smegmatis* on calf-thymus DNA.

2.4. Construction of phylogenetic trees

The amino acid sequences of RbpA, β, β' subunits of RNA polymerase and 11 other proteins were retrieved from genome sequences of organisms belonging to the actinomycete phylum. The other proteins included: α-subunit of RNA polymerase, ω-subunit of RNA polymerase, GroEL1, GroEL2, S12 (30S ribosomal protein), Hsp70 (Heat shock proteins), HisA (phosphoribosyl isomerase A), PyrD (di-hydro-orotate dehydrogenase), EF-G (elongation factor-G), IF-2 (initiation factor-2), glucose-6-phosphate dehydrogenase. The candidate organisms from the actinomycetes have been enumerated in Table 1. Multiple sequence alignment and phylogenetic tree construction were carried out using the software MEGA 4.0.2 (Kumar et al., 2008).

2.5. Computation of pairwise distance matrix from pairwise distances between amino acid sequences

After qualitatitive analysis of the similarity between the phylogenetic trees obtained from the list of species mentioned in Table 1, the quantitative analyses required the computation of pairwise distances between

Table 1
Candidate organisms from actinobacterial phylum used for initial phylogenetic analysis.

Mycobacteria	Corynebacteria	Streptomyces	Other
Mycobacterium tuberculosis	*Corynebacterium aurimucosum*	*Streptomyces coelicolor*	*Nocardia farcinica*
Mycobacterium smegmatis	*Corynebacterium diphtheriae*	*Streptomyces avermitilis*	*Rhodococcus opacus*
Mycobacterium leprae	*Corynebacterium urealyticum*	*Streptomyces lividans*	*Rhodococcus erythropolis*
Mycobacterium bovis	*Corynebacterium pseudogenitalium*	*Streptomyces sviceus*	*Frankia alni*
	Corynebacterium striatum		*Tsukamurella paurometabola*
	Corynebacterium tuberculostearicum		*Actinosynnema mirum*
	Corynebacterium pseudogenitalium		*Gordinia bronchialis*
	Corynebacterium accolens		*Segniliparus rotundus*
	Corynebacterium glutamicum		*Saccharomonospora viridis*
	Corynebacterium matruchotii		*Saccharopolyspora erythrea*
	Corynebacterium glucurunolyticum		*Geodermitophilus obscures*
	Corynebacterium lipophiloflavum		*Thermobispora bispora*
	Corynebacterium jeikium		*Thermobifida fusca*
			Sanguibacter keddieii
			Stackerbrandita nassauensis
			Rhodococcus equi
			Streptosporangium roseum
			Micromonospora aurantica
			Nocardiopsis dassonvillie

pairs of amino acid sequences. The evolutionary distance between a pair of sequences is usually measured by the number of nucleotide (or amino acid) substitutions occurring between them. Evolutionary distances are fundamental for the study of molecular evolution and are useful for phylogenetic reconstructions and the estimation of divergence times. Distance matrices are further used for the calculation of correlation matrices.

The distance matrices were calculated only for those genes that were common among all the species mentioned in Table 1. Therefore, the phylogenetic trees of 16S rRNA, pyruvate dehydrogenase α/β subunits and polyketide synthase were excluded from the exercise of distance calculation. Also, species that did not contain all the genes were omitted to maintain uniformity. The final list of organisms whose sequences were used for subsequent calculations are shown in Table 2. The distance matrices were constructed using the same software MEGA 4.0.2, which was used for the construction of the phylogenetic tree.

2.6. Calculation of the Pearson's coefficients of correlation among the distance matrices

Co-evolution of genes can be quantitatively followed by measuring the similarity scores between the sets of values. A similarity score is obtained by calculating the Pearson's linear correlation coefficient between the two distance matrices (taking two matrices at a time). The matrices should have the same dimensions so as to be comparable. Therefore, as mentioned in the previous section, species that contained all the genes in common were considered and the remaining species were excluded. The distance matrices for the 25 species were constructed. The order of the pairwise distances of the 25 species was uniformly and stringently maintained in each matrix.

The desired pairwise combinations of the matrices were tabulated for which correlation coefficient was to be determined. Calculations of the Pearson's correlation were carried out using Microsoft Excel (MSOffice 2003 edition).

2.7. Testing the significance of the Pearson's coefficients of correlation

The value of the correlation coefficients must be tested for their statistical significance. For this we made use of the Student's *t*-test. A Student's *t*-test is a statistical hypothesis test that follows the Student's *t* distribution for validating the differences between the two sets of values. We exploit this information to validate the correlation coefficients obtained between the various pairs of matrices.

3. Results

3.1. Effect of MsRbpA on the activities of M. smegmatis and E. coli RNA polymerases in the presence of rifampicin at the IC_{50}

The RNA polymerase purified from *E. coli* RL916 was assayed to have a specific activity of 70 nmoles of ^{3}H-UTP/mg/30 min, while that of *M. smegmatis* RNA polymerase from SM07 was 64 nmoles of ^{3}H-UTP/mg/30 min. The IC_{50} of *E. coli* RNA polymerase for rifampicin has been previously reported as 0.10 μg/ml (Fujii et al., 1995) and that of *M. smegmatis* RNA polymerase is 0.05 μg/ml (Mukherjee and Chatterji, 2008). We tested the role of MsRbpA on *E. coli* RNA polymerase in vitro by enriching the transcription assay mixture with increasing ratios of MsRbpA:RNA polymerase. The same set of assays was carried out for *M. smegmatis* RNA polymerase. As can be seen in Fig. 1, 0.10 μg/ml of rifampicin inhibited the activity of *E. coli* RNA polymerase by 50%. When MsRbpA was added in increasing molar ratios to *E. coli* RNA polymerase, no recovery of activity took place (Fig. 1). On the contrary, MsRbpA was able to rescue *M. smegmatis* RNA polymerase at the IC_{50} concentrations of rifampicin (Fig. 1). Since the calf thymus DNA-based transcription assay is a non-specific method to judge the activity, it was important to look into the in vivo role of MsRbpA in *E. coli* and compare the same with *M. smegmatis*. This would also give a clearer picture of any promoter-specific activity of MsRbpA in increasing the rifampicin-tolerance levels of *E. coli*.

Table 2
List of organisms from Actinobacteria used for the quantitative analyses of phylogenetic distances.

Mycobacterium tuberculosis	*Tsukamurella paurometabola*
Mycobacterium smegmatis	*Actinosynnema mirum*
Mycobacterium leprae	*Gordinia bronchialis*
Mycobacterium bovis	*Segniliparus rotundus*
Corynebacterium diphtheriae	*Saccharomonospora viridis*
Corynebacterium aurimucosum	*Saccharopolyspora erythrea*
Streptomyces coelicolor	*Geodermitophilus obscures*
Streptomyces avermitilis	*Thermobispora bispora*
Nocardia farcinia	*Thermomonospora curvata*
Rhodococcus opacus	*Nakamurella multipartite*
Rhodococcus erythropolis	*Salinispora arenicola*
Frankia alni	*Thermobifida fusca*

3.2. Effect of overexpression of MsRbpA on the rifampicin-tolerance of M. smegmatis and E. coli

We chose the acetamidase promoter in order to have sufficient overexpression to allow expression of detectable amount of protein from mycobacterial cells. We cloned MsRbpA under acetamidase promoter in pJAM2 and electroporated the construct pJAM2MsRbpA into competent mc^{2}155 cells. As a control, only pJAM2 vector was also electroporated into competent mc^{2}155 cells (Dey et al., 2010). The vector pJAM2 has a kanamycin resistance marker. The resulting strains Jmc^{2}155 (carrying pJAM2) and JRmc^{2}155 (carrying pJAM2MsRbpA) were screened for the overexpression of MsRbpA in the presence of acetamide. The strains were then tested at different levels of rifampicin. The strain overexpressing MsRbpA, JRmc^{2}155, in the inducing conditions of 2% acetamide grew at rifampicin concentrations of 20 μg/ml, 40 μg/ml and 80 μg/ml, while the strain carrying pJAM2, Jmc^{2}155, was incapable of growing at these concentrations (Fig. 2). Thus, we found that overexpression of MsRbpA leads to increase in the rifampicin tolerance level in an otherwise rifampicin-sensitive strain carrying the vector alone.

