Aromatic Patch in WhiB-Like Transcription Factors Facilitates Primary Sigma Factor Interaction in *Mycobacterium tuberculosis*

Daisy Guiza Beltran[1,#], Tao Wan[1,9,#], Javier Seravalli[1,2], Magdaléna Horová[1], Camden Jones[1], Shanren Li[1,10], Pengchong Ma[3], Chloe Ong[1], Zhifang Lu[1], Donald F Becker[1,2], Jeffrey P. Mower[4], Qiuming Yao[3], Yu Pan[5], Hongfeng Yu[3], Adrie J. C. Steyn[6,7], LiMei Zhang[1,2,8*]

[1]Department of Biochemistry, University of Nebraska-Lincoln, Lincoln, NE, USA

[2]Redox Biology Center, University of Nebraska-Lincoln, Lincoln, NE, USA

[3]School of Computing, University of Nebraska-Lincoln, Lincoln, NE, USA

[4]Department of Agronomy and Horticulture, University of Nebraska-Lincoln, Lincoln, NE, USA

[5]Department of Biological System Engineering, University of Nebraska-Lincoln, Lincoln, NE, USA

[6]Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama, USA

[7]Africa Health Research Institute, University of KwaZulu-Natal, Durban, South Africa

[8]Nebraska Center for Integrated Biomolecular Communication, University of Nebraska-Lincoln, Lincoln, NE, USA

[9]Present address: Biortus Biosciences Co., Ltd, Jiangyin, Jiangsu, 214437, China

[10]Present address: College of Life Sciences, Fujian Normal University, Fuzhou, Fujian, 350117, China

[#]These authors contribute equally

*Correspondence: lzhang30@unl.edu

ABSTRACT

WhiB-like (Wbl) family proteins are a unique family of iron-sulfur ([4Fe-4S]) cluster-bound transcription factors found exclusively in Actinobacteria and actinobacteriophages, including the notoriously persistent pathogen *Mycobacterium tuberculosis* (*Mtb*). Despite their critical roles in cell development, stress response and antibiotic resistance, the mechanisms of gene regulation by the Wbl family proteins are not fully understood due to the lack of a canonical DNA-binding motif in most Wbl proteins. Here, we present structural and biochemical evidence demonstrating that all *Mtb* Wbl proteins bind to the same site in the conserved region 4 of the primary sigma 70 factor facilitated by a previously unrecognized structural motif, the aromatic patch, in the Wbl family. Our phylogenetic findings provide compelling evidence for a complex evolutionary relationship of Wbls between actinobacteria and the associated phages. Together, this work fills a critical gap in our understanding of the function, mechanism and evolutionary origin of Wbls.

Keywords: WhiB-Like (Wbl) family, aromatic patch, primary sigma factor A, iron-sulfur cluster, X-ray crystallography

39 **INTRODUCTION**

40 The *WhiB-like* (*wbl*) proteins are a group of monomeric [4Fe-4S] cluster-containing
41 transcription factors discovered in *Streptomyces coelicolor* (*Sco*) in 1972 and named after the white
42 colonies of the *wbl* mutants due to the disruption of pigment formation (1). Subsequent genetic
43 and bioinformatic analyses find that the Wbl proteins are widely distributed in the phylum
44 Actinobacteria and their associated actinobacteriophages, but not in any other living organisms.
45 Notably, several actinobacterial species of medical and industrial significance encode multiple
46 Wbl paralogues (2-4). For example, seven Wbl members (WhiB1-7) are encoded in *Mtb*, 12 in *Sco*
47 and four in *Corynebacterium glutamicum* (*Cglu*). Wbl proteins play critical roles in diverse
48 biological processes in these actinobacteria, such as cell division and development, responses to
49 redox stress and nutrition starvation, antibiotic resistance and pathogenesis (3,5-21). Because of
50 that, the Wbls proteins have attracted a great deal of attention since their discovery (see recent
51 reviews in (4,22)).

52 Wbl family proteins are characterized by two conserved motifs, the [4Fe-4S]-cluster binding
53 motif composed of four conserved cysteines and the Gly-rich motif (also referred to as the β-turn)
54 (Fig. 1A) (3,23). Based on their sequence similarity, these Wbl proteins are grouped into five major
55 subfamilies, each represented by an *Mtb* Wbl (WhiB1-4, and WhiB7, respectively) and distributed
56 across a wide range of Actinobacterial species (2,4). The rest of the Wbl proteins are more species-
57 specific, such as WhiB5 and WhiB6 found only in Mycobacteria. Several subfamily-specific
58 functional motifs have been reported, most notably the well-characterized "AT-hook" DNA
59 binding motif in the WhiB7 subfamily (24-28). However, most Wbl proteins lack canonical DNA
60 binding motifs, or no other Wbl family-wide motif apart from the two conserved motifs mentioned

61 above has been reported. This knowledge gap has long posed a challenge in understanding how
62 Wbl proteins regulate gene expression.

63 The discovery that WhiB3 requires interaction with the conserved C-terminal region 4 of the
64 primary σ^{70} sigma factor (σ^{A}_{4}) in the RNA polymerase (RNAP) holoenzyme to activate
65 transcription marks a milestone in deciphering the mechanism of transcriptional regulation by
66 Wbls (6). σ^{A}_{4} is highly conserved in the bacterial σ^{70}-family primary sigma factors, responsible
67 for recognizing the -35 promoter element and commonly used as an anchor for many
68 transcription factors for gene activation (29-31). The subsequent studies show that all the *Mtb*
69 Wbls, except for WhiB5, and two Wbl homologs in Streptomyces also bind to σ^{A}_{4} (14,31-33),
70 indicating a molecular mechanism shared by multiple Wbl members across different species.

71 Recent structural characterization of four Wbl proteins by our group and others reveals that,
72 despite the low sequence identity, these Wbl proteins share a similar helical architecture
73 surrounding the [4Fe-4S] binding pocket and bind to the same site on σ^{A}_{4} centered on the H516-
74 P517 pair (Fig. S1A and B) (26-28,34,35). Notably, the hydrophobic interactions-driven
75 molecular interface within the Fe-S cluster binding pocket between these monomeric Wbls and
76 σ^{A}_{4} is distinct from other previously characterized σ^{A}_{4}-dependent transcription factors (*i.e.*, the
77 cyclic AMP receptor protein [CRP] and the fumarate and nitrate reduction regulatory protein
78 [FNR]), which are dimeric and interact with σ^{70}_{4} via surface charge complementarity (6,14,36).
79 Further mechanistic investigations of *Mtb* WhiB7 and *Sve* WhiB2 show that WhiB7 activates gene
80 expression by binding to σ^{A}_{4} and specifically recognizing the target promoter sequence; while
81 WhiB2 co-activates gene expression with another transcription factor, WhiA (26,27,34). However,
82 a comprehensive analysis of the shared motifs and the mode of action in the Wbl family is currently

83 missing.

84 By structural comparison of the reported σ^{A}_{4}-bound Wbl structures (26-28,32,34,35), including
85 three Wbls (WhiB1, WhiB3 and WhiB7) from *Mtb* and the WhiB2 homolog from *Streptomyces*
86 *venezuelae* (*Sve*), we found that Wbls binding to σ^{A}_{4} is facilitated by a conserved patch of aromatic
87 residues (herein referred to as the aromatic patch) in all four complexes, including one residue at
88 the beginning of the helix *h1* (corresponding to W3 in *Mtb* WhiB1, labeled as aromatic patch 1
89 [AP1]), two in the end of helix *h1* (corresponding to F17 and F18 in *Mtb* WhiB1 and labeled as
90 AP2-3), and one in the β-turn connecting *h3* and *h4* (corresponding to W60 in *Mtb* WhiB1 and
91 labeled as AP4) (Fig. 1A, Supplementary Fig. 1). The residues in the aromatic patch are conserved
92 in each Wbl subfamily and at least three of them (including AP2-3) are critical for σ^{A}_{4} binding and
93 Wbl function (Fig. 1A, Supplementary Fig. 2) (26,28,35). In this study, we use multidisciplinary
94 approaches to test whether the aromatic patch is a shared motif in the Wbl family for σ^{A}_{4} binding.
95 We present compelling structural, molecular and biochemical evidence demonstrating that the
96 remaining *Mtb* Wbls (WhiB4, WhiB5 and WhiB6) also bind to the same site on σ^{A}_{4}, and the Wbl:
97 σ^{A}_{4} interaction is facilitated by the aromatic patch. Moreover, our structural modeling,
98 phylogenetic analysis and biochemical characterization of Wbl proteins from diverse
99 Actinobacteria and actinobacteriophages show that this aromatic patch is widely conserved among
100 Wbl family members and uncover complex evolutionary relationships among bacterial and phage
101 Wbl gene family members. Together, our findings identify a previously unrecognized structural
102 feature that underlies a shared molecular mechanism of action in the Wbl family and offer new
103 insights into the evolution and function of this unique family of [4Fe-4S] transcription factors.

104 **RESULTS**

105 ***Aromatic patch at the molecular interface of the WhiB6:$\sigma^A{}_4$ complex***

106 WhiB6 is a Mycobacteria-specific Wbl protein. It regulates the expression of the Type VII
107 secretion systems (ESX-1 and ESX-4) and plays an important role in the reactivation of *Mtb* and
108 other pathogenic mycobacterial species (16,37-41). Compared to other *Mtb* Wbls, WhiB6 lacks
109 two equivalent residues in the aromatic patch, and an unexpected polar residue is found
110 corresponding to AP3, which has been shown to be essential for $\sigma^A{}_4$ binding in all the *Mtb* Wbls
111 that we have tested thus far (Fig. 1A; Supplementary Fig. 2) (26,28,35). Due to the lack of any 3D
112 structural information for either WhiB6 alone or in complex with $\sigma^A{}_4$, it is unclear how WhiB6
113 interacts with $\sigma^A{}_4$ for transcriptional regulation. To address this knowledge gap, we purified and
114 crystallized the $\sigma^A{}_4$-bound WhiB6 to determine the crystal structure of the complex. As previously
115 described, we used the chimera protein of $\sigma^A{}_4$ fused to the RNAP β-subunit flap tip helix (β_{tip}) via
116 an artificial linker (hereby denoted $\sigma^A{}_4$-β_{tip}) to mimic the interaction between σ^A and the β-subunit
117 in the RNAP holoenzyme (see in Materials and Methods, Supplementary Tables S1) (26). The
118 crystal structure of WhiB6:$\sigma^A{}_4$-β_{tip} was refined at 1.8 Å (Supplementary Tables S2).

119 As observed in the other Wbl:$\sigma^A{}_4$ complexes (26-28,34,35), the overall architecture of
120 WhiB6:$\sigma^A{}_4$ is dominated by alpha helices, and the molecular interface between $\sigma^A{}_4$ and WhiB6 is
121 centered around the [4Fe-4S] cluster binding pocket (Fig. 1B). The core helical arrangement of
122 WhiB6 surrounding the Fe-S cluster is surprisingly similar to that of other Wbl:$\sigma^A{}_4$ complexes
123 despite the low sequence identity (<30%). A structural overlay of $\sigma^A{}_4$–bound WhiB6 with WhiB1
124 shows a Cα root mean square deviation ($RMSD_{C\alpha}$) of 1.07 Å for the 37 aligned residues located
125 in helices *h1-3* of WhiB1 (Fig. 1C). In contrast, striking differences are observed in the rest of the
126 structural arrangements that are away from the Fe-S cluster binding pocket, in particular, WhiB6

127 has a significantly shorter helix *h2* and loop *l1* between *h1* and *h2*, but a much longer helix *h4*
128 compared to WhiB1.

