

Chasing the structural diversity of the transcription regulator *Mycobacterium tuberculosis* HigA2William Richardson,<sup>a</sup> Gyun Won Kang,<sup>b</sup> Hee Joong Lee,<sup>b</sup> Kang Mu Kwon,<sup>b</sup> Saron Kim<sup>b</sup> and Hyo Jung Kim<sup>a,b,\*</sup>

Received 26 May 2021

Accepted 28 July 2021

Edited by Z.-J. Liu, Chinese Academy of Sciences, China

**Keywords:** toxin–antitoxin systems; *Mycobacterium tuberculosis*; transcription factors; HigBA; HigA2; crystal structure; structure determination; X-ray crystallography.

**PDB references:** *Mycobacterium tuberculosis* HigA2, 7ewc; 7ewd; 7ewe

**Supporting information:** this article has supporting information at [www.iucrj.org](http://www.iucrj.org)

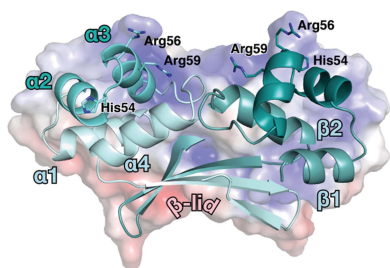
<sup>a</sup>Centre for Biomolecular Sciences, School of Pharmacy, University of Nottingham, Nottingham, United Kingdom, and <sup>b</sup>College of Pharmacy, Woosuk University, Wanju 55338, Republic of Korea. \*Correspondence e-mail: [hyojungkim@woosuk.ac.kr](mailto:hyojungkim@woosuk.ac.kr)

Transcription factors are the primary regulators of gene expression and recognize specific DNA sequences under diverse physiological conditions. Although they are vital for many important cellular processes, it remains unclear when and how transcription factors and DNA interact. The antitoxin from a toxin–antitoxin system is an example of negative transcriptional autoregulation: during expression of the cognate toxin it is suppressed through binding to a specific DNA sequence. In the present study, the antitoxin HigA2 from *Mycobacterium tuberculosis* M37Rv was structurally examined. The crystal structure of *M. tuberculosis* HigA2 comprises three sections: an N-terminal autocleavage region, an  $\alpha$ -helix bundle which contains an HTH motif, and a C-terminal  $\beta$ -lid. The N-terminal region is responsible for toxin binding, but was shown to cleave spontaneously in its absence. The HTH motif performs a key role in DNA binding, with the C-terminal  $\beta$ -lid influencing the interaction by mediating the distance between the motifs. However, *M. tuberculosis* HigA2 exhibits a unique coordination of the HTH motif and no DNA-binding activity is detected. Three crystal structures of *M. tuberculosis* HigA2 show a flexible alignment of the HTH motif, which implies that the motif undergoes structural rearrangement to interact with DNA. This study reveals the molecular mechanisms of how transcription factors interact with partner proteins or DNA.

## 1. Introduction

Bacterial gene regulation is controlled by a multitude of transcription factors that recognize specific DNA sequences and allow controlled cellular responses (Villard, 2004). Given their role in responding to environmental cues such as host colonization and virulence, transcription factors have been intensively studied as drug targets (Liu *et al.*, 2015). Most transcription factors are multidomain proteins that possess a DNA-binding domain and an effector domain for ligand or protein interaction. There are many known structural motifs for DNA binding, including helix–turn–helix, zinc-finger, leucine-zipper and helix–loop–helix motifs. The simplest motif, helix–turn–helix (HTH), comprises two  $\alpha$ -helices with a fixed angle permitting binding to the major groove of DNA (Luscombe *et al.*, 2000). Since the amino-acid side chains of  $\alpha$ -helices are solvent-accessible, they invariably make key interactions with DNA, with their composition influencing the specificity. HTH motif-containing proteins typically form dimers that strengthen their DNA interaction and can mirror the dyad symmetry of their binding site (Brennan & Matthews, 1989).