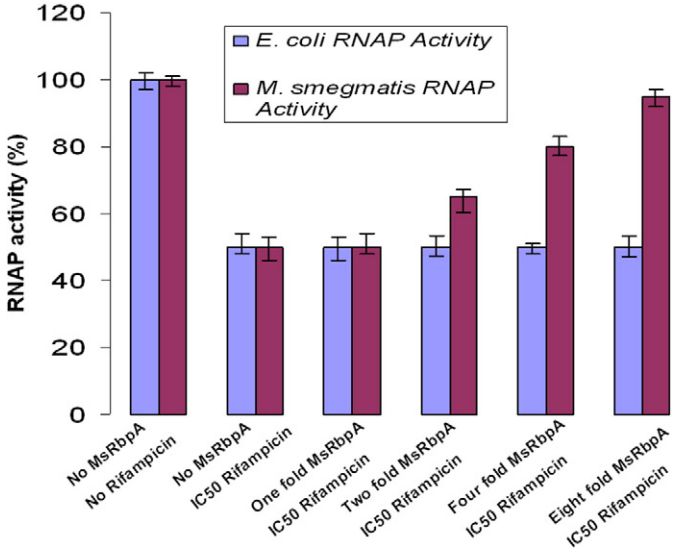

Fig. 1. Assessment of the role of MsRbpA on transcription activity of *E. coli* RNA Polymerase (RNAP) from RL916 (light shade) and *M. smegmatis* RNAP from SM07 (dark shade), at IC_{50} concentration of rifampicin (0.10 μg/ml for *E. coli* RNAP and 0.05 μg/ml for *M. smegmatis* RNAP). The bar-graph has been annotated with respect to the conditions of the assay. It can be seen that MsRbpA rescues the transcription activity of RNAP from *M. smegmatis* in presence of rifampicin, but is indifferent towards *E. coli* RNAP under similar conditions.

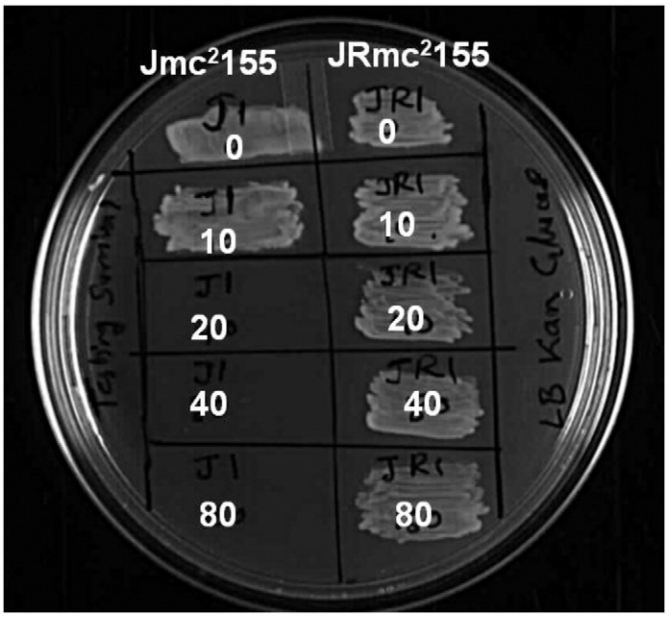

Fig. 2. MB 7H9 broth cultures of Jmc^{2}155 and JRmc^{2}155 cells were grown in presence of rifampicin (0, 10, 20, 40 and 80 μg/ml; shown in white) under inducing conditions of 2% acetamide. The surviving cells were pelleted, resuspended in 5 μl of LB and patched onto LB agar plates supplemented with 25 μg/ml of kanamycin and 2% glucose. The plates were scanned after 24 h of incubation at 37 °C. The *M. smegmatis* strain overexpressing MsRbpA (JRmc^{2}155) showed increase in MIC value for rifampicin as compared to the strain housing the vector backbone (Jmc^{2}155) only.

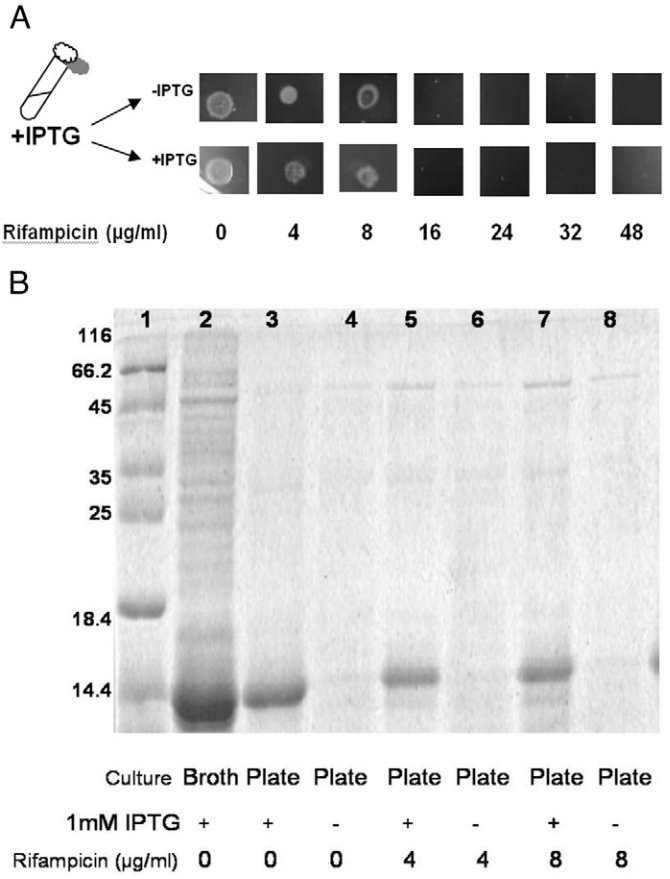

Fig. 3. A: Broth cultures of *E. coli* BL21 (DE3) cells, transformed with pETMsRbpA, were grown in LB (with IPTG) to $OD_{600} = 0.3$. One set was not induced with IPTG (upper panel) and one set was induced with 1 mM IPTG (lower panel). Serial, ten-fold dilutions were spotted (5 μl) onto LB agar plates supplemented with 100 μg/ml of ampicillin. Two series of plates for each set of broth cultures were made, one supplemented with 1 mM IPTG and the other without IPTG. A gradient of rifampicin was maintained (0 μg/ml, 4 μg/ml, 8 μg/ml, 16 μg/ml, 32 μg/ml and 64 μg/ml). The plates were scanned after 16 h of incubation at 37 °C. None of the tested series (under inducing conditions) showed any increase in MIC values for rifampicin.B: The expression state of MsRbpA in *E. coli* BL21 (DE3) transformed with pETMsRbpA. The results show the 15% SDS-PAGE expression analyses of MsRbpA for the series of experiments shown in the lower panel of Fig. 3A. Lane 1 = protein marker (kDa); Lane 2 = broth culture induced with 1 mM IPTG; Lanes 3, 5, and 7 = culture from LB agar with 1 mM IPTG; Lanes 4, 6, and 8 = culture from LB Agar with no IPTG.

The expression of MsRbpA in pETMsRbpA is under the control of T7-promoter fused with *lac* operator, therefore, it can act as a genetic switch to direct the expression of MsRbpA in *E. coli* BL21 (DE3). For this purpose, we transformed *E. coli* BL21 (DE3) cells with pETMsRbpA. Subsequently, we grew the transformed *E. coli* BL21 cells under inducing conditions (1 mM IPTG). The cells from these two sets were plated onto LB agar plates (with 100 μg/ml ampicillin). The plates contained a gradient of rifampicin (0 μg/ml, 4 μg/ml, 8 μg/ml, 16 μg/ml, 32 μg/ml and 64 μg/ml). However, as can be seen from Fig. 3A that overexpression of MsRbpA does not result in an increase in the MIC value of rifampicin for *E. coli*. At this point of time, it can be questioned as to whether any expression of MsRbpA actually took place in *E. coli* when the growth was taking place in IPTG. In parallel, it needs to be shown that there was a switch-off in the expression of MsRbpA in the absence of IPTG. Therefore, the growing colonies (shown in Fig. 3A) were picked, lysed and analyzed on a 15% SDS-PAGE. Fig. 3B depicts the results of MsRbpA in a switched-on or switched-off state. Lanes 4, 6, and 8 show the expression of MsRbpA in a switched-off state, while lanes 3, 5 and 7 show its expression in a switched-on state.

The probable reasons for this indifference on the part of MsRbpA could be:

a) Exclusivity of RbpA in actinobacteria or absence of RbpA-like proteins in *E. coli* (Paget et al., 2001; Newell et al., 2006).
b) Proteobacterial RNA polymerases show divergence from actinobacterial RNA polymerases in their phylogenetic trees (Lane and Darst, 2010).