129 At the molecular interface of WhiB6:$\sigma^A{}_4$ complex, H516 and P517 of $\sigma^A{}_4$ interact with W21
130 and W56 of WhiB6 (corresponding to F17 [AP2] and W60 [AP4] of WhiB1) within the Fe-S
131 cluster binding pocket of the complex (Fig. 1D). Additionally, a conserved Arg, R38, in the WhiB6
132 subfamily caps the Fe-S cluster pocket in the complex. R38 may serve as a functional equivalent
133 of W3 in WhiB1 based on the structural comparison of the two complexes. As previously observed
134 in other Wbl proteins, substitution of either H516 in $\sigma^A{}_4$ or W21 in WhiB6 by an Ala abolishes the
135 complex formation in the pull-down assays (Fig. 1E and F), highlighting the essential role of the
136 aromatic patch in the formation of WhiB6:$\sigma^A{}_4$. We note that the W21A mutation in WhiB6 disrupts
137 the stability of the Fe-S cluster in WhiB6, as evidenced by a decrease in the absorption at ~410 nm
138 and the presence of the absorption peak at ~350 nm when compared to the wildtype WhiB6. T22
139 of WhiB6 in loop *l1*, which corresponds to F18 of WhiB1 (Fig. 1D), is close to the molecular
140 interface of the WhiB6:$\sigma^A{}_4$ complex within the [4Fe-4S] cluster binding pocket with the hydroxyl
141 group of T22 facing away from the [4Fe-4S] cluster. Substitution of T22 with an Ala did not affect
142 $\sigma^A{}_4$ binding in the pull-down assay, confirming a non-essential role of WhiB6 T22 for $\sigma^A{}_4$ binding
143 and consistent with high sequence variability at this position among the WhiB6 homologs
144 (Supplementary Fig. 1).

145 ***WhiB5 binding to $\sigma^A{}_4$ is H516-dependent and facilitated by the aromatic patch***

146 Unlike WhiB6, sequence analysis shows three of the aromatic patch residues are found in the
147 WhiB5 subfamily, despite the low sequence identity (<30%) between WhiB5 and other *Mtb* Wbl

148 proteins (Fig. 1A; Supplementary Fig. 2). Because of the observation, it was puzzling to note that
149 WhiB5 is the only *Mtb* Wbl that did not show σ^A binding in the *in vitro* and *in vivo* binding assays
150 by Feng *et al.* (31). However, as the authors pointed out, since the *in vitro* pull-down assays were
151 carried out under aerobic conditions, oxidation damage of the [4Fe-4S] cluster in WhiB5 may also
152 abolish σ^A binding. Therefore, we revisited the question of whether WhiB5 binds to $\sigma^A{}_4$. The
153 results from our co-expression from *E. coli* and pull-down assays under anaerobic conditions
154 confirm that WhiB5 binds to $\sigma^A{}_4$ and the binding requires $\sigma^A{}_4$ H516 (Fig. 2A), supporting that
155 WhiB5 binds to the same site on $\sigma^A{}_4$ as other *Mtb* Wbls.

156 To gain structural insights into the interaction between WhiB5 and $\sigma^A{}_4$ and into the role of the
157 aromatic patch residues in the interaction, we first attempted crystallographic characterization of
158 WhiB5:$\sigma^A{}_4$ but were unable to produce crystals with this protein complex despite exhaustive trials.
159 We therefore applied structural modeling approaches complemented by experimental validation
160 to test the effects of the aromatic patch residues on WhiB5 binding to $\sigma^A{}_4$. both the homologous
161 model of WhiB5:$\sigma^A{}_4$ by SWISS-MODEL and the *ab initio* model by AlphaFold3 support that
162 WhiB5 adopts a similar helical arrangement as the $\sigma^A{}_4$-bound WhiB1 surrounding the cluster
163 binding pocket (see in the Materials and Methods, Fig. 2A and Supplementary Fig. 4). The
164 conserved aromatic patch residues in the WhiB5 subfamily align well with those in the WhiB1 at
165 the molecular interface within the Fe-S binding pocket, implying a similar role in $\sigma^A{}_4$ binding.
166 Consistent with the structural prediction, substitution of either W12 or F13 (corresponding to AP2
167 and AP3, respectively) by Ala significantly affects the cluster stability in WhiB5 and its binding
168 to $\sigma^A{}_4$ in comparison to the wildtype WhiB5 in the pull-down assays (Fig. 2C-D). We note that a
169 W12A or F13A mutation disrupts the stability of the Fe-S cluster in WhiB5, and we were not able

170 to purify the mutant proteins in the holo-form. Low levels of WhiB5 mutants were detected by
171 SDS PAGE analysis from the affinity pull-down samples using a His_6-tagged $\sigma^A{}_4$ as the bait (Fig.
172 2C) but not in the samples using a Strep-tagged $\sigma^A{}_4$ as the bait (Fig. 2D), suggesting that the
173 residual WhiB5 mutants retained in the pull-down samples from the NiNTA column are likely due
174 to non-specific binding.

175 To assess whether WhiB5 binding to $\sigma^A{}_4$ is physiologically relevant, we examined the
176 interaction between WhiB5 and $\sigma^A{}_4$ using the mycobacterial two-hybrid system assay termed
177 mycobacterial protein fragment complementation (M-PFC) assay. The M-PFC assay is based on
178 the functional reconstitution of two murine dihydrofolate reductase (mDHFR) domains,
179 $mDHFR_{F[1,2]}$ and $mDHFR_{F[3]}$, respectively, by fusing the two domains independently to two
180 interacting proteins. When the proteins of interest interact, mDHFR activity is restored, conferring
181 bacterial resistance to trimethoprim (TRIM), an antibiotic targeting bacterial DHFR (42). The M-
182 PFC assay was also used in the study by Feng *et al.* (31), in which $mDHFR_{F[1,2]}$ was fused to the
183 N-terminus of WhiB5. It is noted that WhiB5 has an unusually short sequence in the N-terminus
184 prior to the first Fe-S cluster-ligating cysteine residue compared to other *Mtb* Wbl proteins (Fig.
185 1A). Indeed, we found that an N-terminal His-tag, but not a C-terminal tag, results in significant
186 formation of inclusion bodies when over-expressing the recombinant WhiB5 protein in *E. coli*
187 (data not shown). To avoid possible interference of WhiB5 folding and to stabilize the
188 conformation of $\sigma^A{}_4$, we constructed the two mDHFR fragments to the C-terminus of WhiB5 and
189 $\sigma^A{}_4$-β_{tip} (denoted WhiB5-$mDHFR_{F[1,2]}$ and $\sigma^A{}_4$-$mDHFR_{F[3]}$, respectively) in our M-PFC assay (see
190 in Materials and Methods). As shown in Fig. 2E, all the strains grow similarly on the 7H10 agar
191 plates without TRIM. In the presence of TRIM, the *Mycobacterium smegmatis* (*Msm*) strain

192 expressing WhiB5-mDHFR$_{F[1,2]}$ and σ^{A}_{4}-mDHFR$_{F[3]}$, but not the negative control, grows on the
193 TRIM agar plate comparably to the positive control. Consistent with the results from the pull-
194 down assays (Fig. 2D), the *Msm* strain co-expressing σ^{A}_{4}-mDHFR$_{F[3]}$ with either WhiB5 mutant
195 (W12A or F13A)-mDHFR$_{F[1,2]}$ shows significant defects in cell growth on the TRIM agar plate
196 when compared to the wildtype WhiB5. These observations cement the physiological relevance of
197 WhiB5 binding to σ^{A}_{4} and underscore the essential role of the aromatic patch in WhiB5 for σ^{A}_{4}
198 binding.

199 Because both the *in vitro* and *in vivo* protein-protein binding assays were carried out with
200 overexpressed proteins, we used isothermal titration calorimetry (ITC) to quantify WhiB5 binding
201 to σ^{A}_{4} under anaerobic conditions and compared the results with those of WhiB1 and WhiB2.
202 WhiB1 and WhiB2 share the same aromatic patch compositions, making them good references for
203 assessing the impact of the missing AP1 residue in WhiB5 on its binding affinity to σ^{A}_{4}. As shown
204 in Fig. 3, all three Wbl proteins bind to σ^{A}_{4} with low-nanomolar dissociation constants (K_d) and
205 the binding reactions are exothermic, which agrees favorably with the aromatic residues-driven
206 interactions at the molecular interface of the Wbl:σ^{A}_{4} complex and provides additional evidence
207 supporting WhiB5 binding to σ^{A}_{4} under physiologically relevant conditions. We note that the
208 binding affinity of WhiB5 to σ^{A}_{4} is about two- to three-fold lower than that of WhiB1 or WhiB2,
209 consistent with the absence of a key aromatic patch residue (AP1) in WhiB5 for σ^{A}_{4} binding (Fig.
210 1A).

211 ***Aromatic patch is required for WhiB4 binding to σ^{A}_{4}***

212 Previous *in vitro* and *in vivo* protein–protein binding assays have shown that WhiB4 binds to

213 σ^{A}_{4}, but the contribution of the aromatic patch residues to σ^{A}_{4} binding has not been examined. Both
214 the homology model and the AlphaFold model of WhiB4:σ^{A}_{4} in our analysis predict that WhiB4
215 adapts a fold surrounding the Fe-S cluster resembling that of WhiB1, with the interaction interface
216 centered on the H516-P517 pair of σ^{A}_{4} (Supplementary Fig. 6). The aromatic patch residues in
217 WhiB4 closely align with those of WhiB1, except for a leucine residue (L45) corresponding to
218 AP2 (F17 in WhiB1) in the 3D structural alignment (Supplementary Fig. 6A).

219 Consistent with the structural prediction, our pull-down assays indicate that WhiB4 binding to
220 σ^{A}_{4} requires H516, as substitution of H516 by an Ala in σ^{A}_{4} abolishes WhiB4 binding (Fig. 4A).
221 Similarly, mutation of F46 (corresponding to F18 in WhiB1) to Ala in WhiB4 abolishes σ^{A}_{4}
222 binding and drastically destabilizes the WhiB4 [4Fe-4S] cluster, evident by a sharp decrease in the
223 410-nm absorbance (Fig. 4B). These observations underline the critical roles the aromatic patch
224 residues in both maintaining [4Fe-4S] cluster integrity and binding to σ^{A}_{4}, as observed for other
225 Wbls (Fig. 1F) (26,35). Notably, an Ala substitution of non-aromatic residue L45 also abolishes
226 WhiB4 binding to σ^{A}_{4}. This observation differs from that of T22 in WhiB6 (Fig. 1E), but aligns
227 well with the sequence analysis showing L45 is highly conserved in the WhiB4 subfamily and the
228 structural models predicting that the side chain of L45 contributes directly to the WhiB4:σ^{A}_{4}
229 interface (Supplementary Figs. 2 and 6A).

230 ***Aromatic patch facilitating Mtb Wbl binding to σ^{A}_{4} is widely conserved in the Wbl family***

231 To date, over 35,000 Wbl sequences have been deposited in the Wbl family database (Pfam
232 ID: PF02467; InterPro ID: IPR034768). Since five *Mtb* Wbl proteins are well distributed in
233 Actinobacteria and σ^{A}_{4} is highly conserved in Actinobacteria, we speculate that the aromatic patch

234 facilitating *Mtb* Wbls' binding to $\sigma^A{}_4$ may be found in many other Wbl family proteins. To test
235 this hypothesis, we analyzed 995 Wbl sequences from the representative proteomes of
236 Actinobacteria and actinobacteriophages with a co-membership threshold of 15% (RP15), as
237 defined in the Pfam database (Pfam ID: PF02467) (see in Materials and Methods). These Wbl
238 proteins, of which 55% are from Actinobacteria and 45% are from Actinobacteriophages, were
239 classified into 29 subfamilies (Fig. 5, Supplementary DataSet 1).