In the present study, we elucidated the crystal structure of *Mycobacterium tuberculosis* H37Rv HigA2 (hereafter referred to as  $M_t$ HigA2), which contains an HTH motif (Sala *et al.*,



2014). HigB and HigA constitute a bacterial toxin–antitoxin (TA) system which is organized into a small operon. The HigA antitoxin represses the transcription of HigBA or forms a stable HigBA complex, thereby preventing the ribonuclease activity of the HigB toxin and the associated cytotoxic events. However, antitoxins are known to be actively degraded, with the liberated toxins leading to increased pathogenicity (Maisonneuve & Gerdes, 2014; Schureck *et al.*, 2016). There are five well known TA systems based on their modes of interaction. Toxins bind to either RNA antitoxins (types I and III) or protein antitoxins (types II, IV and V) (Ghafourian *et al.*, 2014). Type II is well characterized and abundant, and functions by binding of the antitoxin protein to either the DNA or the toxin (Fraikin *et al.*, 2020). The HigBA system is a type II TA system that is found in many pathogens, including *Pseudomonas aeruginosa*, *Proteus vulgaris*, *Vibrio cholerae*, *Streptococcus pneumoniae* and *M. tuberculosis* (Kędzierska & Hayes, 2016).

*M. tuberculosis* is the causative organism of tuberculosis and is a significant contributor to global mortality, with the World Health Organization reporting 1.4 million deaths in 2019 (Fukunaga *et al.*, 2021). *M. tuberculosis* usually colonizes the lungs and can persist in host tissues for decades without leading to disease, and spreads easily through air transmission. Current treatment regimens are lengthy and consist of a 6–9 month course of four antibiotics, with some concern regarding rifampicin-resistant and multidrug-resistant (MDR) *M. tuberculosis* (Mabhula & Singh, 2019). Little is known about the interactions between the host and bacteria during persistent infection or drug resistance, but it has been suggested that TA systems play an essential role. The genome of pathogenic *M. tuberculosis* H37Rv has 79 TA systems, while other nonpathogenic mycobacterial genomes possess only 5–10 TA systems (Sala *et al.*, 2014). The risk of active tuberculosis infection increases when the toxin is released from the antitoxin (Gupta, 2009; Ramage *et al.*, 2009). Among the 79 TA systems in *M. tuberculosis* H37Rv, there are 38 type II TA systems, which include two HigBA systems:  $M_t$ HigBA2 and  $M_t$ HigBA3. Of the two,  $M_t$ HigBA2 is categorized among ten of the 79 TA systems which are induced drastically in drug-tolerant persister cells. In addition,  $M_t$ HigBA2 is known to be important for survival in lung tissue (Stewart *et al.*, 2005; Jain *et al.*, 2007). In this study, we determined the crystal structure of  $M_t$ HigA2, a transcription factor from the tuberculosis-causing pathogen *M. tuberculosis* H37Rv.  $M_t$ HigA2 exploits structural characteristics to interact with the toxin or DNA. This study presents a better understanding of how a multi-functional transcription factor regulates its function through structural changes. This work should contribute new insights into pathogenic bacterial physiology and pathogenicity.

## 2. Materials and methods

### 2.1. Cloning, expression and purification

The gene encoding the  $M_t$ HigA2 antitoxin was amplified from *M. tuberculosis* H37Rv genomic DNA by PCR using