Thus, loss of function of actinobacterial MsRbpA on proteobacterial RNA polymerase (from *E. coli*) hint towards a lack of interaction across these species. Additionally, we have proof that RbpA interacts with RNA polymerase in *Streptomyces coelicolor* (Newell et al., 2006) and MsRbpA interacts with RNA polymerase in *M. smegmatis* (Dey et al., 2010, 2011). Also MtbRbpA had a partial effect on increasing the MIC of rifampicin for a Δ*rbpA* strain of *S. coelicolor*, indicating a limited conservation of interaction among actinomycetes (Newell et al., 2006).

3.3. Co-evolution of RNA polymerase and RbpA in Actinobacteria

Co-evolution is prevalent in species at the organismic and molecular levels. It stands as an important function in the evolution of species and manifests itself in the host–parasite and predator–prey interactions. Proteins and their interacting partners also form important pairs that must co-evolve to maintain their specificity. The sequence changes in one partner must be complemented by corresponding changes in the other partner so as to maintain its functionality. Otherwise the interaction between the proteins is lost along with its function. Evolutionary studies on interacting proteins have also revealed the co-evolution of binding partners (Goh et al., 2000). The same approach was extrapolated in our study of the co-evolution of RbpA with actinobacterial RNA polymerase using bioinformatics approaches. As a control, phylogenetic trees of 9 other genes (apart from RbpA and the subunits of actinobacterial RNA polymerase) were constructed. These genes

included, glucose-6-phosphate dehydrogenase, GroEL1, GroEL2 and the six other anciently conserved proteins (Lake et al., 2009).

Phylogenetic analysis has shown that the trees of RbpA, RNA polymerase β and RNA polymerase β' subunits (Fig. 4A, B, and C) share a similarity in their appearance. As a control, the phylogenetic tree for the gene glucose-6-phosphate dehydrogenase (from same set of species) did not show a similar appearance (Fig. 4G). We have also analyzed the phylogenetic trees of RNA polymerase subunits α

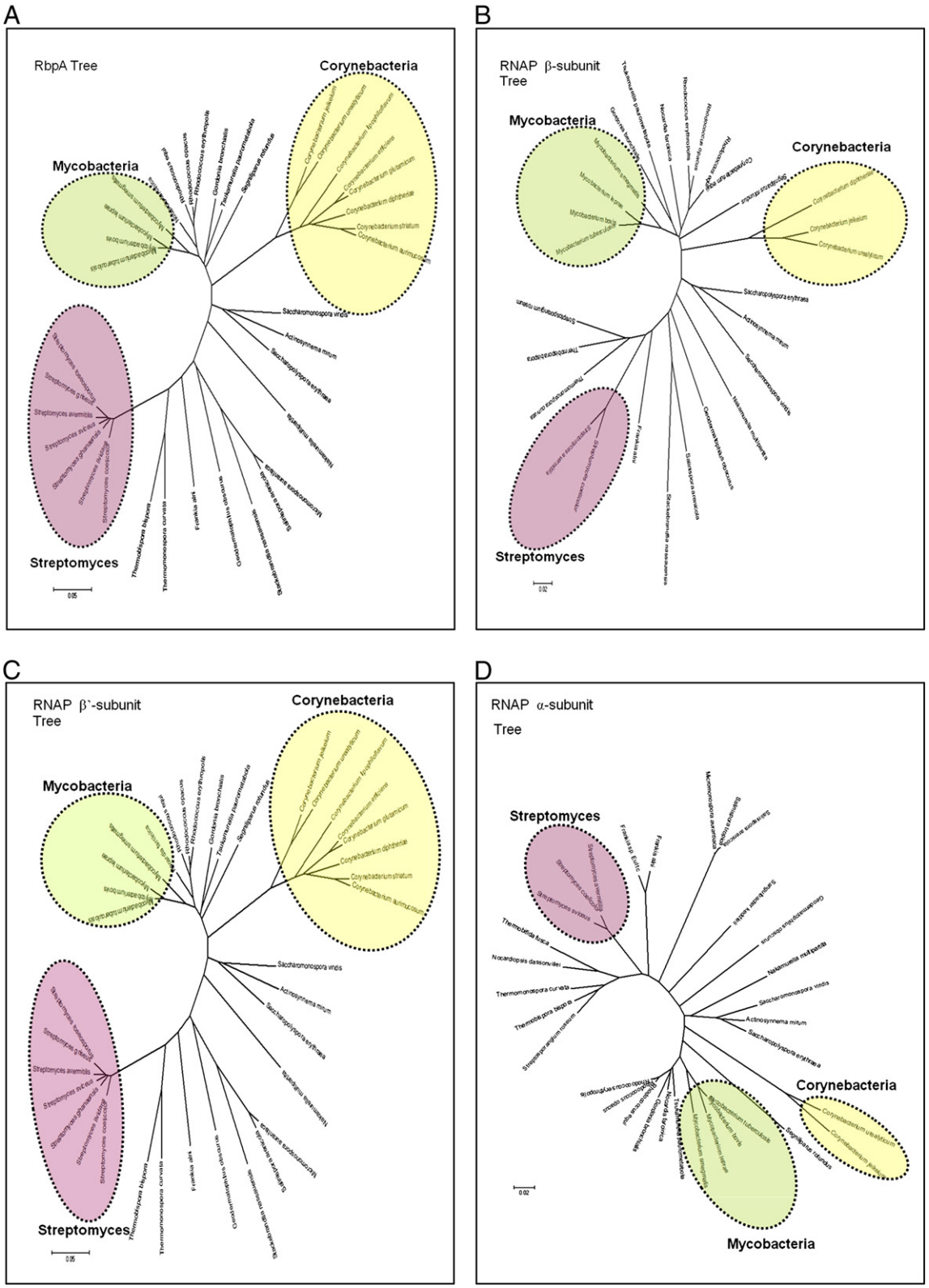

Fig. 4. A. Phylogenetic tree of RbpA in Actinobacteria.B: Phylogenetic tree of RNA polymerase β-subunit in Actinobacteria.C: Phylogenetic tree of RNA polymerase β'-subunit in Actinobacteria.D: Phylogenetic tree of RNA polymerase α-subunit.E: Phylogenetic tree of RNA polymerase ω-subunit.F: Phylogenetic tree of GroEL1.G: Phylogenetic tree of glucose-6-phosphate dehydrogenase in Actinobacteria.

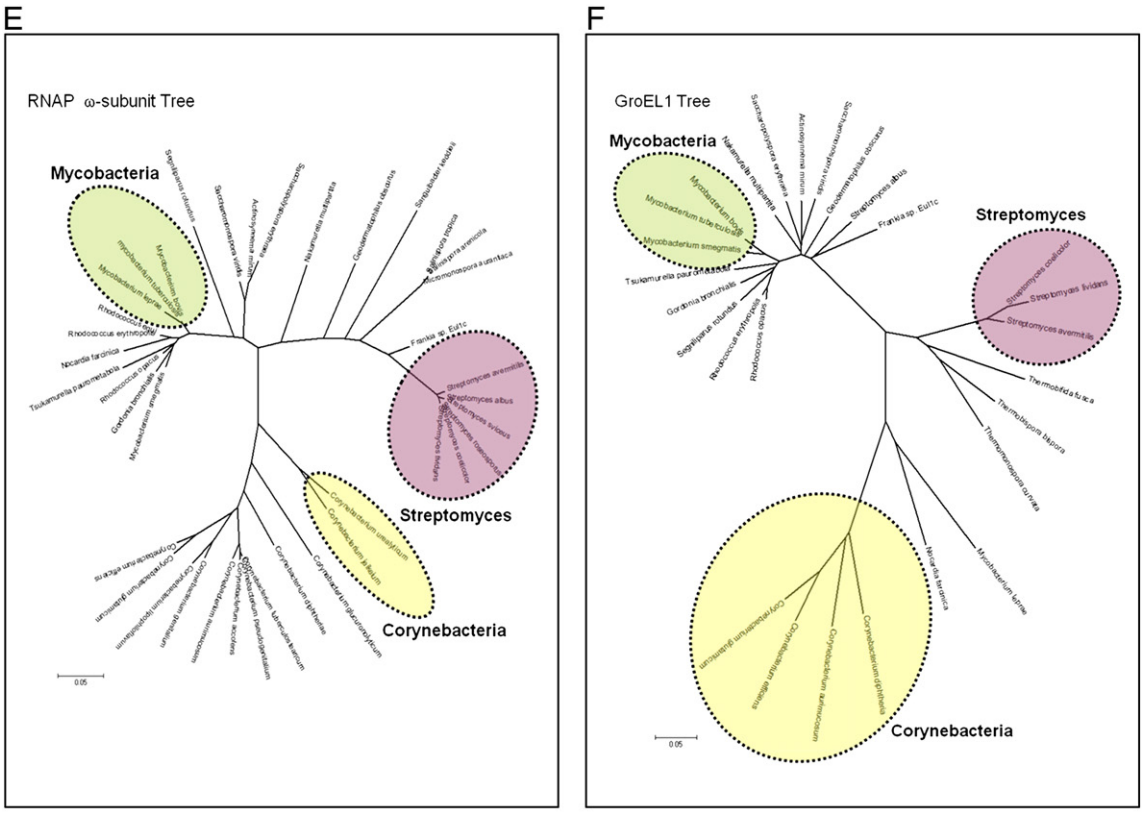

Fig. 4 (*continued*).