240 Our phylogenetic analysis confirms that five *Mtb* Wbls (WhiB1–4 and WhiB7) represent the
241 major Wbl subfamilies and reveals an intricate evolutionary relationship between bacterial Wbls
242 and those from actinobacteriophages. As shown in Fig. 5, the five major Wbl subfamilies, marked
243 by the representative *Mtb* Wbls, account for ~80% of the total sequences from diverse
244 actinobacterial species and associated phages, which is in line with the previous bioinformatic
245 analysis (2). Among them, the WhiB2 subfamily is the largest Wbl subfamily and is found most
246 widely distributed in Actinobacteria (2). Our analysis also shows that the WhiB2 subfamily has a
247 broad mix of members from Actinobacteria (33%) and actinobacteriophages (67%), which include
248 the WhiB2 homolog from Mycobacteriophage TM4 that has been shown to inhibit the function of
249 WhiB2 in *Mycobacterium smegmatis* (*Msm*) (43). The other four major subfamilies are dominated
250 by members from Actinobacteria. The relatively fewer actinobacteriophage sequences in these
251 subfamilies are often located within the earliest diverging clades (highlighted by red arrowheads
252 in Fig. 5), while the bacterial clades occupy more derived positions. This pattern suggests that
253 many of these homologs were transferred from phages to bacteria, supporting the notion that the
254 phages may have served as vectors for the horizontal gene transfer (HGT) of Wbl proteins among
255 Actinobacteria as proposed in the previous studies (2,25,44). Conversely, the phylogenetic analysis

256 also identifies some phage Wbl members nested within otherwise bacterial clades (red stars in Fig.
257 5), providing clear examples of HGT from bacteria to phage and underscoring the complex
258 interplay of HGT in spreading *wbl* genes between Actinobacteria and their related phages. The 24
259 remaining subfamilies are much smaller in size, and the majority (19 out of 24) of them are
260 dominated by Wbls from actinobacteriophages. The two remaining *Mtb* Wbls (WhiB5 and WhiB6)
261 are grouped together with several unique Wbls from actinobacteriophages into one subfamily
262 (labeled as No. 29 in Supplementary DataSet 1, Fig. 5), which is otherwise closely related with the
263 WhiB7 subfamily.

264 As we speculated, sequence alignment of the 995 representative Wbl sequences indicates the
265 aromatic patch residues in *Mtb* Wbls for $\sigma^A{}_4$ binding is well conserved among the Wbl homologs,
266 together with the two known motifs (the four [4Fe-4S] cluster-ligating Cys and the β-turn) (Fig.
267 6A). Considering the low sequence identity among the Wbl members, we generated AlphaFold
268 structural models for these Wbl members to better quantify the conservation of the aromatic patch
269 in the Wbl family by 3D structural alignment against $\sigma^A{}_4$-bound WhiB1 (PDB ID:6ONO) (see in
270 Materials and Methods). Our results show that, among the 995 sequences examined, 984 out of
271 the 995 Wbls (98.7%) possess at least two aromatic residues (either Trp, Phe, Tyr or His)
272 corresponding to AP1-4 and at least one corresponding to AP2-3 (Fig. 6B, Supplementary DataSet
273 1). Moreover, eight out of the 11 outliers possess an aromatic patch similar to that observed in the
274 *Mtb* Wbl proteins when considering the residues next to the aromatic patch at the relative -1 to 1
275 position, with the remaining three lacking an aromatic patch as defined (Supplementary Fig. 8A-
276 C). These findings indicate the newly identified aromatic patch is widely conserved in the Wbl
277 family. During the structural comparison, we also found three Wbls with only three Cys at the

278 expected [4Fe-4S] cluster binding pocket in the AlphaFold models. Two of these Wbl sequences
279 (UniProt IDs: A0A1Q9SLC3 and A0A1Q9SWN8) are obsolete, while the third one (UniProt ID:
280 A0A1D8EX64) from *Mycobacterium phage Tortellini* misses a Cys corresponding to the first Fe-
281 S cluster-ligating Cys (corresponding to C9 in WhiB1) (Supplementary Fig. 8D). Blast search
282 shows that A0A1D8EX64 shares 99% sequence identity with another Wbl protein (UniProt ID:
283 A0A2U8UHK0) from *Mycobacterium phage Xavia*, which has an additional Cys in the N-terminus
284 corresponding to C9 of WhiB1. Therefore, it is unclear whether these Wbl outliers are due to the
285 uncertainty of gene sequencing or structural prediction, or they represent true evolutionary variants
286 of the Wbl family.

287 To provide experimental evidence that Wbls from actinobacteriophages also bind to $\sigma^A{}_4$ in a
288 manner similar to *Mtb* Wbls, we randomly selected three phage Wbls (UniProt IDs:
289 A0A0K2FNL9, A0A3G3M9S7 and Q857R7) from the WhiB1 subfamily and two other relatively
290 larger species-specific Wbl subfamilies (Fig. 5), and examined their interactions with $\sigma^A{}_4$. The
291 pull-down assays confirm that, like the *Mtb* Wbls, all three phage Wbls bind to $\sigma^A{}_4$ in a His516-
292 dependent manner (Supplementary Fig. 9). Moreover, as previously observed in *Mtb Wbls*, the
293 aromatic patches in these phage Wbls vary in residue composition and differ in contribution to $\sigma^A{}_4$
294 binding and the [4Fe-4S] cluster stability. For example, Q857R7 belongs to the WhiB1 subfamily
295 with an aromatic patch closely resembling that of WhiB1 (Fig. 5, Supplementary Fig. 10A).
296 Consistently, the mutagenesis and pull-down assays show that the two aromatic patch residues F20
297 and F21 (corresponding to AP2 and AP3) are essential for $\sigma^A{}_4$ binding but do not affect the [4Fe-
298 4S] cluster stability, mirroring the roles of F17 and F18 in WhiB1 (Supplementary Fig. 10B and
299 C). In contrast, A0A3G3M9S7 shows low sequence identity (<25%) with *Mtb* WhiB1 and the

300 AlphaFold model of σ^{A}_{4}-bound A0A3G3M9S7 does not align well with WhiB1 ($RMSD_{C\alpha}$ of 2.68
301 Å for the 44 aligned residues; Supplementary Fig. 11A). Nonetheless, sequence alignment and
302 homology modeling indicate that A0A3G3M9S7 retains three of the four WhiB1 aromatic patch
303 residues (Supplementary Fig. 11B), with a lack of AP2 similar to WhiB4. Mutating either W12 or
304 Y26 in A0A3G3M9S7 not only abolishes σ^{A}_{4} binding but also destabilizes the [4Fe-4S] cluster as
305 well as overall protein folding (Supplementary Fig. 11C-F). A0A0K2FNL9, on the other hand,
306 contains three aromatic patch residues (W3, F17 or F18) aligning with AP1-3 in WhiB1 as well as
307 two additional A0A3G3M9S7-spcific aromatic patch residues (W4 and F58) at the interaction
308 interface of the A0A3G3M9S7:σ^{A}_{4} model (Supplementary Fig. 12A). A single mutation of either
309 AP2 or AP3 in A0A3G3M9S7 (F17A or F18A) or a double mutation of the A0A3G3M9S7-
310 specific aromatic patch residues (W4A-F58A) has little effect on σ^{A}_{4} binding in the pull-down
311 assays (Supplementary Fig. 12B and C). However, a triple mutation (W4A-F17A-F58A)
312 completely abolishes σ^{A}_{4} binding and destabilizes the [4Fe-4S] cluster in A0A3G3M9S7. These
313 observations parallel our previous finding that the WhiB3 subfamily-specific aromatic patch
314 residue W15 compensates for the loss of the canonical aromatic patch residue AP1 (W17) in σ^{A}_{4}
315 binding (Supplementary Fig. 2) (28). Collectively, these results provide compelling experimental
316 evidence that these phage Wbls bind to the same site on σ^{A}_{4} facilitated by the aromatic patch and
317 expand our knowledge of the diversity and adaptability of the aromatic patch in Wbl family
318 proteins.

319 **Discussion**

320 Wbl family proteins have gained increasing interest in the past decade because of their critical
321 roles in the many biological processes in Actinobacteria that are of significant biomedical and

322 economic implications, and their unique mode of action as the first family of monomeric Fe-S
323 transcription factors. However, the mechanism of action of the Wbl family proteins remains
324 enigmatic due to the lack of knowledge on the conserved functional motifs other than the [4Fe-4S]
325 cluster binding motif and the β-turn. Herein, we apply multidisciplinary approaches to uncover a
326 previously unrecognized motif, aromatic patch, in the Wbl family that facilitates $\sigma^A{}_4$ binding to
327 regulate the genes involved in diverse biological processes and provide invaluable insights into
328 the evolutionary history of the Wbl proteins among actinobacteria and associated phages. Multiple
329 Wbl paralogs are present in *Mtb* and many other actinobacterial species, where they play nonredundant
330 roles (4,22). Previous studies have shown that these paralogs are regulated at both the transcriptional and
331 protein levels (8,12,45). Our earlier studies on WhiB1 and WhiB7 have demonstrated that the absence of
332 an equivalent aromatic patch (AP) residue increases the oxygen sensitivity of the Wbl Fe–S cluster (26,35).
333 In this study, we further show that the lack of an equivalent aromatic patch residue (AP1) in WhiB5
334 correlates with its lower binding affinity to $\sigma^A{}_4$. Together, these findings support an additional mechanism
335 for modulating Wbl activities by fine-tuning the Fe-S reactivity and Wbl:$\sigma^A{}_4$ interaction.

336 The unusual hydrophobic residue-driven, tight interaction between Wbl proteins and $\sigma^A{}_4$ is in
337 stark contrast to the previously characterized $\sigma^{70}{}_4$-dependent transcription activators that interact
338 with $\sigma^{70}{}_4$ via surface charge complementarity (6,14,36). Rather, its mode of $\sigma^A{}_4$ binding resembles
339 that of the "phage appropriator" AsiA of the bacteriophage T4, which hijacks the host RNAP via
340 binding to $\sigma^{70}{}_4$ and co-activates phage genes with another phage regulator MotA (46-48).
341 Consistently, the phylogenetic analysis by us and other groups reveals a complex evolutionary
342 relationship of Wbl homologs between actinobacteria and their associated phages. The *wbl* genes
343 are believed to have been laterally distributed in Actinobacteria (44), while their origin remains

344 unknown. Broad analyses of the Wbl family proteins indicate that some members are widespread
345 among Actinobacteria, whereas others are more sporadically distributed (2). The variability in the
346 distribution of multiple Wbl paralogs suggests that HGT may have played a role in their spread,
347 and these transfer events were possibly mediated by phages, which also harbor Wbl genes
348 (2,25,44). In fact, the Wbl family proteins were found to be among the most abundant regulator
349 proteins encoded by actinobacteriophage genomes, and phylogenetic analysis revealed a complex
350 evolutionary relationship of Wbl homologs between actinobacteria and their associated phages and
351 plasmids, supporting a history of HGT (49), although the low sampling of phage homologs limits
352 strong conclusions about the directionality of HGT. To further explore the evolutionary
353 relationships among the bacterial and phage Wbl gene family members, our phylogenetic analysis
354 included many additional phage Wbl sequences. Like previous studies, our tree shows a complex
355 clustering of phage and bacterial Wbl homologs, lending further support for HGT of homologs
356 between phage and bacteria. This finding implies that bacterial hosts may have sequestered the
357 *wbl* gene from actinobacteriophages and repurposed the function of the genes for their own
358 benefits during the early evolutionary events, followed by redistribution of the *wbl* genes among
359 the Actinobacteria via phage-mediated HGT, in line with the previous hypothesis by Weinbauer *et*
360 *al.* (50).

361 The results from this work and previous studies show that most Wbl characterized thus far,
362 except for WhiB7, lack a recognizable DNA binding motif and do not bind to a specific target
363 DNA sequence despite that they all bind to the same site on $\sigma^A{}_4$. These findings raise a new
364 question about how Wbls regulate gene expression via binding to σ^A in the RNAP. Recent studies
365 on the WhiB2 homolog from *Sve* offer a working model for how other Wbl proteins without a

366 canonical DNA binding motif may regulate gene expression via binding to σ^A (15,34,51). These
367 studies show that *Sve* WhiB co-activates the genes involved in cell division and sporulation via
368 binding to σ^A in the RNAP and another regulator WhiA. The single-particle cryo-electron
369 microscopy (cryoEM) structure of the WhiA/WhiB-dependent transcription initiation complex
370 shows that *Sve* WhiB acts as a mediator that binds to WhiA on one side and σ^A on the other side
371 (34). The WhiA and $\sigma^A{}_4$ in the RNAP, joined by WhiB2, define promoter specificity of the
372 WhiA/WhiB2-dependent genes. *Sve* WhiB only shows non-specific DNA contacts, explaining the
373 absence of the recognizable DNA binding motif in the WhiB2 subfamily. Further molecular and
374 structural characterization of the Wbl-dependent transcriptional machinery is needed for an in-
375 depth understanding of the diverse roles of Wbl proteins in stress-responsive transcriptional
376 regulation.