5'-CCAGGGAGCAGCCTCGATGGCGATGACACTACGG GGACATGGAC-3' and 5'-CCAGGGAGCAGCCTCGCTAT GCCAGGGTGAATGTCTCATCTCC-3' as the forward and reverse primers, respectively. A plasmid for  $M_t$ HigA2 was prepared using a ligation-independent cloning (LIC) strategy based on a locally engineered pET-15b vector as described previously (Aslanidis & de Jong, 1990; Eschenfeldt *et al.*, 2009; Jeong *et al.*, 2012; Kim *et al.*, 2018; Kim, 2020). The amplified DNA of  $M_t$ HigA2 was inserted into an engineered vector containing an additional thioredoxin (Trx) tag linked by a Tobacco etch virus (TEV) protease cleavage site (Parks *et al.*, 1994). The resulting construct consisted of an N-terminal hexahistidine tag, a TEV cleavage site, GAAS for LIC and the  $M_t$ HigA2 gene. Each recombinant plasmid was transformed into *Escherichia coli* DH5 $\alpha$  cells and verified by DNA sequencing. For expression, the recombinant plasmid was transformed into *E. coli* C41 cells. The cells were grown in Luria broth (LB) medium supplemented with ampicillin (50  $\mu$ g ml<sup>-1</sup>) at 37°C. Expression of recombinant Trx- $M_t$ HigA2 protein was induced by the addition of 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) upon reaching an OD<sub>600</sub> of 0.5 and the culture was grown at 37°C for an additional 4 h. The cells were harvested by centrifugation at 4500g at 4°C. The cell pellet was resuspended and lysed on ice by sonication in lysis buffer A (50 mM Tris–HCl pH 7.5, 500 mM NaCl), and the lysate was clarified by centrifugation at 20 000g for 1 h at 4°C. The cleared supernatant was applied onto an Ni<sup>2+</sup>-nitrilotriacetate (Ni–NTA) affinity column (Qiagen, Germany) and eluted with elution buffer containing 250 mM imidazole. To remove the Trx tag,  $M_t$ HigA2 was incubated with TEV protease at a 10:1 molar ratio of  $M_t$ HigA2:TEV protease for 1 h at 4°C. Since TEV protease possesses a hexahistidine tag, only cleaved  $M_t$ HigA2 was retrieved using an Ni–NTA affinity column. The purified protein was analyzed with >95% purity by SDS–PAGE and was concentrated to 8 mg ml<sup>-1</sup> by ultra-filtration in 3000 Da molecular-mass cutoff spin columns (Millipore, USA).

### 2.2. Crystallization, data collection and structure determination

Crystals of  $M_t$ HigA2 were grown by sitting-drop vapour diffusion at 20°C using a 96-well crystallization plate. Initial crystallization conditions were established using screening kits from Hampton Research (Crystal Screen, Crystal Screen 2, Index, PEG/Ion and Natrix), Molecular Dimensions (ProPlex, JCSG-plus and Structure Screen I and II) and Emerald Bio-Systems (Wizard I, II, III and IV). For the optimal growth of  $M_t$ HigA2 crystals, 1  $\mu$ l (8 mg ml<sup>-1</sup>)  $M_t$ HigA2 solution was mixed with 1  $\mu$ l precipitant solution and equilibrated against a 1 ml reservoir of the precipitant solution. The best crystals of  $M_t$ HigA2 were obtained using three conditions: (i) 11%(w/v) PEG 20K, 0.1 M MES pH 6.5; (ii) 10%(w/v) PEG 8K, 8%(v/v) ethylene glycol, 0.1 M HEPES pH 7.5; and (iii) 25%(w/v) PEG 4K, 0.2 M ammonium sulfate, 0.1 M sodium acetate trihydrate pH 4.6. Crystals were transferred to a cryoprotectant solution containing 20%(v/v) glycerol in each crystallization condition