and ω, GroEL1 (Fig. 4D, E and F) as well as other genes (as mentioned in the Materials and methods; see Supplementary material). In order to ascertain that the observed similarity was not anecdotal, it was important to calculate the statistical relationship between tree similarities. For this purpose, we computed the pairwise distances between the members of each phylogenetic tree (for the same set of species; the values of the phylogenetic distances have been enlisted in Table 3A to D). Similar data analyses were carried out for

Table 3

A to G. Distance matrices for phylogenetic trees of RbpA, RNA polymerase β-subunit, RNA polymerase β'-subunit RNA polymerase α-subunit, RNA polymerase ω-subunit, GroEL1 and glucose-6-phosphate dehydrogenase from the selected actinobacterial species (the numbers in the matrices represent the following species: 1. *Mycobacterium tuberculosis;* 2. *Mycobacterium smegmatis*; 3. *Mycobacterium leprae*; 4. *Mycobacterium bovis*; 5. *Corynebacterium diphtheria*; 6. *Corynebacterium aurimucosum*; 7. *Streptomyces coelicolor*; 8. *Streptomyces avermitilis;* 9. *Nocardia farcinica*; 10. *Rhodococcus opacus;* 11. *Rhodococcus erythropolis*; 12. *Frankia alni*, 13. *Tsukamurella paurometabola*; 14. *Actinosynnema mirum*; 15. *Gordinia bronchialis*; 16. *Segniliparus rotundus*; 17. *Saccharomonospora viridis*; 18. *Saccharopolyspora erythrea*; 19. *Geodermitophilus obscures*; 20. *Thermobispora bispora*; 21. *Thermonospora curvata*; 22. *Nakamurella multipartite*; 23. *Salinospora arenicola*; 24. *Thermobifida fusca*).

A: Pairwise phylogenetic distances for RbpA in Actinobacteria

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
[1]	0.089																							
[2]	0.032	0.101																						
[3]	0	0.089	0.032																					
[4]	0.339	0.384	0.369	0.339																				
[5]	0.416	0.432	0.449	0.416	0.213																			
[6]	0.536	0.592	0.573	0.536	0.518	0.536																		
[7]	0.536	0.592	0.573	0.536	0.518	0.536	0																	
[8]	0.174	0.174	0.213	0.174	0.309	0.384	0.592	0.592																
[9]	0.213	0.149	0.226	0.213	0.324	0.416	0.573	0.573	0.149															
[10]	0.267	0.187	0.281	0.267	0.339	0.369	0.554	0.554	0.213	0.077														
[11]	0.592	0.611	0.631	0.592	0.672	0.693	0.416	0.416	0.611	0.592	0.592													
[12]	0.174	0.124	0.187	0.174	0.354	0.432	0.573	0.573	0.161	0.137	0.187	0.592												
[13]	0.354	0.384	0.4	0.354	0.483	0.554	0.5	0.5	0.369	0.384	0.4	0.536	0.369											
[14]	0.295	0.281	0.309	0.295	0.369	0.4	0.592	0.592	0.267	0.267	0.281	0.573	0.213	0.432										
[15]	0.267	0.239	0.281	0.267	0.324	0.416	0.5	0.5	0.213	0.239	0.239	0.554	0.253	0.449	0.309									
[16]	0.369	0.432	0.384	0.369	0.483	0.611	0.449	0.449	0.432	0.432	0.466	0.536	0.416	0.226	0.536	0.483								
[17]	0.354	0.416	0.4	0.354	0.449	0.573	0.432	0.432	0.369	0.416	0.449	0.536	0.4	0.137	0.449	0.432	0.187							
[18]	0.466	0.483	0.518	0.466	0.611	0.672	0.536	0.536	0.5	0.5	0.536	0.432	0.536	0.483	0.592	0.554	0.536	0.483						
[19]	0.518	0.573	0.554	0.518	0.672	0.693	0.432	0.432	0.611	0.611	0.592	0.416	0.631	0.518	0.592	0.573	0.432	0.5	0.554					
[20]	0.631	0.651	0.651	0.631	0.693	0.759	0.466	0.466	0.693	0.631	0.631	0.416	0.693	0.592	0.693	0.651	0.466	0.536	0.536	0.281				
[21]	0.449	0.4	0.466	0.449	0.536	0.672	0.573	0.573	0.449	0.4	0.466	0.518	0.416	0.384	0.483	0.483	0.339	0.339	0.554	0.518	0.5			
[22]	0.483	0.536	0.518	0.483	0.573	0.651	0.466	0.466	0.554	0.554	0.592	0.483	0.536	0.432	0.573	0.573	0.416	0.384	0.432	0.5	0.536	0.466		
[23]	0.483	0.536	0.518	0.483	0.573	0.651	0.466	0.466	0.554	0.554	0.592	0.483	0.536	0.432	0.573	0.573	0.416	0.384	0.432	0.5	0.518	0.466	0.011	
[24]	0.573	0.592	0.631	0.573	0.611	0.611	0.466	0.466	0.592	0.573	0.573	0.466	0.651	0.536	0.631	0.611	0.518	0.536	0.466	0.339	0.295	0.536	0.518	0.518

B: Pairwise phylogenetic distances for RNA polymerase β-subunit in Actinobacteria

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
[1]	0.074																							
[2]	0.041	0.069																						
[3]	0	0.074	0.041																					
[4]	0.248	0.239	0.249	0.248																				
[5]	0.248	0.227	0.24	0.248	0.154																			
[6]	0.248	0.247	0.25	0.248	0.313	0.316																		
[7]	0.244	0.242	0.247	0.244	0.31	0.307	0.042																	
[8]	0.113	0.106	0.108	0.113	0.226	0.225	0.232	0.224																
[9]	0.121	0.091	0.116	0.121	0.239	0.229	0.243	0.236	0.086															
[10]	0.124	0.102	0.116	0.124	0.236	0.229	0.247	0.235	0.099	0.035														
[11]	0.215	0.213	0.219	0.215	0.299	0.305	0.174	0.174	0.212	0.216	0.231													
[12]	0.136	0.108	0.126	0.136	0.237	0.229	0.25	0.25	0.13	0.113	0.121	0.235												
[13]	0.171	0.153	0.164	0.171	0.265	0.259	0.224	0.224	0.15	0.148	0.149	0.196	0.178											
[14]	0.143	0.108	0.123	0.143	0.241	0.233	0.265	0.259	0.123	0.104	0.115	0.234	0.112	0.162										
[15]	0.141	0.124	0.134	0.141	0.251	0.239	0.251	0.25	0.13	0.119	0.13	0.225	0.147	0.173	0.146									
[16]	0.174	0.171	0.179	0.174	0.27	0.269	0.243	0.249	0.166	0.177	0.177	0.215	0.186	0.14	0.193	0.186								
[17]	0.159	0.143	0.151	0.159	0.26	0.251	0.222	0.223	0.139	0.147	0.153	0.178	0.169	0.095	0.162	0.179	0.126							
[18]	0.202	0.195	0.197	0.202	0.294	0.282	0.202	0.194	0.195	0.204	0.213	0.173	0.218	0.166	0.22	0.205	0.21	0.178						
[19]	0.222	0.228	0.218	0.222	0.308	0.291	0.162	0.166	0.216	0.234	0.232	0.146	0.237	0.199	0.247	0.242	0.224	0.19	0.195					
[20]	0.229	0.234	0.231	0.229	0.308	0.308	0.167	0.164	0.222	0.241	0.236	0.134	0.245	0.2	0.243	0.241	0.22	0.199	0.172	0.111				
[21]	0.197	0.192	0.197	0.197	0.291	0.288	0.255	0.245	0.194	0.186	0.194	0.212	0.2	0.179	0.202	0.206	0.193	0.177	0.19	0.225	0.216			
[22]	0.236	0.235	0.24	0.236	0.3	0.299	0.218	0.217	0.223	0.231	0.234	0.177	0.243	0.2	0.258	0.231	0.227	0.192	0.186	0.207	0.193	0.231		