377 **Methods**

378 ***Bacterial strains, plasmid, and growth conditions***

379 The bacterial strains and plasmids used in this study are listed in supplementary Table S1. All
380 the plasmids were confirmed by DNA sequencing before being transformed into *E. coli* BL21-
381 Gold (DE3) strain for protein expression.

382 All *E. coli* strains were grown in Luria–Bertani (LB) media and at 37 °C, 200 rpm, unless
383 otherwise specified. *Mycobacterium smegmatis* MC^2 155 (*Msm*) and the related strains were grown
384 in either Middlebrook 7H9 broth or on 7H10 agar (BD Difco™) supplemented with 10% v/v ADS
385 (2% dextrose, 5% bovine serum albumin and 0.85% NaCl), 0.2% v/v glycerol and 0.05% v/v
386 Tween 80 (Sigma). When appropriate, the media were supplemented with antibiotics at the

387 following concentrations: ampicillin, 100 μg/ml; hygromycin, 100 μg/ml for *E. coli* and 50 μg/ml
388 for *Msm*; kanamycin, 50 μg/ml for *E. coli* and 25 μg/ml for *Msm*; spectinomycin, 100 μg/ml for
389 *E. coli*.

390 ***General Procedures for protein overexpression and purification***

391 The overexpression and purification of the proteins of interest in this study have been
392 previously described unless otherwise stated (35). In brief, the *E. coli* BL21-Gold (DE3) strains
393 containing the desirable plasmids were used to overexpress the proteins studied in this manuscript.
394 Induction was initiated with the addition of 100 μM isopropyl β-d-thiogalactopyranoside (IPTG).
395 For protein constructs containing a [4Fe-4S] cluster, 100 μM ferric ammonium sulfate was added
396 before the induction.

397 The Wbl-containing proteins and complexes with a His-tag were purified using an affinity
398 HisTrap HP column (GE Health Care Life Sciences) in an anaerobic chamber (Coy Labs). For
399 proteins that tend to have high amounts of DNA contamination, an extra wash using a buffer
400 containing 20 mM Tris-HCl, pH 8.0, 1M NaCl, and 1 mM DTT was added during the purification
401 to reduce DNA contamination. Imidazole in the eluted fractions was removed by passing the eluate
402 through either a desalting column or a Superdex 200 column (GE Healthcare Life Sciences). The
403 recombinant $\sigma^A{}_4$-related proteins without a Wbl were purified with procedures similar to those for
404 the purification of the Wbl proteins, except that the purification was carried out under aerobic
405 conditions at 4°C.

406 The purified proteins were quantified by UV-visible (UV-Vis) spectroscopy and Bradford
407 assays. The UV-Vis spectra of the purified proteins were recorded using an HP 8452a diode array

408 UV-visible spectrophotometer (Agilent Technologies Inc.). The absorption at 410 nm,
409 characteristic of proteins containing $[4Fe\text{-}4S]^{2+}$ clusters (52,53), was used to estimate the
410 occupancy of the Fe-S cluster in the protein samples containing a Wbl protein. Protein
411 concentrations were estimated either by the Pierce Bradford Assay Kit (Thermo Fisher Scientific)
412 or UV-visible absorption spectroscopy. The iron content in the purified proteins containing holo-
413 Wbl was quantified as described (54). Unless otherwise specified, all the purified proteins in 50
414 mM Tris-HCl pH 8, 100 mM NaCl, and 1 mM DTT were stored in a Liquid Nitrogen Dewar or -
415 80°C until use.

416 ***Crystallization***

417 The plasmids pET21-MtbWhiB6 and pET28-His_6-$\sigma^A_{C82}\beta_{tip}$ were used to express and purify
418 the WhiB6:σ^A_{C82}-β_{tip} complex. Initial crystallization screens of the WhiB6:σ^A_{C82}-β_{tip} complex were
419 carried out at 18°C in a Coy anaerobic chamber using the sitting-drop vapor diffusion method,
420 followed by optimization of the crystallization hits. High-quality crystals were obtained by mixing
421 1 µl of WhiB6:σ^A_{C82}-β_{tip} at 50 mg/ml in 50 mM Tris-HCl, pH 8.0, 100 mM NaCl and 1 mM DTT
422 with an equal volume of the reservoir solution containing 0.1 M Na Citrate pH 5.5, 20% PEG3000.
423 All the crystals were briefly soaked in the reservoir solution supplemented with 20% glycerol for
424 cryoprotection before flash-frozen in liquid nitrogen.

425 ***X-ray crystallographic data collection, structural determination, and analysis***

426 X-ray diffraction data were collected at the beamline 12-2 of the Stanford Synchrotron
427 Radiation Light source from single crystals maintained at 100 K using a 6M Pixel Array Detector.
428 A single crystal was used to collect both the single anomalous dispersion (SAD) data at the Fe K-

429 edge absorption peak (7140 eV), and the native data for final refinement was collected at 13,000
430 eV. The diffraction data were indexed, integrated, and scaled with HKL2000 (55). The
431 experimental phases were determined by MR-SAD with Phenix.AutoSol using WhiB7:σ^A_{C82}-β_{tip}
432 crystal structure (PDB: 7KUG) as a searching model (26,56). Model building and structure
433 refinement were carried out using COOT and Phenix (57,58). Two copies of the WhiB6: σ^A_{C82}-β_{tip}
434 complex molecules were found in each asymmetric unit. The data collection and refinement
435 statistics are summarized in Table S2.

436 The gene sequence encoding 116-aa *Mtb* WhiB6 reported in the MycoBrowser portal was used
437 to construct the plasmid for overexpression of the WhiB6:σ^A_{C82}-β_{tip} complex and for crystallization
438 (59). We subsequently noticed that the two studies on the transcription analysis indicate the open
439 reading frame of WhiB6 shifts 60 bp downstream, starting at M21 of WhiB6 reported in the
440 MycoBrowser portal (60,61). We therefore renumbered the residue number accordingly.

441 ***Co-expression and in vitro pull-down assays***

442 To test the interaction between the Wbls protein (wildtype or mutant) and σ^A_4 (wildtype or
443 mutant), the two plasmids encoding *Wbls* (wildtype or mutant as specified) and σ^A_{CTD} (wildtype
444 or mutant), respectively, were transformed into *E. coli* BL21-Gold (DE3) for protein co-expression
445 and affinity purification as previously described (62). The plasmids used for the co-expression and
446 pull-down assays are listed in Table S1. A HisTrap™ HP column was used for the pull-down
447 assays using a His-tagged protein as the bait, while a Strep-Tactin®XT 4Flow® column (IBA
448 Lifesciences) was used for the pull-down assay using a Strep-tagged protein as the bait. The
449 purified protein samples were analyzed by UV-Visible spectroscopy and SDS-PAGE.

450 ***Isothermal titration calorimetry (ITC)***

451 For isothermal titration calorimetry (ITC) experiments, the plasmids pET28-6Hisσ^{A}_{C112}-β_{tip},
452 pETDuet-6His-SUMO-MtbWhiB1, pETDuet-6His-SUMO-MtbWhiB2 and pETDuet-6His-
453 SUMO-MtbWhiB5 were used to express and purify 6His-tagged σ^{A}_{C112}-β_{tip} (His_6-σ^{A}_{C112}-β_{tip}) and
454 the Wbl proteins with a N-terminal His_6-SUMO tag, respectively, from *E. coli* as described above.
455 Prior to the experiments, all the proteins were exchanged to the buffer containing 50 mM Tris pH
456 8.0, 100 mM NaCl, and 0.5 mM tris (2-carboxyethyl) phosphine (TCEP) by passing through a
457 desalting column in an anaerobic chamber (Coy Labs).

458 The isothermal titration calorimetry (ITC) experiments were conducted at 25 °C in a VP-ITC
459 calorimeter (MicroCal Malvern Panalytical LLC) inside an anaerobic chamber (Belle
460 Technologies). For each experiment, 1.4 ml of 10 μM His_6-σ^{A}_{C112}-β_{tip} was loaded into the sample
461 cell. Following thermal equilibration, a series of 5-μl 75 μM His_6-SUMO-Wbl was injected into
462 the sample well at intervals of 300 seconds. A 2 μL pre-injection was included at the start of each
463 titration. The reference buffer titrated to was used for background subtraction. The analysis for the
464 ITC was done using MicroCal ITC Origin software.

465 ***Mycobacterial Protein Fragment Complementation (M-PFC) assay***

466 The interaction between WhiB5 and *Mtb* σ^{A}_{4} was analyzed in *Msm* using the M-PFC assay
467 (42). The plasmids used for the M-PFC assays are listed in Table S1. Briefly, the *Saccharomyces*
468 *cerevisiae* leucine-zipper sequence homodimerization domain (GCN4), which was fused to the N
469 terminus of dihydrofolate reductase fragment ($mDHFR_{F[1,2]}$ and $mDHFR_{F[3]}$) in pUAB100 and
470 pUAB200 was replaced by WhiB5 or *Mtb* σ^{A}_{C82}-β_{tip}, respectively, to express WhiB5-Gly_{10}-

471 $mDHFR_{F[1,2]}$ and σ^{A}_{C82}-β_{tip}-Gly_{10}-$mDHFR_{F[3]}$. The resulting plasmids, pUAB100-WhiB5 and
472 pUAB200-σ^{A}_{C82}-β_{tip}, were co-transformed into *Msm* for testing the interaction between WhiB5
473 and σ^{A}_{4} in *Msm*. As previously described (42), the *Msm* strain carrying pUAB100 and pUAB200
474 was used as the positive control, while the strain with pUAB300 and pUAB400 was used as the
475 negative control (42). All the strains were grown in 7H9 broth supplemented with 50 μg/ml of
476 hygromycin and 25 μg/ml of kanamycin until saturation, followed by subculture with initial
477 OD_{600nm} of 0.05, and incubated at 37 °C, 200 rpm until OD_{600nm} of 0.6-1. Each of the samples was
478 diluted to OD_{600nm} of 0.4. 2 μl of the Msm cultures from ten-fold serial dilutions were spotted on
479 7H10 broth agar plates containing 50 μg/ml of hygromycin, 25 μg/ml of kanamycin, and
480 trimethoprim (TRIM) as indicated. The images were taken after five days of incubation at 37 °C.

481 ***Bioinformatic analysis***

482 The 1048 Wbl sequences from the representative proteomes of Actinobacteria and
483 actinobacteriophages with a co-membership threshold of 15% (RP15) were downloaded from the
484 Pfam database (ID: PF02467) (63). We subsequently excluded the disqualified Wbl sequences,
485 including those with less than four Cys residues per the definition of Wbl family or with non-
486 standard amino acid (such as "X"). We also excluded the sequences outside of the range of 75–
487 250 aa that are likely incomplete sequences or contain multiple domains. The resulting 995
488 sequences were used in the following bioinformatic analysis (Supplementary DataSet 1).

489 For the phylogenetic analysis, multiple sequence alignment of the resulting 995 sequences was
490 conducted by ClustalW (64). Fasttree2 and TreeCluster were used for tree structure generation and
491 cluster assignment, respectively (65,66). The classing method (avg clade) in the TreeCluster

492 program was used for the cluster assignment. The five major clusters were named after the Mtb
493 Wbl proteins in the cluster, while the automatically assigned cluster IDs remain for the rest of the
494 clusters.

495 ***3D structural modeling and alignment***

496 AlphaFold2 was used for the structural modeling of the representative Wbl proteins described
497 above, with the template data up to 2022 (67). The top-ranked AlphaFold model for each Wbl
498 sequence was used for 3D structural alignment against WhiB1 (PDB ID: 6ONO) by the PyMol
499 Molecular Graphics System v3.0 (https://pymol.org).