Table 1

Crystal data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	Form I	Form II	Form III
Data collection			
Beamline	BL-5C, PLS	BL-7A, PLS	I04, DLS
Wavelength (Å)	0.98	0.98	0.98
Resolution range (Å)	40.0–2.1	40.0–3.2	77.0–3.4
Space group	$P2_12_12_1$	$P4_32_12$	$P3_121$
<i>a</i> , <i>b</i> , <i>c</i> (Å)	30.610, 89.955, 114.961	67.621, 67.621, 190.646	80.890, 80.890, 153.910
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Observations (total/unique)	379872/20447	14009/7877	47017/9603
Completeness (%)	99.8 (99.7)	100 (100)	99.5 (99.2)
$R_{\text{merge}}^\dagger$	0.079 (0.261)	0.105 (0.757)	0.112 (2.663)
$CC_{1/2}$	0.99 (0.98)	0.98 (0.89)	0.99 (0.51)
Multiplicity	6.0 (6.2)	7.4 (7.4)	4.9 (5.1)
$\langle I/\sigma(I) \rangle$	29.0 (15.5)	24.7 (7.0)	7.9 (0.9)
Refinement			
$R_{\text{work}}/R_{\text{free}}^\ddagger$ (%)	20.9/24.7	21.5/26.4	20.6/27.7
Average <i>B</i> value (Å <sup>2</sup> )	35.4	99.6	126.4
R.m.s.d., bond lengths (Å)	0.010	0.009	0.009
R.m.s.d., angles (°)	1.693	1.697	1.849
Ramachandran plot (%)			
Favoured	98.88	89.32	82.74
Allowed	1.12	7.83	13.74
Disallowed	0	2.85	4.09

$^\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the *i*th measured intensity of reflection *hkl* and  $\langle I(hkl) \rangle$  is the mean of all measured intensities of reflection *hkl*.  $^\ddagger R_{\text{work}} = \sum_{hkl} ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_{hkl} |F_{\text{obs}}|$ , where  $F_{\text{obs}}$  is the observed structure-factor amplitude and  $F_{\text{calc}}$  is the structure factor calculated from the model.  $R_{\text{free}}$  is computed in the same manner as  $R_{\text{work}}$  but from a test set containing 5% of the data, which were excluded from the refinement calculation.

and were flash-cooled in a stream of nitrogen at 100 K. Diffraction data were collected on the BL-5C experimental station at Pohang Light Source (PLS), Korea and the I04 experimental station at Diamond Light Source (DLS), UK. The data sets were processed and scaled using *XDS* and the *CCP4* suite (Kabsch, 2010; Winn *et al.*, 2011). The three crystal forms belonged to space groups  $P2_12_12_1$ ,  $P4_32_12$  and  $P3_121$ , respectively. The structure packed in space group  $P2_12_12_1$  was solved by molecular replacement with *Phaser*, using an ensemble search model generated from PDB entries 1y7y, 3b7h, 3kxa and 2a6c by *MrBUMP* in *CCP4* (McCoy *et al.*, 2007; Keegan *et al.*, 2018; McGeehan *et al.*, 2005; Ren *et al.*, 2010). It produced a marginal solution with three monomers in the asymmetric unit with a log-likelihood gain (LLG) of 134. The model was then rebuilt automatically with *ARP/wARP* and manually with *Coot*, and refined with *REFMAC5* and *Phenix* (Emsley *et al.*, 2010; Liebschner *et al.*, 2019; Murshudov *et al.*, 2011; Afonine *et al.*, 2012; Langer *et al.*, 2008).

For structural determination of the  $P4_32_12$  and  $P3_121$  crystal forms, molecular replacement was used in *Phaser* using the  $M_{\text{HigA2}}$  structure packed in space group  $P2_12_12_1$  as the template. Iterative cycles of model building were performed using *Coot*, followed by refinement in *REFMAC5* and *Phenix* (Emsley *et al.*, 2010; Liebschner *et al.*, 2019; Murshudov *et al.*, 2011). The crystal contained two dimers (four monomers) per asymmetric unit when the structure was packed in space groups  $P2_12_12_1$  and  $P4_32_12$ . Packing in space group  $P3_121$  showed one more monomer, with a total of five monomers in the asymmetric unit, and the unpaired monomer forms a crystallographic dimer with twofold symmetry. A portion of

the data (5%) were set aside before refinement. The final crystallographic statistics are summarized in Table 1. Structural alignments and figures were generated using *PyMOL* (<http://www.pymol.org>) and *UCSF Chimera* (Pettersen *et al.*, 2004).