(continued on next page)

Table 3 (*continued*)

[23]	0.227	0.228	0.233	0.227	0.295	0.288	0.214	0.213	0.219	0.225	0.231	0.165	0.236	0.195	0.243	0.219	0.22	0.186	0.18	0.2	0.18	0.22	0.025	
[24]	0.24	0.248	0.243	0.24	0.305	0.308	0.19	0.187	0.228	0.241	0.239	0.176	0.249	0.241	0.26	0.247	0.235	0.236	0.218	0.139	0.121	0.255	0.217	0.213

C: Pairwise phylogenetic distances for RNA polymerase β'-subunit from Actinobacteria

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
[1]	0.08																							
[2]	0.038	0.096																						
[3]	0	0.08	0.038																					
[4]	0.275	0.275	0.286	0.275																				
[5]	0.273	0.275	0.285	0.273	0.19																			
[6]	0.312	0.311	0.322	0.312	0.406	0.403																		
[7]	0.314	0.314	0.326	0.314	0.412	0.405	0.029																	
[8]	0.102	0.091	0.112	0.102	0.272	0.286	0.297	0.3																
[9]	0.118	0.105	0.123	0.118	0.285	0.294	0.301	0.304	0.071															
[10]	0.124	0.109	0.126	0.124	0.288	0.299	0.304	0.305	0.082	0.031														
[11]	0.29	0.288	0.295	0.29	0.392	0.402	0.225	0.229	0.282	0.279	0.277													
[12]	0.128	0.128	0.142	0.128	0.269	0.282	0.305	0.308	0.11	0.12	0.128	0.292												
[13]	0.218	0.204	0.22	0.218	0.305	0.329	0.295	0.299	0.192	0.178	0.186	0.29	0.218											
[14]	0.129	0.133	0.145	0.129	0.286	0.284	0.312	0.311	0.115	0.116	0.121	0.294	0.112	0.219										
[15]	0.156	0.158	0.167	0.156	0.281	0.291	0.345	0.346	0.152	0.163	0.17	0.325	0.144	0.246	0.158									
[16]	0.204	0.194	0.213	0.204	0.312	0.32	0.291	0.293	0.179	0.181	0.189	0.275	0.212	0.154	0.21	0.232								
[17]	0.201	0.194	0.203	0.201	0.319	0.333	0.282	0.281	0.175	0.17	0.179	0.273	0.195	0.157	0.203	0.226	0.158							
[18]	0.276	0.267	0.28	0.276	0.382	0.378	0.228	0.234	0.269	0.269	0.269	0.24	0.279	0.255	0.285	0.312	0.26	0.243						
[19]	0.323	0.318	0.332	0.323	0.389	0.398	0.205	0.212	0.308	0.318	0.325	0.246	0.319	0.304	0.322	0.339	0.293	0.298	0.267					
[20]	0.31	0.304	0.318	0.31	0.396	0.396	0.206	0.213	0.303	0.31	0.31	0.229	0.312	0.301	0.311	0.33	0.288	0.288	0.25	0.133				
[21]	0.253	0.247	0.263	0.253	0.339	0.362	0.319	0.331	0.245	0.255	0.267	0.302	0.258	0.255	0.272	0.279	0.247	0.247	0.292	0.329	0.327			
[22]	0.296	0.296	0.309	0.296	0.397	0.39	0.264	0.271	0.296	0.305	0.304	0.276	0.298	0.28	0.315	0.331	0.287	0.285	0.238	0.274	0.264	0.311		
[23]	0.296	0.297	0.311	0.296	0.397	0.391	0.261	0.268	0.297	0.306	0.305	0.275	0.3	0.28	0.317	0.333	0.287	0.285	0.239	0.27	0.261	0.309	0.009	
[24]	0.337	0.333	0.345	0.337	0.419	0.422	0.234	0.242	0.335	0.339	0.341	0.26	0.34	0.332	0.343	0.361	0.316	0.314	0.282	0.173	0.162	0.356	0.281	0.279

D: Pairwise distances for RNA polymerase α subunit

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
[1]	0.053																							
[2]	0.497	0.487																						
[3]	0	0.053	0.497																					
[4]	0.604	0.582	0.48	0.604																				
[5]	0.56	0.546	0.48	0.56	0.231																			
[6]	0.361	0.352	0.477	0.361	0.497	0.497																		
[7]	0.37	0.361	0.457	0.37	0.514	0.511	0.049																	
[8]	0.47	0.467	0.304	0.47	0.391	0.376	0.432	0.425																
[9]	0.095	0.093	0.497	0.095	0.571	0.539	0.352	0.346	0.461															
[10]	0.097	0.086	0.487	0.097	0.564	0.532	0.344	0.338	0.457	0.027														
[11]	0.175	0.175	0.504	0.175	0.582	0.571	0.37	0.352	0.467	0.18	0.173													
[12]	0.111	0.097	0.497	0.111	0.571	0.542	0.361	0.361	0.467	0.093	0.097	0.19												
[13]	0.125	0.123	0.487	0.125	0.597	0.575	0.367	0.355	0.457	0.139	0.139	0.17	0.139											
[14]	0.102	0.088	0.497	0.102	0.578	0.557	0.346	0.349	0.47	0.091	0.088	0.188	0.111	0.149										
[15]	0.109	0.113	0.504	0.109	0.586	0.539	0.344	0.338	0.457	0.091	0.093	0.188	0.12	0.166	0.111									
[16]	0.132	0.137	0.501	0.132	0.586	0.542	0.37	0.37	0.464	0.166	0.156	0.175	0.153	0.1	0.168	0.178								
[17]	0.142	0.144	0.497	0.142	0.589	0.578	0.338	0.344	0.477	0.156	0.156	0.175	0.163	0.111	0.156	0.178	0.118							
[18]	0.144	0.137	0.477	0.144	0.575	0.567	0.332	0.323	0.464	0.142	0.139	0.151	0.142	0.12	0.146	0.158	0.139	0.12						
[19]	0.361	0.361	0.497	0.361	0.525	0.525	0.26	0.257	0.464	0.358	0.358	0.361	0.358	0.37	0.352	0.346	0.358	0.358	0.364					
[20]	0.391	0.379	0.507	0.391	0.532	0.539	0.252	0.247	0.454	0.376	0.376	0.373	0.364	0.376	0.364	0.361	0.367	0.355	0.358	0.19				
[21]	0.134	0.125	0.48	0.134	0.567	0.549	0.341	0.332	0.444	0.134	0.132	0.173	0.139	0.102	0.144	0.168	0.116	0.137	0.12	0.344	0.364			
[22]	0.188	0.18	0.48	0.188	0.567	0.567	0.341	0.338	0.477	0.188	0.188	0.163	0.18	0.158	0.183	0.213	0.18	0.178	0.134	0.352	0.361	0.163		
[23]	0.19	0.183	0.484	0.19	0.564	0.567	0.344	0.341	0.474	0.19	0.19	0.163	0.183	0.161	0.185	0.215	0.183	0.18	0.137	0.355	0.364	0.166	0.004	
[24]	0.388	0.385	0.487	0.388	0.511	0.532	0.252	0.241	0.444	0.379	0.373	0.379	0.394	0.394	0.382	0.367	0.388	0.376	0.394	0.255	0.244	0.388	0.388	0.385

Table 3 (*continued*)