500 The homologous models of the Wbl:$\sigma^A{}_4$ complexes were generated by SWISS-MODEL using
501 WhiB1:$\sigma^A{}_4$ (PDB ID: 6ONO) as a template, while the *ab initio* models of the complexes were
502 generated by AlphaFold3 using the online server (35,68,69).

503 ***Data visualization***

504 The representative homologs for the five major Wbl subfamilies (WhiB1-4 and WhiB7) used
505 for the alignments are modified from the study by *Chandra, et al.* (2), and listed in the legend of
506 Figure S1. The homologs of WhiB5 and WhiB6 are selected from the represented mycobacterial
507 species as noted in the legend of Figure S1. Sequence alignments were performed using Clustal
508 Omega and the ESpript online server (https://espript.ibcp.fr), and the sequence logo was generated
509 using WebLogo (70-72).

510 The phylogenetic tree structure and protein clusters from the bioinformatic analysis were
511 visualized by iTOL (73). 3D structure figures were prepared with the PyMol Molecular Graphics
512 System v3.0 (https://pymol.org).

513 **Data Availability**

514 Atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank
515 (PDB) under the accession code 8D5V for the WhiB6:σ^{A}_{4}-β_{tip} complex, and released for public
516 access with the validation report.

517 The phylogenetic tree with highlights of the Wbl subfamilies and the actinobacteriophages Wbls
518 is available at iTOL.

519 **Supplemental Information**

520 Supplemental documents can be found online at the journal website.

521 **Acknowledgments**

522 The authors thank Dr. Mark A. Wilson at the University of Nebraska-Lincoln for the insightful
523 discussions. We thank the staff at beamlines 12-2 of SSRL for their assistance during the X-ray
524 diffraction data collection. The authors would like to acknowledge the following funding sources:
525 This work was supported by grants from the National Institutes of Health (R35 GM138157) and
526 the National Science Foundation CAREER Award (CLP 1846908 for the work related to WhiB1)
527 to L-M.Z., and the NIH (T32 GM136593) to D.G.B. through the T32 Predoctoral Training
528 Fellowship. The content is solely the responsibility of the authors and does not necessarily
529 represent the official views of the National Institutes of Health and the National Science
530 Foundation. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator
531 Laboratory, is supported by the U.S. Department of Energy, Office of Science, Office of Basic
532 Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular
533 Biology Program is supported by the DOE Office of Biological and Environmental Research, and

534 by the National Institutes of Health, National Institute of General Medical Sciences (including
535 P41GM103393). This work was completed utilizing the Holland Computing Center of the
536 University of Nebraska, which receives support from the UNL Office of Research and Innovation
537 and the Nebraska Research Initiative.

538 **Author contributions**

539 L-M.Z. directed the project. D.G.B., T.W., C. J., S. L., M.H and Z. L. performed molecular
540 biology work, expressed, purified, and characterized the proteins. T.W. carried out the
541 crystallographic study. J.S. and D.G.B. conducted the ITC assay, with assistance from D.B. Y.P.,
542 H.Y, D.G.B. and L-M.Z. carried out the structural modeling and structural analysis. Q.Y., P. M.
543 and C.O. conducted the phylogenetic analysis and tree figure generation, J.P.M. provided expertise
544 in the phylogenetic tree interpretation. A. S. provided the plasmids used for M-PFC assays. D.G.B,
545 T.W, J.P.M. and L-M.Z. wrote the manuscript. All authors contributed to the manuscript
546 preparation.

547 **Conflict of interest**

548 The authors declare no competing interests.

549 **References**

550 1. Chater, K. F. (1972) A morphological and genetic mapping study of white colony
551 mutants of *Streptomyces coelicolor*. *J Gen Microbiol* **72**, 9-28
552 2. Chandra, G., and Chater, K. F. (2014) Developmental biology of Streptomyces from the
553 perspective of 100 actinobacterial genome sequences. *FEMS Microbiol Rev* **38**, 345-379
554 3. Davis, N. K., and Chater, K. F. (1992) The *Streptomyces coelicolor* whiB gene encodes a
555 small transcription factor-like protein dispensable for growth but essential for
556 sporulation. *Mol Gen Genet* **232**, 351-358
557 4. Bush, M. J. (2018) The actinobacterial WhiB-like (Wbl) family of transcription factors.
558 *Mol Microbiol* **110**, 663-676

559 5. Molle, V., Palframan, W. J., Findlay, K. C., and Buttner, M. J. (2000) WhiD and WhiB,
560 homologous proteins required for different stages of sporulation in *Streptomyces*
561 *coelicolor A3(2)*. *J Bacteriol* **182**, 1286-1295
562 6. Steyn, A. J., Collins, D. M., Hondalus, M. K., Jacobs, W. R., Jr., Kawakami, R. P., and
563 Bloom, B. R. (2002) *Mycobacterium tuberculosis* WhiB3 interacts with RpoV to affect
564 host survival but is dispensable for *in vivo* growth. *Proc Natl Acad Sci USA* **99**, 3147-
565 3152
566 7. Morris, R. P., Nguyen, L., Gatfield, J., Visconti, K., Nguyen, K., Schnappinger, D., Ehrt,
567 S., Liu, Y., Heifets, L., Pieters, J., Schoolnik, G., and Thompson, C. J. (2005) Ancestral
568 antibiotic resistance in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci USA* **102**,
569 12200-12205
570 8. Geiman, D. E., Raghunand, T. R., Agarwal, N., and Bishai, W. R. (2006) Differential
571 gene expression in response to exposure to antimycobacterial agents and other stress
572 conditions among seven *Mycobacterium tuberculosis* whiB-like genes. *Antimicrob*
573 *Agents Chemother* **50**, 2836-2841
574 9. Singh, A., Guidry, L., Narasimhulu, K. V., Mai, D., Trombley, J., Redding, K. E., Giles,
575 G. I., Lancaster, J. R., Jr., and Steyn, A. J. (2007) *Mycobacterium tuberculosis* WhiB3
576 responds to O_2 and nitric oxide via its [4Fe-4S] cluster and is essential for nutrient
577 starvation survival. *Proc Natl Acad Sci USA* **104**, 11562-11567
578 10. Smith, L. J., Stapleton, M. R., Fullstone, G. J., Crack, J. C., Thomson, A. J., Le Brun, N.
579 E., Hunt, D. M., Harvey, E., Adinolfi, S., Buxton, R. S., and Green, J. (2010)
580 *Mycobacterium tuberculosis* WhiB1 is an essential DNA-binding protein with a nitric
581 oxide-sensitive iron-sulfur cluster. *Biochem J* **432**, 417-427
582 11. Chawla, M., Parikh, P., Saxena, A., Munshi, M., Mehta, M., Mai, D., Srivastava, A. K.,
583 Narasimhulu, K. V., Redding, K. E., Vashi, N., Kumar, D., Steyn, A. J., and Singh, A.
584 (2012) Mycobacterium tuberculosis WhiB4 regulates oxidative stress response to
585 modulate survival and dissemination in vivo. *Mol Microbiol* **85**, 1148-1165
586 12. Larsson, C., Luna, B., Ammerman, N. C., Maiga, M., Agarwal, N., and Bishai, W. R.
587 (2012) Gene expression of *Mycobacterium tuberculosis* putative transcription factors
588 *whiB1-7* in redox environments. *PLoS One* **7**, e37516
589 13. Casonato, S., Cervantes Sanchez, A., Haruki, H., Rengifo Gonzalez, M., Provvedi, R.,
590 Dainese, E., Jaouen, T., Gola, S., Bini, E., Vicente, M., Johnsson, K., Ghisotti, D., Palu,
591 G., Hernandez-Pando, R., and Manganelli, R. (2012) WhiB5, a transcriptional regulator
592 that contributes to *Mycobacterium tuberculosis* virulence and reactivation. *Infect Immun*
593 **80**, 3132-3144
594 14. Burian, J., Yim, G., Hsing, M., Axerio-Cilies, P., Cherkasov, A., Spiegelman, G. B., and
595 Thompson, C. J. (2013) The mycobacterial antibiotic resistance determinant WhiB7 acts
596 as a transcriptional activator by binding the primary sigma factor SigA (RpoV). *Nucleic*
597 *Acids Res* **41**, 10062-10076
598 15. Bush, M. J., Chandra, G., Bibb, M. J., Findlay, K. C., and Buttner, M. J. (2016) Genome-
599 wide chromatin immunoprecipitation sequencing analysis shows that WhiB is a
600 transcription factor that cocontrols its regulon with WhiA to initiate developmental cell
601 division in *Streptomyces*. *mBio* **7**, e00523-00516

602 16. Chen, Z., Hu, Y., Cumming, B. M., Lu, P., Feng, L., Deng, J., Steyn, A. J., and Chen, S.
603 (2016) Mycobacterial WhiB6 differentially regulates ESX-1 and the Dos regulon to
604 modulate granuloma formation and virulence in Zebrafish. *Cell Rep* **16**, 2512-2524
605 17. Bosserman, R. E., Nguyen, T. T., Sanchez, K. G., Chirakos, A. E., Ferrell, M. J.,
606 Thompson, C. R., Champion, M. M., Abramovitch, R. B., and Champion, P. A. (2017)
607 WhiB6 regulation of ESX-1 gene expression is controlled by a negative feedback loop in
608 *Mycobacterium marinum*. *Proc Natl Acad Sci USA* **114**, E10772-E10781
609 18. Wu, J., Ru, H. W., Xiang, Z. H., Jiang, J., Wang, Y. C., Zhang, L., and Liu, J. (2017)
610 WhiB4 Regulates the PE/PPE Gene Family and is Essential for Virulence of
611 Mycobacterium marinum. *Sci Rep* **7**, 3007
612 19. Lee, D. S., Kim, P., Kim, E. S., Kim, Y., and Lee, H. S. (2018) *Corynebacterium*
613 *glutamicum* WhcD interacts with WhiA to exert a regulatory effect on cell division genes.
614 *Antonie Van Leeuwenhoek* **111**, 641-648
615 20. Chawla, M., Mishra, S., Anand, K., Parikh, P., Mehta, M., Vij, M., Verma, T., Singh, P.,
616 Jakkala, K., Verma, H. N., AjitKumar, P., Ganguli, M., Narain Seshasayee, A. S., and
617 Singh, A. (2018) Redox-dependent condensation of the mycobacterial nucleoid by
618 WhiB4. *Redox Biol* **19**, 116-133
619 21. Lin, C., Tang, Y., Wang, Y., Zhang, J., Li, Y., Xu, S., Xia, B., Zhai, Q., Li, Y., Zhang, L.,
620 and Liu, J. (2022) WhiB4 Is Required for the Reactivation of Persistent Infection of
621 *Mycobacterium marinum* in Zebrafish. *Microbiol Spectr* **10**, e0044321
622 22. Guiza Beltran, D., Wan, T., and Zhang, L. (2024) WhiB-like proteins: Diversity of
623 structure, function and mechanism. *Biochim Biophys Acta Mol Cell Res* **1871**, 119787
624 23. Soliveri, J. A., Gomez, J., Bishai, W. R., and Chater, K. F. (2000) Multiple paralogous
625 genes related to the Streptomyces coelicolor developmental regulatory gene *whiB* are
626 present in Streptomyces and other actinomycetes. *Microbiology* **146 (Pt 2)**, 333-343
627 24. Burian, J., Ramon-Garcia, S., Sweet, G., Gomez-Velasco, A., Av-Gay, Y., and
628 Thompson, C. J. (2012) The mycobacterial transcriptional regulator *whiB7* gene links
629 redox homeostasis and intrinsic antibiotic resistance. *J Biol Chem* **287**, 299-310
630 25. Saini, V., Farhana, A., and Steyn, A. J. (2012) *Mycobacterium tuberculosis* WhiB3: a
631 novel iron-sulfur cluster protein that regulates redox homeostasis and virulence. *Antioxid*
632 *Redox Signal* **16**, 687-697
633 26. Wan, T., Horova, M., Beltran, D. G., Li, S., Wong, H. X., and Zhang, L. M. (2021)
634 Structural insights into the functional divergence of WhiB-like proteins in
635 *Mycobacterium tuberculosis*. *Mol Cell* **81**, 2887–2900
636 27. Lilic, M., Darst, S. A., and Campbell, E. A. (2021) Structural basis of transcriptional
637 activation by the *Mycobacterium tuberculosis* intrinsic antibiotic-resistance transcription
638 factor WhiB7. *Mol Cell* **81**, 2875-2886 e2875
639 28. Wan, T., Horova, M., Khetrapal, V., Li, S., Jones, C., Schacht, A., Sun, X., and Zhang, L.
640 (2023) Structural basis of DNA binding by the WhiB-like transcription factor WhiB3 in
641 *Mycobacterium tuberculosis*. *The Journal of biological chemistry* **299**, 104777
642 29. Gomez, M., Doukhan, L., Nair, G., and Smith, I. (1998) sigA is an essential gene in
643 Mycobacterium smegmatis. *Molecular Microbiology* **29**, 617-628
644 30. Lee, D. J., Minchin, S. D., and Busby, S. J. (2012) Activating transcription in bacteria.
645 *Annu Rev Microbiol* **66**, 125-152