### 2.3. PDB codes

Protein coordinates and structure factors have been deposited in the RCSB PDB as entries 7ewc, 7ewd and 7ewe.

## 3. Results

### 3.1. The $\beta$ -lid anchors the dimeric state in the crystal structure of *M. tuberculosis* HigA2

The crystal structure of  $M_{\text{HigA2}}$  was determined from three distinct crystal forms (Supplementary Fig. S1). Form I belonged to space group  $P2_12_12_1$  and was determined to 2.0 Å resolution, form II packed in space group  $P4_32_12$  and was determined to 3.2 Å resolution, and form III was solved in space group  $P3_121$  to 3.4 Å resolution. Due to a lack of electron density at the N-terminus, ~25 residues are undefined in each monomer despite the relatively high resolution. The N-terminally cleaved monomer consists of four consecutive  $\alpha$ -helices ( $\alpha 1$ ,  $\alpha 2$ ,  $\alpha 3$  and  $\alpha 4$ ) and two antiparallel  $\beta$ -strands ( $\beta 1$  and  $\beta 2$ ), named the  $\alpha$ -helix bundle and  $\beta$ -lid after the previously determined HipB anti-toxin structure (Schumacher *et al.*, 2009). The  $\alpha$ -helix bundle contains an HTH motif that is required for DNA binding, comprising a preceding helix ( $\alpha 2$ ) and a recognition helix ( $\alpha 3$ ) (Matthews *et al.*, 1982; Wintjens & Rooman, 1996). In the  $M_{\text{HigA2}}$  crystal structure, positively charged residues on the HTH motif, His54, Arg56 and Arg59, are oriented towards the expected DNA-binding region. The role of dimerization is delegated to the C-terminal  $\beta$ -lid interface, which comprises two antiparallel  $\beta$ -strands from each monomer that stack to form a four-stranded antiparallel  $\beta$ -sheet, referred to as a  $\beta$ -lid, due to its curvature. A substantial number of  $\beta$ -lid hydrogen bonds and salt bridges result in a tight dimerization network [Figs. 1(a) and 1(b)].

Although dimerization is largely dependent on the  $\beta$ -lid, some limited interaction is observed between the  $\alpha$ -helix bundles through the hydrogen bonding of backbone atoms. However, the interaction distances are diverse between the structures determined from the crystal forms. All dimers share the  $\beta$ -lid as their major dimeric interface, but the position of the  $\alpha$ -helix bundle varies. When the different dimers are aligned based on a single HTH motif from one monomer, the corresponding motif in the dimer pair shows a different relative position. The flexibility in dimerization leads to drastic

changes in the locations of positively charged residues on the HTH motif, including His54, Arg56 and Arg59. The greatest change in distance is that between the N<sup>δ</sup> atoms of His54, which move by ~7 Å [Figs. 1(c) and 1(d)]. This implies that flexibility in dimerization would have a huge impact on DNA bending. To predict the physiological dimer, each form was submitted to the PISA server to calculate the strength of the dimer interface. Interestingly, all of the forms had similarly favourable dissociation energies ( $\Delta G$ ): -14, -15 and -17 kcal mol<sup>-1</sup> for forms I, II and III, respectively. This suggests that the flexibility of the  $\beta$ -lid does not impact the formation of a stable dimer and is likely to contribute to function.

### 3.2. Autocleavage of *M. tuberculosis* HigA2

M<sub>T</sub>HigA2 is a 101-amino-acid protein (11 kDa) that forms a dimeric state in solution. However, the N-terminal ~25 residues are unstructured in all three observed crystal forms. Interestingly, this agrees with our previous structural observations on M<sub>T</sub>HigA3, which was also determined to have a

cleaved N-terminus. Although crystallization was attempted in the presence of a covalently linked Trx at the N-terminus (~13 kDa), the whole Trx was not packed in the crystal, implying that the protein tends to be cleaved spontaneously. A similar result was observed for *V. cholerae* HigA, which was crystallized with and without the cognate HigB toxin. The N-terminal ~25 residues are not structured when crystallized without the toxin. These cleaved residues were found to interact with the toxin through the elucidation of a toxin–antitoxin complex crystal structure (Hadži *et al.*, 2017). The same pattern is shown in the crystal structure of *S. pneumoniae* HigA, indicating that the missing N-terminus in the M<sub>T</sub>HigA2 structure is important for interaction with the M<sub>T</sub>HigB2 toxin (Kang *et al.*, 2020).