E: Pairwise distances for RNA polymerase ω subunit

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
[1]	0.053																							
[2]	0.026	0.08																						
[3]	0	0.053	0.026																					
[4]	0.331	0.386	0.368	0.331																				
[5]	0.296	0.349	0.296	0.296	0.137																			
[6]	0.331	0.368	0.368	0.331	0.445	0.425																		
[7]	0.331	0.368	0.368	0.331	0.445	0.425	0.026																	
[8]	0.094	0.094	0.123	0.094	0.405	0.368	0.349	0.349																
[9]	0.053	0.039	0.08	0.053	0.386	0.349	0.349	0.349	0.08															
[10]	0.053	0.053	0.08	0.053	0.386	0.349	0.331	0.331	0.066	0.039														
[11]	0.262	0.296	0.296	0.262	0.405	0.349	0.094	0.108	0.279	0.279	0.262													
[12]	0.094	0.094	0.094	0.094	0.405	0.368	0.386	0.386	0.137	0.08	0.094	0.349												
[13]	0.137	0.182	0.167	0.137	0.296	0.331	0.296	0.296	0.198	0.182	0.167	0.246	0.23											
[14]	0.08	0.053	0.08	0.08	0.425	0.349	0.386	0.386	0.108	0.053	0.066	0.314	0.094	0.167										
[15]	0.182	0.214	0.214	0.182	0.386	0.331	0.368	0.368	0.23	0.214	0.198	0.279	0.262	0.198	0.23									
[16]	0.137	0.182	0.167	0.137	0.296	0.331	0.296	0.296	0.198	0.182	0.167	0.246	0.23	0.053	0.198	0.198								
[17]	0.152	0.198	0.182	0.152	0.349	0.349	0.314	0.314	0.214	0.198	0.152	0.262	0.246	0.039	0.182	0.198	0.08							
[18]	0.262	0.279	0.296	0.262	0.368	0.349	0.182	0.198	0.296	0.262	0.262	0.152	0.279	0.246	0.296	0.314	0.246	0.262						
[19]	0.331	0.386	0.368	0.331	0.465	0.425	0.246	0.23	0.425	0.386	0.405	0.182	0.425	0.314	0.405	0.349	0.331	0.349	0.279					
[20]	0.368	0.405	0.405	0.368	0.486	0.465	0.262	0.262	0.425	0.405	0.405	0.246	0.465	0.314	0.425	0.368	0.314	0.349	0.279	0.198				
[21]	0.214	0.246	0.246	0.214	0.331	0.368	0.296	0.296	0.262	0.23	0.23	0.246	0.279	0.167	0.279	0.314	0.182	0.214	0.214	0.386	0.368			
[22]	0.349	0.349	0.349	0.349	0.405	0.386	0.23	0.214	0.368	0.349	0.349	0.182	0.386	0.296	0.349	0.368	0.279	0.331	0.198	0.246	0.262	0.314		
[23]	0.349	0.349	0.349	0.349	0.405	0.386	0.23	0.214	0.368	0.349	0.349	0.182	0.386	0.296	0.349	0.368	0.279	0.331	0.198	0.246	0.262	0.314	0	
[24]	0.331	0.368	0.368	0.331	0.425	0.425	0.23	0.23	0.386	0.368	0.368	0.182	0.425	0.262	0.386	0.331	0.262	0.296	0.23	0.152	0.108	0.314	0.214	0.214

F: Pairwise distances for GroEL1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
[1]	0.053																							
[2]	0.497	0.487																						
[3]	0	0.053	0.497																					
[4]	0.604	0.582	0.48	0.604																				
[5]	0.56	0.546	0.48	0.56	0.231																			
[6]	0.361	0.352	0.477	0.361	0.497	0.497																		
[7]	0.37	0.361	0.457	0.37	0.514	0.511	0.049																	
[8]	0.47	0.467	0.304	0.47	0.391	0.376	0.432	0.425																
[9]	0.095	0.093	0.497	0.095	0.571	0.539	0.352	0.346	0.461															
[10]	0.097	0.086	0.487	0.097	0.564	0.532	0.344	0.338	0.457	0.027														
[11]	0.175	0.175	0.504	0.175	0.582	0.571	0.37	0.352	0.467	0.18	0.173													
[12]	0.111	0.097	0.497	0.111	0.571	0.542	0.361	0.361	0.467	0.093	0.097	0.19												
[13]	0.125	0.123	0.487	0.125	0.597	0.575	0.367	0.355	0.457	0.139	0.139	0.17	0.139											
[14]	0.102	0.088	0.497	0.102	0.578	0.557	0.346	0.349	0.47	0.091	0.088	0.188	0.111	0.149										
[15]	0.109	0.113	0.504	0.109	0.586	0.539	0.344	0.338	0.457	0.091	0.093	0.188	0.12	0.166	0.111									
[16]	0.132	0.137	0.501	0.132	0.586	0.542	0.37	0.37	0.464	0.166	0.156	0.175	0.153	0.1	0.168	0.178								
[17]	0.142	0.144	0.497	0.142	0.589	0.578	0.338	0.344	0.477	0.156	0.156	0.175	0.163	0.111	0.156	0.178	0.118							
[18]	0.144	0.137	0.477	0.144	0.575	0.567	0.332	0.323	0.464	0.142	0.139	0.151	0.142	0.12	0.146	0.158	0.139	0.12						
[19]	0.361	0.361	0.497	0.361	0.525	0.525	0.26	0.257	0.464	0.358	0.358	0.361	0.358	0.37	0.352	0.346	0.358	0.358	0.364					
[20]	0.391	0.379	0.507	0.391	0.532	0.539	0.252	0.247	0.454	0.376	0.376	0.373	0.364	0.376	0.364	0.361	0.367	0.355	0.358	0.19				
[21]	0.134	0.125	0.48	0.134	0.567	0.549	0.341	0.332	0.444	0.134	0.132	0.173	0.139	0.102	0.144	0.168	0.116	0.137	0.12	0.344	0.364			
[22]	0.188	0.18	0.48	0.188	0.567	0.567	0.341	0.338	0.477	0.188	0.188	0.163	0.18	0.158	0.183	0.213	0.18	0.178	0.134	0.352	0.361	0.163		
[23]	0.19	0.183	0.484	0.19	0.564	0.567	0.344	0.341	0.474	0.19	0.19	0.163	0.183	0.161	0.185	0.215	0.183	0.18	0.137	0.355	0.364	0.166	0.004	
[24]	0.388	0.385	0.487	0.388	0.511	0.532	0.252	0.241	0.444	0.379	0.373	0.379	0.394	0.394	0.382	0.367	0.388	0.376	0.394	0.255	0.244	0.388	0.388	0.385

G: Pairwise phylogenetic distances for glucose-6-phosphate dehydrogenase from Actinobacteria

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
[1]	0.098																							
[2]	0.114	0.142																						

(*continued on next page*)

Table 3 (*continued*)

[3]	0	0.098	0.114																												
[4]	2.135	2.048	2.197	2.135																											
[5]	2.165	2.048	2.197	2.165	0.211																										
[6]	1.036	1.036	1.109	1.036	2.165	2.165																									
[7]	1.397	1.427	1.472	1.397	2.299	2.335	1.488																								
[8]	0.155	0.151	0.22	0.155	1.894	1.942	1.036	1.457																							
[9]	0.155	0.102	0.18	0.155	2.165	2.165	1.057	1.537	0.151																						
[10]	0.167	0.118	0.197	0.167	2.02	2.048	1.046	1.504	0.155	0.094																					
[11]	1.057	1.046	1.109	1.057	2.076	2.105	0.513	1.504	1.036	1.036	0.977																				
[12]	0.247	0.215	0.275	0.247	2.165	2.135	1.016	1.427	0.265	0.256	0.224	0.987																			
[13]	1.109	1.131	1.131	1.109	2.02	2.105	1.057	1.68	1.057	1.165	1.109	0.931	1.165																		
[14]	0.247	0.197	0.275	0.247	2.105	2.105	1.046	1.504	0.224	0.224	0.18	1.036	0.159	1.12																	
[15]	0.279	0.284	0.348	0.279	2.197	2.165	1.057	1.472	0.303	0.289	0.294	0.968	0.338	1.067	0.294																
[16]	2.135	2.105	2.23	2.135	0.338	0.318	2.165	2.264	1.994	2.197	2.048	2.105	2.165	2.299	2.165	2.23															
[17]	2.105	2.02	2.23	2.105	0.308	0.298	2.264	2.335	1.942	2.135	1.994	2.197	2.135	2.264	2.076	2.165	0.197														
[18]	2.264	2.197	2.299	2.264	0.94	0.968	2.135	2.741	2.165	2.264	2.197	2.197	2.264	2.264	2.299	2.264	0.931	0.949													
[19]	1.942	1.87	2.048	1.942	0.422	0.374	2.135	2.197	1.803	1.994	1.894	2.135	2.02	1.942	1.918	2.048	0.328	0.333	0.844												
[20]	2.105	1.994	2.23	2.105	0.353	0.358	2.197	2.165	1.918	2.135	1.968	2.135	2.02	2.165	1.994	2.135	0.284	0.261	0.968	0.247											
[21]	2.048	1.968	2.105	2.048	0.395	0.363	2.197	2.135	1.894	1.968	2.048	2.105	1.994	2.165	1.942	2.048	0.289	0.261	0.987	0.318	0.318										
[22]	1.109	1.109	1.165	1.109	2.048	2.105	0.913	1.488	1.099	1.131	1.109	0.844	1.099	0.996	1.109	1.077	2.135	2.135	2.105	1.968	2.048	2.105									
[23]	1.131	1.143	1.177	1.131	2.105	2.135	0.913	1.52	1.12	1.154	1.143	0.852	1.12	0.996	1.131	1.088	2.165	2.165	2.105	2.02	2.105	2.135	0.102								
[24]	2.076	2.048	2.165	2.076	0.379	0.369	2.299	2.23	2.02	2.135	2.048	2.299	2.135	2.299	2.105	2.165	0.289	0.247	0.869	0.233	0.238	0.298	2.135	2.135							

the remaining set of trees obtained from the phylogenetic analyses of the other genes (mentioned in Materials and methods; see Supplementary material).