646 31. Feng, L., Chen, Z., Wang, Z., Hu, Y., and Chen, S. (2016) Genome-wide characterization
647 of monomeric transcriptional regulators in *Mycobacterium tuberculosis*. *Microbiology*
648 **162**, 889-897
649 32. Kudhair, B. K., Hounslow, A. M., Rolfe, M. D., Crack, J. C., Hunt, D. M., Buxton, R. S.,
650 Smith, L. J., Le Brun, N. E., Williamson, M. P., and Green, J. (2017) Structure of a Wbl
651 protein and implications for NO sensing by *M. tuberculosis*. *Nat Commun* **8**, 2280
652 33. Stewart, M. Y. Y., Bush, M. J., Crack, J. C., Buttner, M. J., and Le Brun, N. E. (2020)
653 Interaction of the Streptomyces Wbl protein WhiD with the principal sigma factor σ^{HrdB}
654 depends on the WhiD [4Fe-4S] cluster. *J Biol Chem* **295**, 9752-9765
655 34. Lilic, M., Holmes, N. A., Bush, M. J., Marti, A. K., Widdick, D. A., Findlay, K. C., Choi,
656 Y. J., Froom, R., Koh, S., Buttner, M. J., and Campbell, E. A. (2023) Structural basis of
657 dual activation of cell division by the actinobacterial transcription factors WhiA and
658 WhiB. *Proc Natl Acad Sci USA* **120**, e2220785120
659 35. Wan, T., Li, S., Beltran, D. G., Schacht, A., Zhang, L., Becker, D. F., and Zhang, L.
660 (2020) Structural basis of non-canonical transcriptional regulation by the σ^{A}-bound iron-
661 sulfur protein WhiB1 in *M. tuberculosis*. *Nucleic acids research* **48**, 501-516
662 36. Lonetto, M. A., Rhodius, V., Lamberg, K., Kiley, P., Busby, S., and Gross, C. (1998)
663 Identification of a contact site for different transcription activators in region 4 of the
664 *Escherichia coli* RNA polymerase σ^{70} subunit. *J Mol Biol* **284**, 1353-1365
665 37. Abdallah, A. M., Weerdenburg, E. M., Guan, Q., Ummels, R., Borggreve, S., Adroub, S.
666 A., Malas, T. B., Naeem, R., Zhang, H., Otto, T. D., Bitter, W., and Pain, A. (2019)
667 Integrated transcriptomic and proteomic analysis of pathogenic mycobacteria and their
668 esx-1 mutants reveal secretion-dependent regulation of ESX-1 substrates and WhiB6 as a
669 transcriptional regulator. *PLoS One* **14**, e0211003
670 38. Bosserman, R. E., Nguyen, T. T., Sanchez, K. G., Chirakos, A. E., Ferrell, M. J.,
671 Thompson, C. R., Champion, M. M., Abramovitch, R. B., and Champion, P. A. (2017)
672 WhiB6 regulation of ESX-1 gene expression is controlled by a negative feedback loop in
673 Mycobacterium marinum. *Proc Natl Acad Sci U S A* **114**, E10772-E10781
674 39. Solans, L., Aguilo, N., Samper, S., Pawlik, A., Frigui, W., Martin, C., Brosch, R., and
675 Gonzalo-Asensio, J. (2014) A specific polymorphism in Mycobacterium tuberculosis
676 H37Rv causes differential ESAT-6 expression and identifies WhiB6 as a novel ESX-1
677 component. *Infect Immun* **82**, 3446-3456
678 40. Sala, C., Odermatt, N. T., Soler-Arnedo, P., Gulen, M. F., von Schultz, S., Benjak, A.,
679 and Cole, S. T. (2018) EspL is essential for virulence and stabilizes EspE, EspF and
680 EspH levels in Mycobacterium tuberculosis. *PLoS Pathog* **14**, e1007491
681 41. Vijayaraj, M., Abhinand, P. A., and Ragunath, P. K. (2019) Virtual screening of a MDR-
682 TB WhiB6 target identified by gene expression profiling. *Bioinformation* **15**, 557-567
683 42. Singh, A., Mai, D., Kumar, A., and Steyn, A. J. (2006) Dissecting virulence pathways of
684 *Mycobacterium tuberculosis* through protein-protein association. *Proc Natl Acad Sci USA*
685 **103**, 11346-11351
686 43. Rybniker, J., Nowag, A., van Gumpel, E., Nissen, N., Robinson, N., Plum, G., and
687 Hartmann, P. (2010) Insights into the function of the WhiB-like protein of
688 mycobacteriophage TM4 – a transcriptional inhibitor of WhiB2. *Mol Microbiol* **77**, 642-
689 657

690 44. Chater, K. F., and Chandra, G. (2006) The evolution of development in Streptomyces
691 analysed by genome comparisons. *FEMS Microbiol Rev* **30**, 651-672
692 45. Raju, R. M., Jedrychowski, M. P., Wei, J. R., Pinkham, J. T., Park, A. S., O'Brien, K.,
693 Rehren, G., Schnappinger, D., Gygi, S. P., and Rubin, E. J. (2014) Post-translational
694 regulation via Clp protease is critical for survival of *Mycobacterium tuberculosis*. *PLoS*
695 *Pathog* **10**, e1003994
696 46. Lambert, L. J., Wei, Y., Schirf, V., Demeler, B., and Werner, M. H. (2004) T4 AsiA
697 blocks DNA recognition by remodeling σ^{70} region 4. *EMBO J* **23**, 2952-2962
698 47. Gregory, B. D., Nickels, B. E., Garrity, S. J., Severinova, E., Minakhin, L., Urbauer, R. J.
699 B., Urbauer, J. L., Heyduk, T., Severinov, K., and Hochschild, A. (2004) A regulator that
700 inhibits transcription by targeting an intersubunit interaction of the RNA polymerase
701 holoenzyme. *Proc Natl Acad Sci USA* **101**, 4554-4559
702 48. Shi, J., Wen, A., Zhao, M., You, L., Zhang, Y., and Feng, Y. (2019) Structural basis of
703 sigma appropriation. *Nucleic Acids Res* **47**, 9423-9432
704 49. Sharma, V., Hardy, A., Luthe, T., and Frunzke, J. (2021) Phylogenetic Distribution of
705 WhiB- and Lsr2-Type Regulators in Actinobacteriophage Genomes. *Microbiol Spectr* **9**,
706 e0072721
707 50. Weinbauer, M. G., and Rassoulzadegan, F. (2004) Are viruses driving microbial
708 diversification and diversity? *Environ Microbiol* **6**, 1-11
709 51. Bush, M. J., Bibb, M. J., Chandra, G., Findlay, K. C., and Buttner, M. J. (2013) Genes
710 Required for Aerial Growth, Cell Division, and Chromosome Segregation Are Targets of
711 WhiA before Sporulation in *Streptomyces venezuelae*. *mBio* **4**
712 52. Johnson, M. K. (1998) Iron-sulfur proteins: new roles for old clusters. *Curr Opin Chem*
713 *Biol* **2**, 173-181
714 53. Crack, J. C., Smith, L. J., Stapleton, M. R., Peck, J., Watmough, N. J., Buttner, M. J.,
715 Buxton, R. S., Green, J., Oganesyan, V. S., Thomson, A. J., and Le Brun, N. E. (2011)
716 Mechanistic insight into the nitrosylation of the [4Fe-4S] cluster of WhiB-like proteins. *J*
717 *Am Chem Soc* **133**, 1112-1121
718 54. Crack, J. C., Green, J., Thomson, A. J., and Le Brun, N. E. (2014) Techniques for the
719 production, isolation, and analysis of iron-sulfur proteins. *Methods Mol Biol* **1122**, 33-48
720 55. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in
721 oscillation mode. *Methods Enzymol* **276**, 307-326
722 56. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N.,
723 Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J.,
724 Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S.,
725 Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based
726 system for macromolecular structure solution. *Acta Crystallogr D* **66**, 213-221
727 57. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N.,
728 Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J.,
729 Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S.,
730 Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based
731 system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**,
732 213-221

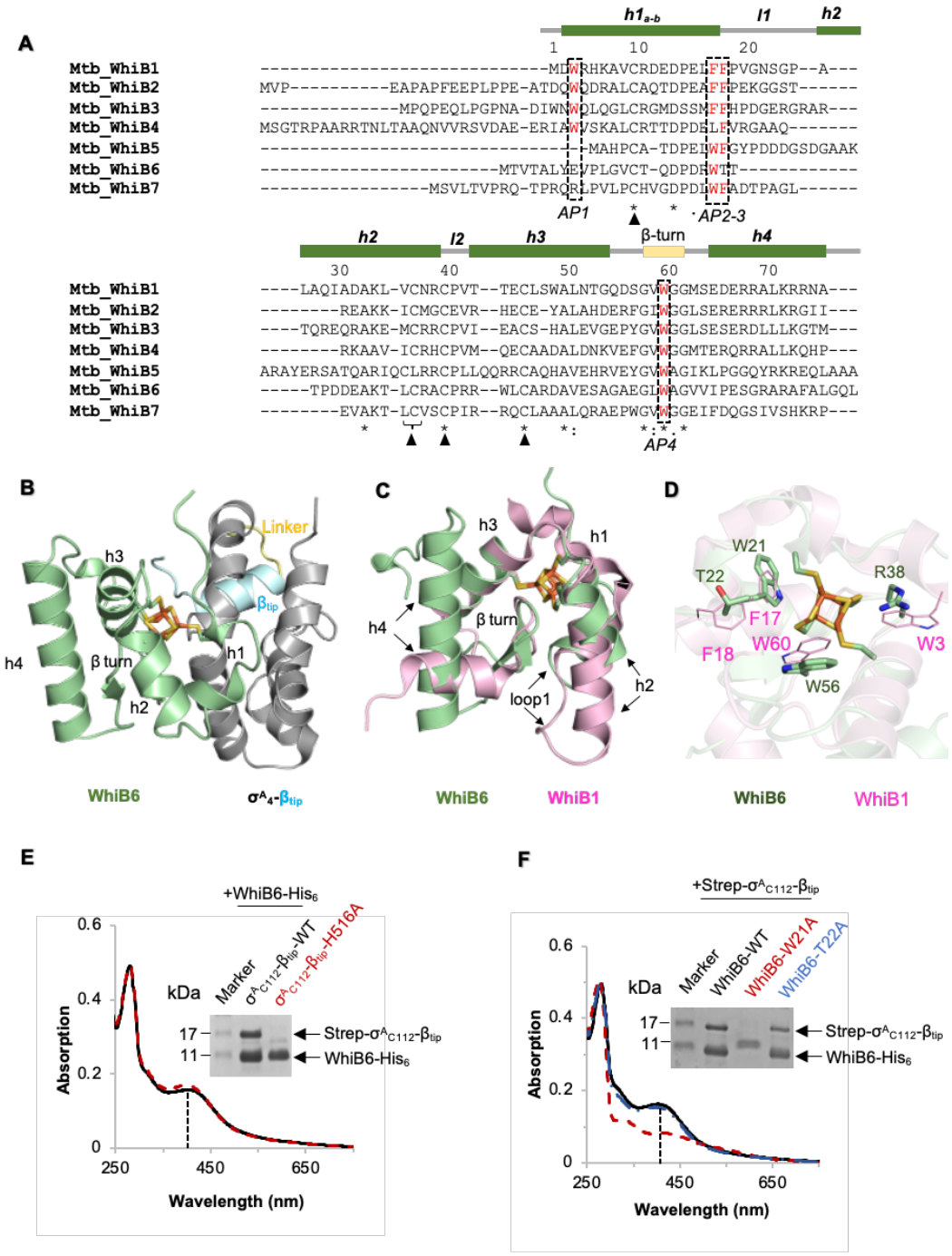
58. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
59. Kapopoulou, A., Lew, J. M., and Cole, S. T. (2011) The MycoBrowser portal: a comprehensive and manually annotated resource for mycobacterial genomes. *Tuberculosis (Edinb)* **91**, 8-13
60. Shell, S. S., Wang, J., Lapierre, P., Mir, M., Chase, M. R., Pyle, M. M., Gawande, R., Ahmad, R., Sarracino, D. A., Ioerger, T. R., Fortune, S. M., Derbyshire, K. M., Wade, J. T., and Gray, T. A. (2015) Leaderless Transcripts and Small Proteins Are Common Features of the Mycobacterial Translational Landscape. *PLoS Genet* **11**, e1005641
61. Cortes, T., Schubert, Olga T., Rose, G., Arnvig, Kristine B., Comas, I., Aebersold, R., and Young, Douglas B. (2013) Genome-wide mapping of transcriptional start sites defines an extensive leaderless transcriptome in *Mycobacterium tuberculosis*. *Cell Reports* **5**, 1121-1131
62. Wan, T., Horova, M., Beltran, D. G., Li, S., Wong, H. X., and Zhang, L. M. (2021) Structural insights into the functional divergence of WhiB-like proteins in Mycobacterium tuberculosis. *Mol Cell* **81**, 2887-2900 e2885
63. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., and Bateman, A. (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res* **49**, D412-D419
64. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948
65. Price, M. N., Dehal, P. S., and Arkin, A. P. (2010) FastTree 2 - approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490
66. Balaban, M., Moshiri, N., Mai, U., Jia, X., and Mirarab, S. (2019) TreeCluster: Clustering biological sequences using phylogenetic trees. *PLoS One* **14**, e0221068
67. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589
68. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-W303
69. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A. J., Bambrick, J., Bodenstein, S. W., Evans, D. A., Hung, C. C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Zemgulyte, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A. I., Cowie, A., Figurnov, M., Fuchs, F. B., Gladman, H., Jain, R., Khan, Y. A., Low, C. M.