Since the autocleavage of M<sub>T</sub>HigA2 has been confirmed to be biologically important, size-exclusion chromatography (SEC) was employed to determine when the protein is cleaved. As M<sub>T</sub>HigA2 lacks tryptophan, it possesses a low molar extinction coefficient (1490 M<sup>-1</sup> cm<sup>-1</sup>), but it is still detectable due to a single tyrosine (Tyr79) [Fig. 2(a)]. However, the Trx-M<sub>T</sub>HigA2 construct showed an indicative

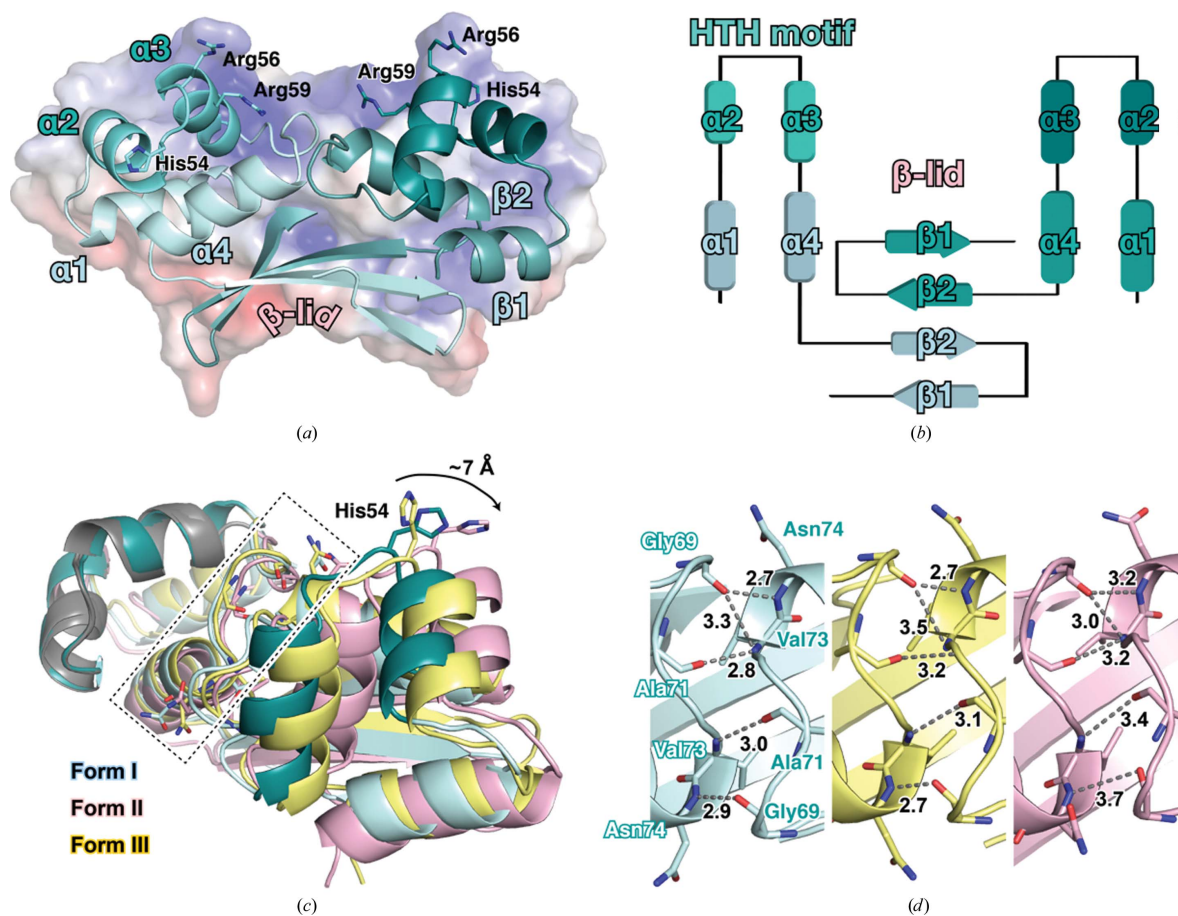


Figure 1

Crystal structure of M<sub>T</sub>HigA2. (a) Crystal structure of the M<sub>T</sub>HigA2 dimer in ribbon representation. The two monomers are coloured light and dark cyan. The HTH motif is highlighted in darker colours. Three positively charged side chains on the HTH motif are indicated as sticks. Potential surface charge is indicated in the background, where surfaces are coloured between -10 kcal mol<sup>-1</sup> (red) and +10 kcal mol<sup>-1</sup> (blue). (b) Secondary-structure diagram of the M<sub>T</sub>HigA2 dimer. The colours correspond to those in (a). (c) The three forms