The distance matrices were then arrayed alongside each other and Pearson's coefficient of correlation was calculated for each pair of distance matrices. Statistical values were tabulated and analyzed for their significance using Student's *t*-test (Tables 4 and 5). Similar work carried out on the phylogenetic trees of NuoE and NuoF subunits of the *E. coli* NADH dehydrogenase complex displayed a discernible similarity (0.86 in a 0–1 scale). The two subunits interact tightly as observed in the crystal structure of the complex (PDB id: 2fug). A similar range of values were observed for the subunits of RNA polymerase in actinobacteria, when compared with each other (Table 6). From Table 7, we observed that correlation coefficient between RbpA and the different subunits of actinobacterial RNA polymerase are 0.82 (for α and RbpA), 0.85 (for β and RbpA), 0.89 (for β' and RbpA) and 0.81 (for ω and RbpA). So the phylogenetic trees of RbpA and RNA polymerase subunits showed a strong correlation between them. Therefore, co-evolution of these interacting proteins, i.e., RbpA and RNA polymerase, gets strongly emphasized here. The important observation to be noted here is that the phylogenetic trees of EF-G and ribosomal protein S12 share very low correlation with RbpA (Table 8). This low correlation is sustained when they are compared with RNA polymerase subunits also. On the other hand, the phylogenetic trees of Hsp70 and IF-2 show a high degree of correlation with RbpA, but again, it can be seen that they share a high degree of correlation with the trees of RNA polymerase subunits also. This relationship appears to be reflexive in the case of these proteins.

A final validation of correlation coefficients was done to ensure positive correlation between the trees. Therefore, correlation coefficients were tested to check their significance.

3.4. Significance of the value of correlation coefficient

The calculated value of 't' is compared with the table value of 't' at $\alpha = 0.05$ and 298° of freedom i.e. 1.64. Since, the calculated value is greater than the table t-value, the null hypothesis is rejected. In other words, the alternate hypothesis is accepted. So the phylogenetic distance matrices are significantly correlated.

The Pearson's correlation coefficients and the Student's 't' test values, both approve the hypothesis that the two interacting proteins, viz. RbpA and RNA polymerase have co-evolved during the course of evolution in actinobacteria.

4. Discussion

We have reported previously (Dey et al., 2010, 2011) that MsRbpA rescues mycobacterial RNA polymerase from the transcription inhibition caused by rifampicin in vitro. As a corroboration of this work, it was shown previously (Dey et al., 2010) and in the present study, that the induction of MsRbpA in vivo causes an increase in the rifampicin-tolerance levels (MIC) of *M. smegmatis*.

In this manuscript, we report the assessment of MsRbpA for its role in phenotypic tolerance to rifampicin in a heterologous system of *E. coli*. The results both in vitro and in vivo, have shown that the rescuing effect from rifampicin is not shown on *E. coli* RNA polymerase and on the intrinsic rifampicin-resistance level of *E. coli*. These results, in spite of being negative, might have implications on the evolution of RNA polymerase in *M. smegmatis* and *E. coli*. It has recently been predicted by Lane and Darst (2010) that RNA polymerase has had a divergent evolution in the bacterial kingdom. Especially, Proteobacteria and Actinobacteria have had a diametrically divergent evolution. Notwithstanding that there have been sporadic occasions where the opinion about protein–protein interactions (involving RNA polymerase) not being conserved across species has been

Table 4
Pearson's coefficients of correlation between the individual pairs of phylogenetic distance matrices.

	α	β	β'	ω	RbpA	HisA	Hsp70	G6pd	S12	EF-g	If-2	Groel1	Groel2	PyrD
α	–	0.896	0.917	0.859	0.824	0.769	0.778	0.395	0.185	0.152	0.876	0.427	0.465	0.767
β		–	0.960	0.896	0.846	0.784	0.822	0.408	0.136	0.227	0.845	0.515	0.522	0.744
β'			–	0.902	0.890	0.730	0.809	0.435	0.202	0.129	0.905	0.456	0.485	0.749
ω				–	0.807	0.683	0.759	0.399	0.155	0.130	0.791	0.506	0.457	0.737
RbpA					–	0.590	0.790	0.427	0.272	0.109	0.855	0.329	0.448	0.682
HisA						–	0.645	0.321	0.026	0.285	0.623	0.593	0.448	0.641
Hsp70							–	0.345	0.101	0.228	0.673	0.340	0.489	0.696
G6pd								–	0.136	0.051	0.459	0.148	0.191	0.254
S12									—	−0.012	0.335	−0.15	0.07	0.219
Ef-g										–	0.002	0.147	0.269	0.277
If-2											–	0.326	0.437	0.702
Groel1												–	0.485	0.335
Groel2													–	0.376
PyrD														–

Table 5
Calculated t-values for the correlation coefficients.

	α	β	β'	ω	RbpA	HisA	Hsp70	G6pd	S12	EF-g	If-2	Groel1	Groel2	PyrD
α	–	34.832	39.685	28.964	25.105	20.767	21.377	7.422	3.25	2.655	31.353	8.152	9.067	20.635
β		–	59.186	34.832	27.391	21.802	24.917	7.714	2.37	4.024	27.277	10.371	10.565	19.222
β'			–	36.066	33.695	18.439	23.759	8.34	3.56	2.246	36.724	8.845	9.574	19.515
ω				–	23.59	16.142	20.124	7.512	2.708	2.263	22.318	10.127	8.869	18.823
RbpA					–	12.614	22.243	8.152	4.879	1.893	28.459	6.014	8.65	16.098
HisA						–	14.57	5.851	0.449	5.133	13.749	12.713	8.65	14.417
Hsp70							–	6.345	1.752	4.042	15.707	6.241	9.677	16.733
G6pd								–	2.37	0.882	8.919	2.583	3.359	4.533
S12									–	−0.207	6.138	−2.59	1.211	3.875
Ef-g										–	0.035	2.565	4.821	4.024
If-2											–	5.953	8.387	6.138
Groel1												–	9.574	6.965
Groel2													–	7.005
PyrD														–

expressed (Steffen and Ullmann, 1998; Mencía et al., 1998; Lohrke et al., 1999), RNA polymerase from different species may also have different properties (Artsimovitch et al., 2000). Thus, MsRbpA can serve as a differential marker for RNA polymerase from *M. smegmatis* and *E. coli*.

Our bioinformatics-based statistical analyses show high correlation coefficients and significant Student's t-values for RNA polymerase subunits and RbpA from Actinobacteria. This indicates that the divergent evolution of RNA polymerase among the phylum actinomycetes is highly correlated with divergent evolution of RbpA, existing exclusively in the same phylum. Lower values of correlation coefficients between RNA polymerase subunits and anciently conserved proteins served as a negative control for our analyses (Table 8). They suggest that though speciation is an important phenomenon in the course of evolution, the high correlation coefficient between RNA polymerase and RbpA is due to co-evolution and not just speciation.

Therefore, it appears that rifampicin being a metabolite from soil actinomycete may have pre-exposed itself to other soil bacteria in the course of evolution. This might have led to a phenotypic defense system comprising of RbpA, co-evolving with the actinobacterial RNA polymerase.

Table 6
Correlation between the subunits of RNA polymerase.

Pairs of proteins	Correlation coefficients	Significance scores
α and β	0.896	34.832
α and β'	0.917	39.685
α and ω	0.859	28.964
β and β'	0.96	59.186
β and ω	0.896	34.832
β' and ω	0.902	36.066

Table 7
Correlation between RbpA and subunits of RNA polymerase from actinobacteria.

RNA polymerase subunits	Correlation coefficients	Significance scores
α	0.824	25.105
β	0.846	27.391
β'	0.890	33.695
ω	0.807	23.59

Acknowledgments

AD thanks CSIR and IISc for fellowship. ARV acknowledges JNCASR, Bangalore, for summer research fellowship. The authors acknowledge the suggestions of Prof. N. Srinivasan (MBU, IISc) and Prof. N.V. Joshi (CES, IISc) during the course of this study. The authors thank Dr. Anshu Malhotra for carefully proofreading the manuscript.

Table 8
Correlation coefficients between RbpA and anciently conserved proteins from actinobacteria.