R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E. D., Zielinski, M., Zidek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., and Jumper, J. M. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493-500

70. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190
71. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320-324
72. Sievers, F., and Higgins, D. G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**, 135-145
73. Letunic, I., and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**, W293-W296

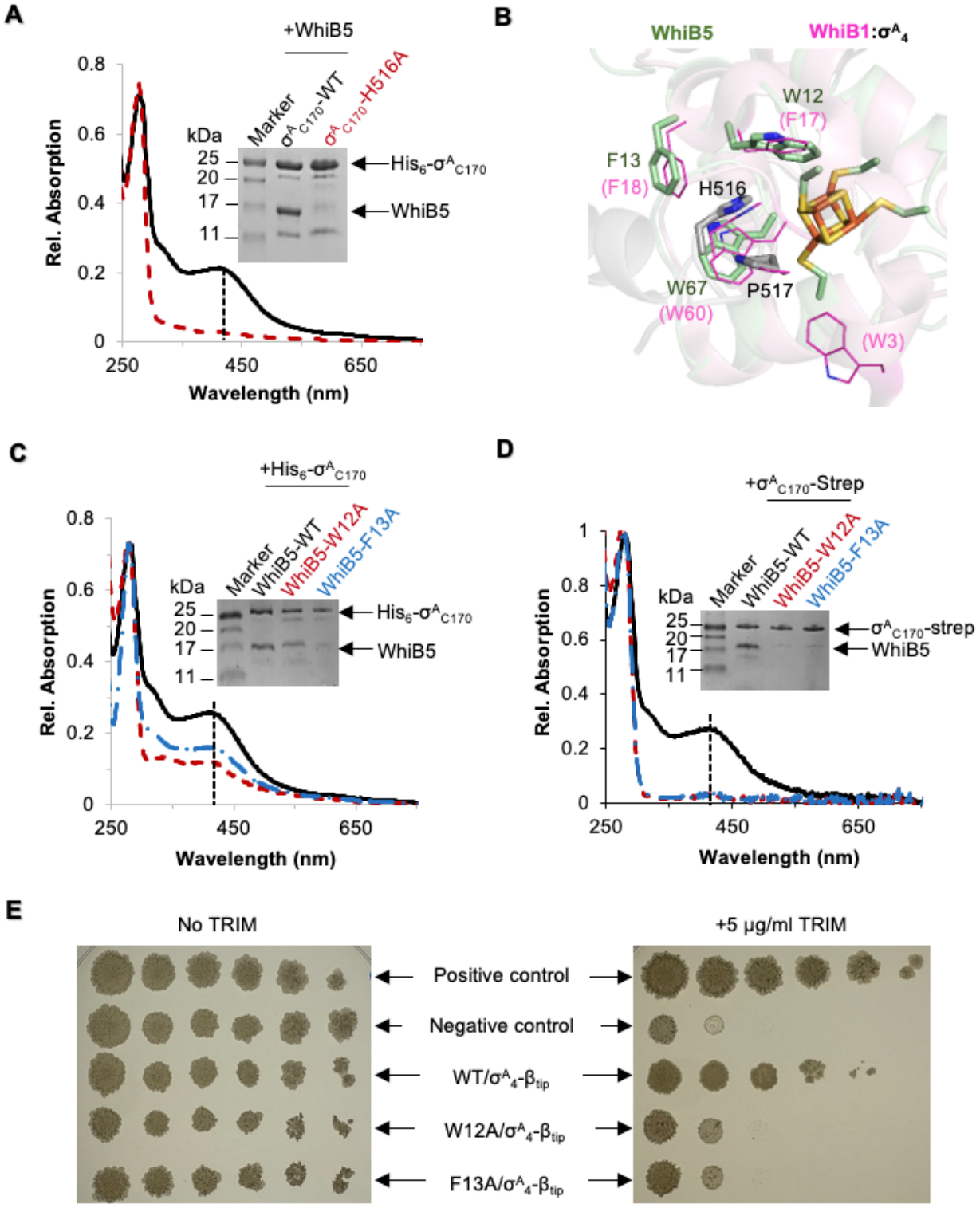
Figures

Figure 1.



793 **Figure 1. The aromatic patch (AP) facilitates *Mtb* WhiB6 binding to $\sigma^A{}_4$.** (A) Sequence
794 alignment of the seven *Mtb* Wbl proteins. The two previously identified motifs in the Wbl family,
795 including the four conserved cysteines (the black triangle) in the [4Fe-4S] cluster binding motif
796 and the Gly-rich β-turn, are highlighted. The assignment of the secondary structures (helices *h1-*
797 *h4*, loop *l1-l2*, β-turn) is based on the $\sigma^A{}_4$-bound WhiB1 (PDB ID:6ONO). The residues in the
798 aromatic patch (AP) in *Mtb* Wbls corresponding to W3, F17, F18, and W60 (referred to as AP1-
799 4, respectively) in WhiB1 for $\sigma^A{}_4$ binding are highlighted in red fonts within the black frame. **(**B)
800 Cartoon representations of the WhiB6:$\sigma^A{}_4$-β_{tip} complex. The key pair of residues, H516-P517, of
801 $\sigma^A{}_4$ for WhiB6 binding is highlighted in sticks. (C-D) Structural overlay of $\sigma^A{}_4$-bound WhiB6 (pale
802 green) with WhiB1 (pink, PDB ID:6ONO) and a close-up view of the Wbl:$\sigma^A{}_4$ interface around
803 the Fe-S cluster binding pocket. The AP1-4 of WhiB1 and corresponding residues in WhiB6 are
804 highlighted in sticks, together with the H516-P517 pair of $\sigma^A{}_4$ in the WhiB6:$\sigma^A{}_4$ complex. $\sigma^A{}_4$ and
805 the [4Fe-4S] cluster of WhiB1 are not shown for clarity. In all structures, the helices *h1-4* and β-
806 turn are indicated. (E and F) UV-visible spectra and SDS-PAGE analyses of the pull-down samples
807 of WhiB6 and $\sigma^A{}_4$-β_{tip} (either wildtype [WT] or mutant as indicated). The color of the UV-visible
808 spectra matches that of the sample labels. The intensity of the absorption peak around 410 nm,
809 highlighted with dashed lines, is indicative of the occupancy of the [4Fe-4S] cluster in the purified
810 samples. The absorption spectra were normalized at 280 nm. The uncropped SDS PAGE images
811 are shown in Fig. S3.