of the M<sub>T</sub>HigA2 crystal structure when the HTH motifs are aligned. The arrangement of the second HTH motif is diverse. The side chain of His54 moves ~7 Å depending on the crystal form. (d) Detailed view of the dimerization interface enclosed with a dotted box in (c). Hydrogen bonds are indicated as grey dotted lines and distances are given in Å.

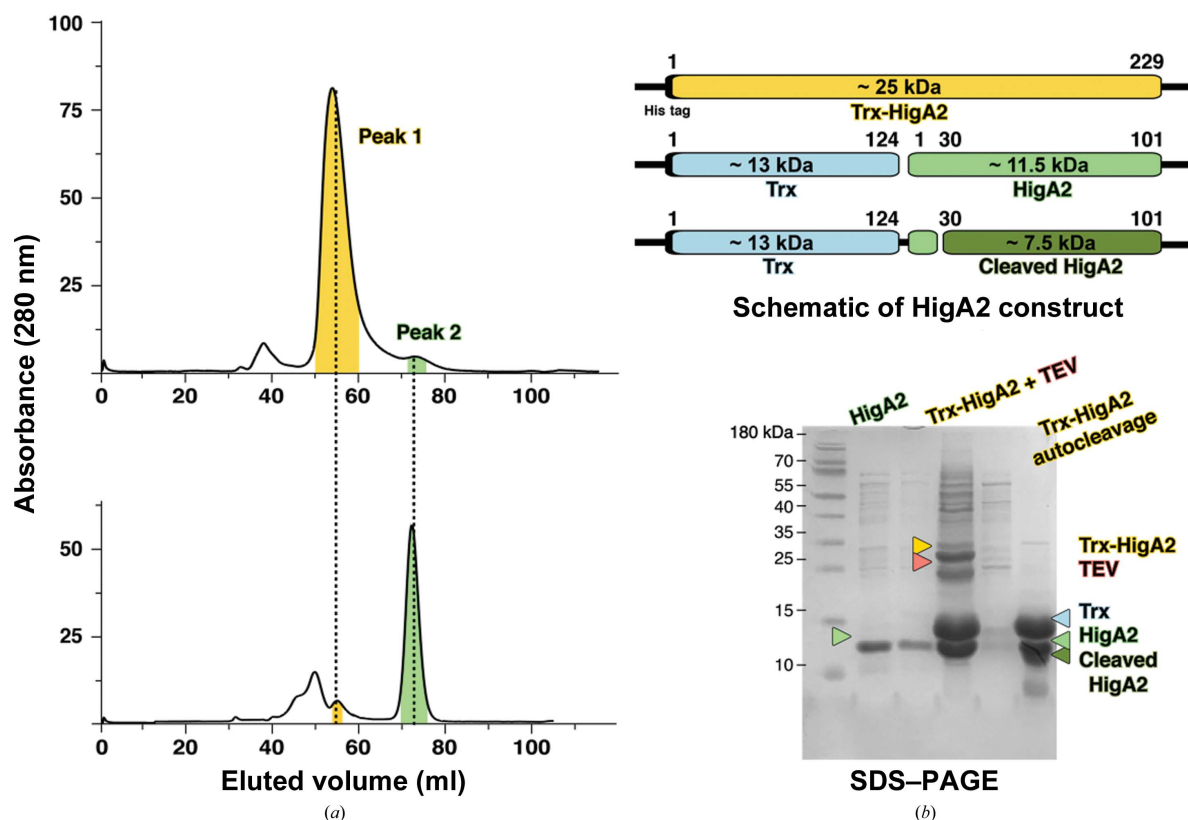
absorbance level since it has an additional N-terminal Trx tag. Cleavage using TEV protease results in Trx (13 kDa) and  $Mt$ HigA2 (11.5 kDa), and a similar result is observed when the protein is preserved at 4–20°C overnight without TEV protease. In both cases,  $Mt$ HigA2 is cleaved further within hours [Fig. 2(b)]. The predicted cleavage site resides between residues 20 and 29, where a significant number of charged residues are present (eight out of ten) [Fig. 3(a)]. Disorder analysis with *DisEMBL* supports the observation in the crystal structure that the N-terminus is intrinsically disordered and is spontaneously cleaved in the absence of the toxin (Linding *et al.*, 2003).