Anciently conserved proteins	Correlation coefficients	Significance scores
Hsp70	0.790	22.243
S12	0.272	4.879
HisA	0.590	12.614
Elongation factor G	0.109	1.893
Initiation factor 2	0.855	28.459
pyrD	0.682	16.098

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.atg.2012.03.001.

References

Alberts BM. The DNA enzymology of protein machines. Cold Spring Harbor Symposia on Quantitative Biology 1984;49:1–12.

Alberts BM. The cell as a collection of protein machines: preparing the next generation of molecular biologists. Cell 1998;92:291–4.

Artsimovitch I, Svetlov V, Anthony L, Burgess RR, Landick R. RNA polymerases from *Bacillus subtilis* and *Escherichia coli* differ in recognition of regulatory signals in vitro. Journal of Bacteriology 2000;182:6027–35.

Atwell S, Ultsch M, De Vos AM, Wells JA. Structural plasticity in a remodeled protein–protein interface. Science 1997;278:1125–8.

Berg HC. The rotary motor of bacterial flagella. Annual Review of Biochemistry 2003;72:19–54.

Brar LK, Rajdev P, Raychaudhuri AK, Chatterji D. Langmuir monolayer as a tool toward visualization of a specific DNA–protein complex. Langmuir 2005;21:10671–5.

Carpousis AJ. The RNA degradosome of *Escherichia coli*: an mRNA-degrading machine assembled on RNaseR. Annual Review of Microbiology 2007;61:71–87.

Delarue M. Amino acyl-tRNA synthetases. Current Opinion in Structural Biology 1995;5:48–55.

Dey A, Verma AK, Chatterji D. Role of an RNA polymerase interacting protein, MsRbpA, from *Mycobacterium smegmatis* in phenotypic tolerance to rifampicin. Microbiology 2010;156:873–83.

Dey A, Verma AK, Chatterji D. Molecular insights into the mechanism of phenotypic tolerance to rifampicin conferred on mycobacterial RNA polymerase by MsRbpA. Microbiology 2011;157:2056–71.

Fryxell KJ. The coevolution of gene family trees. Trends in Genetics 1996;12:364–9.

Fujii K, Saito H, Tomioka H, Mae T, Hosoe K. Mechanism of action of antimycobacterial activity of the new benzoxazinorifamycin KRM-1648. Antimicrobial Agents and Chemotherapy 1995;39:1489–92.

Galan JE, Collmer A. Type III secretion machines: bacterial devices for protein delivery into host cells. Science 1999;284:1322–8.

Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE. Co-evolution of proteins with their interaction partners. Journal of Molecular Biology 2000;299:283–93.

Ishihama A. Subunits of *Escherichia coli* RNA polymerase in function and structure. Journal of Cellular Physiology 1969;74:223–4.

Ishihama A, Ito K. Subunits of RNA polymerase in function and structure. II. Reconstitution of *Escherichia coli* RNA polymerase from isolated subunits. Journal of Molecular Biology 1972;72:111–23.

Ishihama A, Fukuda R. Ito K. Subunits of RNA polymerase in function and structure. IV. Enhancing the role of sigma in the subunit assembly of *Escherichia coli* RNA polymerase. Journal of Molecular Biology 1973;79:127–36.

Jespers L, Lijnen HR, Vanvetswinkel S, Van Hoef B, Brepoels K, Cohen D, De Maeyer M. Guiding a dock mode by phage-display: selection of correlated mutations at the staphylokinase–plasmin interface. Journal of Molecular Biology 1999;290:471–9.

Khorana HG. Bacteriorhodopsin, a membrane protein that uses light to translocate protons. The Journal of Biological Chemistry 1988;263:7439–42.

Kouyama T, Kinosita Jr K, Ikegami A. Structure and function of bacteriorhodopsin. Advances in Biophysics 1988;24:123–75.

Kumar S, Dudley J, Nei M, Tamura K. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Briefings in Bioinformatics 2008;9:299–306.

Lake JA, Skophammer RG, Herbold CW, Servin JA. Genome beginnings: rooting the tree of life. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences 2009;364:2177–85.

Lane WJ, Darst SA. Molecular evolution of multisubunit RNA polymerases: sequence analysis. Journal of Molecular Biology 2010;324:177–92.

Lohrke SM, Nechaev S, Yang H, Severinov K, Jin SJ. Transcriptional activation of *Agrobacterium tumefaciens* virulence gene promoters in *Escherichia coli* requires the *A. tumefaciens* RpoA gene, encoding the alpha subunit of RNA polymerase. Journal of Bacteriology 1999;181:4533–9.

Lowe PA, Hager DA, Burgess RR. Purification and properties of the sigma subunit of *Escherichia coli* DNA-dependent RNA polymerase. Biochemistry 1979;18:1344–52.

Marcaida MJ, DePristo MA, Chandran V, Carpousis AJ, Luisi BF. The RNA degradosome: life in the fast lane of adaptive molecular evolution. Trends in Biochemical Sciences 2006;31:359–65.

Mencía M, Monsalve M, Rojo F, Salas M. Substitution of the C-terminal domain of the *Escherichia coli* RNA polymerase alpha subunit by that from *Bacillus subtilis* makes the enzyme responsive to a *Bacillus subtilis* transcriptional activator. Journal of Molecular Biology 1998;275:177–85.

Moyle WR, Campbell RK, Myers RV, Bernard MP, Han Y, Wang X. Co-evolution of ligand-receptor pairs. Nature 1994;402:83–6.

Mukherjee R, Chatterji D. Stationary phase induced alterations in mycobacterial RNA polymerase assembly: a cue to its phenotypic resistance towards rifampicin. Biochemical and Biophysical Research Communications 2008;369:899–904.

Newell KV, Thomas DP, Brekasis D, Paget MSB. The RNA polymerase binding protein, RbpA, confers basal levels of rifampicin resistance on *Streptomyces coelicolor*. Molecular Microbiology 2006;60:687–96.

Nomura M. Assembly of bacterial ribosomes. Science 1973;179:864–73.

Norcum MT, Wolfe CL, Warrington JA. Three-dimensional working model of the multienzyme aminoacyl–tRNA synthetase complex determined by computational microscopy. Microscopy and Microanalysis 2005;11:164–5.

O'Donnell ME, Kornberg A. Dynamics of DNA polymerase III holoenzyme of *Escherichia coli* in replication of a multiprimed template. The Journal of Biological Chemistry 1985;260:12875–83.

Pages S, Belaich A, Belaich JP, Morag E, Lamed R, Shoham Y, Bayer EA. Species-specificity of the cohesion–dockerin interaction between *Clostridium thermocellum* and *Clostridium cellulolyticum*: prediction of the specificity determinant of the dockerin domain. Proteins 1997;29:517–27.

Paget MSB, Molle V, Cohen G, Aharonowitz Y, Buttner MJ. Defining the disulfide stress response in *Streptomyces coelicolor* A3(2): identification of the σR regulon. Molecular Microbiology 2001;42:1007–20.

Pazos F, Valencia A. Protein co-evolution, co-adaptation, and interactions. The EMBO Journal 2008;27:2648–55.

Pazos F, Helmer-Citterich M, Ausiello G, Valencia A. Correlated mutations contains information about protein–protein interaction. Journal of Molecular Biology 1997;271:511–23.

Pijper A. Bacterial flagella and motility. Nature 1948;161:200.

Steffen P, Ullmann A. Hybrid *Bordetella pertussis–Escherichia coli* RNA polymerases: selectivity of promoter activation. Journal of Bacteriology 1998;180:1567–9.

Sykes MT, Williamson JR. A complex assembly landscape of the 30S ribosomal subunit. Annual Review of Biophysics 2009;38:197–215.

Triccas JA, Parish T, Britton WJ, Gicquel B. An inducible expression system permitting the efficient purification of a recombinant antigen from *Mycobacterium smegmatis*. FEMS Microbiology Letters 1998;167:151–6.

van Kesteren RE, Tensen CP, Smit AB, van Minnen J, Kolakowski LF, Meyerhoff W, Richter D, van Heerikhuizen H, Vreugdenhil E, Geraerts WP. Co-evolution of ligand-receptor pairs in the vasopressin/oxytocin superfamily of bioactive peptides. The Journal of Biological Chemistry 1996;271:3619–26.

Yao NY, O'Donnell M. Replisome structure and conformational dynamics underlie fork progression past obstacles. Current Opinion in Cell Biology 2009;21:336–43.

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. Crystal structure of *Thermus aquaticus* core RNAP at 3.3 Å resolution. Cell 1999;98:811–24.