812 **Figure 2**

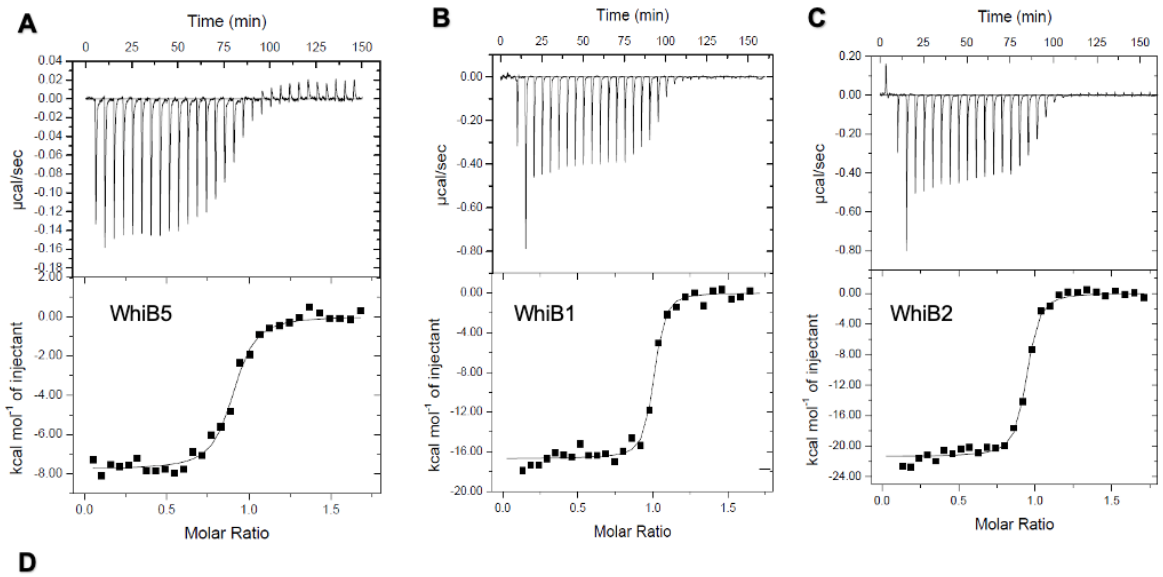


813

814 **Figure 2. Characterization of WhiB5 binding to σ^A_4.** (A) (C) and (D) UV-visible spectra and
815 SDS-PAGE analyses of the samples from co-expression and affinity purification to test WhiB5
816 binding to σ^A_4 (wildtype or mutant as indicated). For Panels (A) and (C), His_6-σ^A_{C170} (wildtype or
817 mutant as indicated) was used as the bait in the pull-down assays. For Panel (D), the Strep-tagged
818 σ^A_{C170} (σ^A_{C170}-Strep) was used as the bait. In all cases, the absorption spectra were normalized at
819 280 nm. The colors of the UV-Visible spectra correspond to those of the SDS-PAGE sample labels.
820 The intensity of the absorption peak around 410 nm, highlighted with dashed lines, is indicative
821 of the occupancy of the [4Fe-4S] cluster in the purified samples. The uncropped SDS PAGE
822 images are shown in Supplementary Fig. 5. (B) Homology model of the σ^A_4-bound WhiB5 by
823 SWISS-Model (pale green) overlaid with that of WhiB1 (PDB ID:6ONO, pink). The [4Fe-4S]
824 clusters and the aromatic patch residues in Wbls, and the H516-P517 pair in σ^A_4 are shown in
825 sticks. (E) Detection of the interactions between *Mtb* WhiB5 (wildtype and mutants as indicated)
826 and σ^A_4 in *Mycobacterium smegmatis* (*Msm*) by Mycobacterial Protein Fragment
827 Complementation (M-PFC) assay (see in Materials and Methods). Positive control, the *Msm* strain
828 co-expressing the *Saccharomyces cerevisiae* leucine-zipper sequence homodimerization domain
829 GCN4, with the fragment of $mDHFR_{F[1,2]}$ and $mDHFR_{F[3]}$, respectively, fused to the C-terminus
830 of GCN4 (GCN4-$mDHFR_{F[1,2]}$ and GCN4-$mDHFR_{F[3]}$); Negative control, the *Msm* strain
831 expressing the two fragments $mDHFR_{F[1,2]}$ and $mDHFR_{F[3]}$ separately. The *Msm* strain co-
832 expressing the recombinant proteins, WhiB5-$mDHFR_{F[1,2]}$ and σ^A_{C82}-β_{tip}-$mDHFR_{F[3]}$, to test the
833 interaction between WhiB5 (wildtype or mutant as indicated) and σ^A_4. All the *Msm* strains were
834 grown on the 7H10 broth agar plates in the absence or presence of 5 μg/ml trimethoprim (TRIM)
835 as indicated. Cell growth in the presence of TRIM is indicative of the protein-protein interaction.

836 **Figure 3**

837



D

Sample	Ka ($x10^7$ M^{-1})	K_d (nM)	ΔH (kcal/mol)	−TΔS (kcal/mol)	n	Chi^2/DoF
WhiB1	8.9 ± 2.3	12.0± 3.0	-16.68 ± 0.18	7.18	0.981 ± 0.005	4.11E5
WhiB2	6.1 ± 1.1	16.9± 3.0	-21.36 ± 0.19	13.39	0.919 ± 0.004	3.804E5
WhiB5	2.1 ± 0.4	50.3± 10.0	-7.76 ± 0.10	2.77	0.884 ± 0.007	1.008E5

838

839 **Figure 3. Isothermal titration calorimetry (ITC) assay of the interactions between $\sigma^A{}_4$ and**
840 ***Mtb* Wbls.** (A-C) Representative results for Isothermal titration calorimetry (ITC) assay of the
841 interactions between $\sigma^A{}_4$ and *Mtb* Wbls as indicated. The calorimetric titration is shown in the top
842 panel, with the integrated injection heat derived from the titration and the best-fit curve (dash lines)
843 of a simple 1:1 interaction model shown in the bottom panel. (D) Summary of the binding
844 parameters from the best fit, including the association constant (K_a), dissociation constant (K_d),
845 changes in enthalpy (ΔH) and entropy (ΔS) upon binding, and the goodness of the fit (Chi^2/DoF).

846

847

848 **Figure 4**

849

850

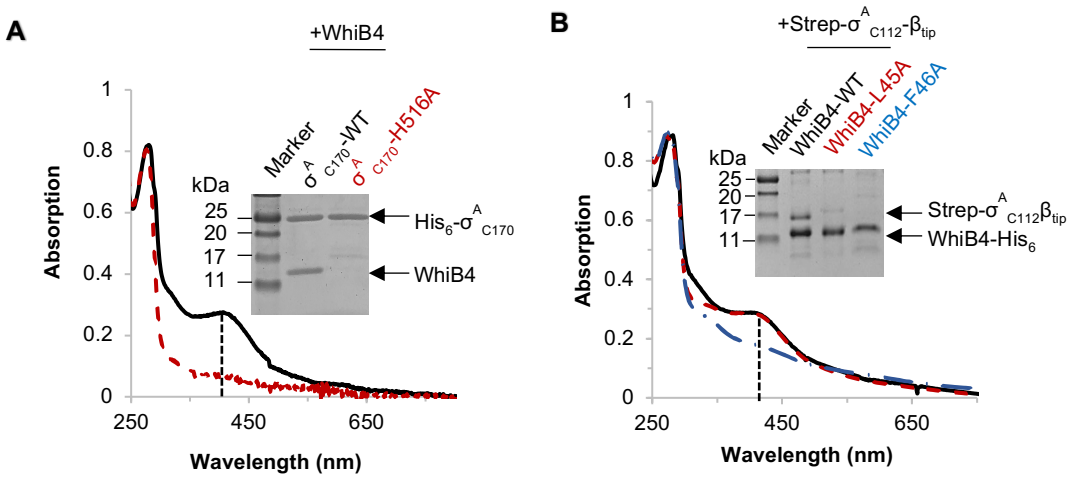
851

852

853

854

855



856 **Figure 4. Role of the aromatic patch residues in WhiB4 binding to σ^A_4.** (A) and (B) UV-visible

857 spectra and SDS-PAGE analyses of the samples from co-expression and purification to test the

858 interaction between σ^A_4 and WhiB4 (either wildtype or mutant as indicated). His_6-σ^A_{C170} (wildtype

859 or mutant as indicated) was used as the bait in (A) to test the interaction with the tagless WhiB4.

860 For Panel (B), WhiB4-His_6 (wildtype or mutant as indicated) was used as the bait to test the

861 interaction with Strep-σ^A_4-β_{tip}. In both cases, the absorption spectra were normalized at 280 nm.

862 The intensity of the absorption peak around 410 nm, highlighted with dashed lines, is indicative

863 of the occupancy of the [4Fe-4S] cluster in the purified samples. The colors of the UV-Vis spectra

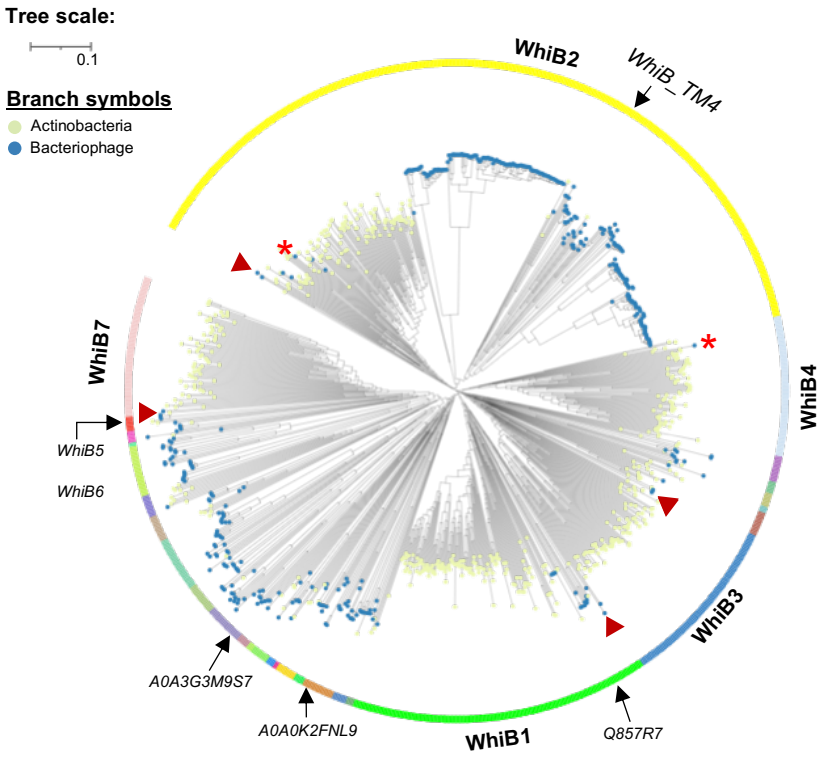
864 match those of the SDS-PAGE sample labels. The uncropped SDS-PAGE images are shown in

865 Supplementary Fig. 7.

866 **Figure 5**

867

868 869 870 871 872 873 874 875 876 877 878



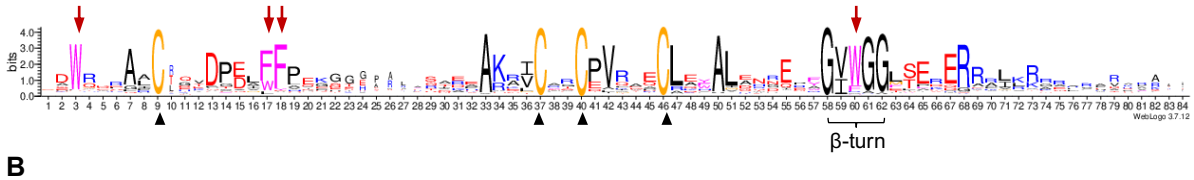
879 **Figure 5. Phylogenetic analysis of the Wbl proteins from the representative genomes of**
880 **Actinobacteria and actinobacteriophages.** The phylogenetic tree of the 995 Wbl family proteins
881 from the representative genomes of Actinobacteria and actinobacteriophages using the Pfam
882 PF02467 (RP15), with highlights of the 29 Wbl subfamilies and the Wbls from
883 actinobacteriophages. The five major Wbl subfamilies are labeled with the representative Wbls in
884 *Mtb* (WhiB1-4 and WhiB7, respectively). The remaining smaller subfamilies are color-coded and
885 assigned sequential numbers in the tree. Examples of the phage Wbl at the early diverging point
886 of the clades are highlighted by the red arrowheads, while those nested within otherwise bacterial

887 clades are highlighted by the red stars. The three phage Wbls (Uniprot IDs: A0A0K2FNL9,

888 Q857R7 and A0A3G3M9S7) tested for σ^{A}_{4} in this study are indicated on the tree.

Figure 6

A



B

	AP1	AP2	AP3	AP4	Match 1+ @AP2-3[c]	Match 2+ @AromPatch[d]
Survey 1[a]	822 (82.6%) (808 W, 8 F, 4 Y, 2 H)	914 (91.9%) (232 W, 670 F, 7 Y, 5 H)	893 (89.7%) (5 W, 831 F, 45 Y, 12 H)	928 (93.3%) (784 W, 126 F, 17 Y, 1 H)	991 (99.6%)	984 (98.9%)
Survey 2[b]	824 (82.8%) (810 W, 8 F, 6 Y, 3 H)	915 (92.0%) (232 W, 672 F, 7 Y, 4 H)	892 (89.6%) (5 W, 831 F, 46 Y, 10 H)	936 (94.1%) (784 W, 132 F, 19 Y, 1 H)	993 (99.8%)	992 (99.7%)

Figure 6. Conservation of the aromatic patch residues across the Wbl proteins from the representative genomes of Actinobacteria and actinobacteriophages. (A) The sequence logo of the 995 Wbl family proteins from the representative genomes of Actinobacteria and actinobacteriophages (Pfam PF02467, RP15) was generated by WebLogo. The four conserved cysteines for [4Fe-4S] cluster binding are highlighted by black arrowheads, and those in the aromatic patch indicated by red arrows. (B) Summary of the distribution of the aromatic patch in the Wbl family based on 3D structural alignment against *Mtb* WhiB1. [a]Survey 1 includes the number of the aromatic residues (W, F, Y or H as indicated) corresponding to the aromatic patch residues AP1-4 (W3, F17, F18 and W60, respectively) in *Mtb* WhiB1 (PDB code:6ONO) by structural alignment. [b]Survey 2 includes all the aromatic residues (W, F, Y or H as indicated) at the position -1 to +1 relative to the aromatic patch residues AP1-4 by structural alignment. [c]Total number of Wbls containing at least one aromatic residue corresponding to AP2-3 (F17 or F18 in

911 *Mtb* WhiB1). [d]Total number of Wbls containing at least two aromatic residues corresponding to

912 the aromatic patch in WhiB1.