### 3.3. Comparison of *M. tuberculosis* HigA2 with related proteins

A total of three  $Mt$ HigBA pairs have been identified in *M. tuberculosis* H37Rv, including HigBAC and HigBA1. In the present study, we determined the structure of  $Mt$ HigA2 and compared it with the previously elucidated  $Mt$ HigA3 structure (Eickhoff *et al.*, 2009). Both share the same secondary-structure topology ( $\alpha 1$ – $\alpha 2$ – $\alpha 3$ – $\alpha 4$ – $\beta 1$ – $\beta 2$ ), with an apparently similar fold, but show a high backbone r.m.s.d. and a low *DALI* Z-score (2.6 Å and 8.4, respectively). The *DALI* server reveals other strong matches in type II TA systems: *E. coli* HipB and *V. vulnificus* transcription factor. Despite low structural sequence identity, they all share a HTH motif in

each monomer, but the  $\beta$ -lid is only preserved in *E. coli* HipB and  $Mt$ HigA3.

When the structures and sequences are aligned, the unique characteristics of  $Mt$ HigA2 are revealed. The first notable difference between the  $Mt$ HigA proteins and the other two proteins lies in the N-terminus. There are no corresponding sequences for this region in *E. coli* HipB and *V. vulnificus* transcription factor. Since the N-terminal residues of  $Mt$ HigA2 and  $Mt$ HigA3 are cleaved readily, the N-terminus of  $Mt$ HigA1 is also expected to be cleaved. The sequence homology is poor in the N-terminal region but the expected cleavage region shows similarity, with prominent charged residues. Therefore, in the absence of the toxin the N-terminus of  $Mt$ HigA adopts an intrinsically disordered form that is prone to spontaneous autocleavage. A similar characteristic is detected in *V. cholerae* HigA, showing the convergence of charged residues at the N-terminal expected cleavage site. Another distinctive feature is found in the HTH motif, which is responsible for nucleic acid binding. Previous structural studies of  $Mt$ HigA3 and *E. coli* HipB specified the DNA-interacting residues, which are predominantly positively charged amino acids such as Lys69 in  $Mt$ HigA3 and Lys38 in *E. coli* HipB. The side chains of both residues protrude to the surface to interact with the negatively charged phosphate or bases of DNA. When superimposed, all four structures show positively charged residues in each corresponding region, Lys40 in *V. vulnificus* transcription factor and Arg56 in  $Mt$ HigA2, implying charge conservation

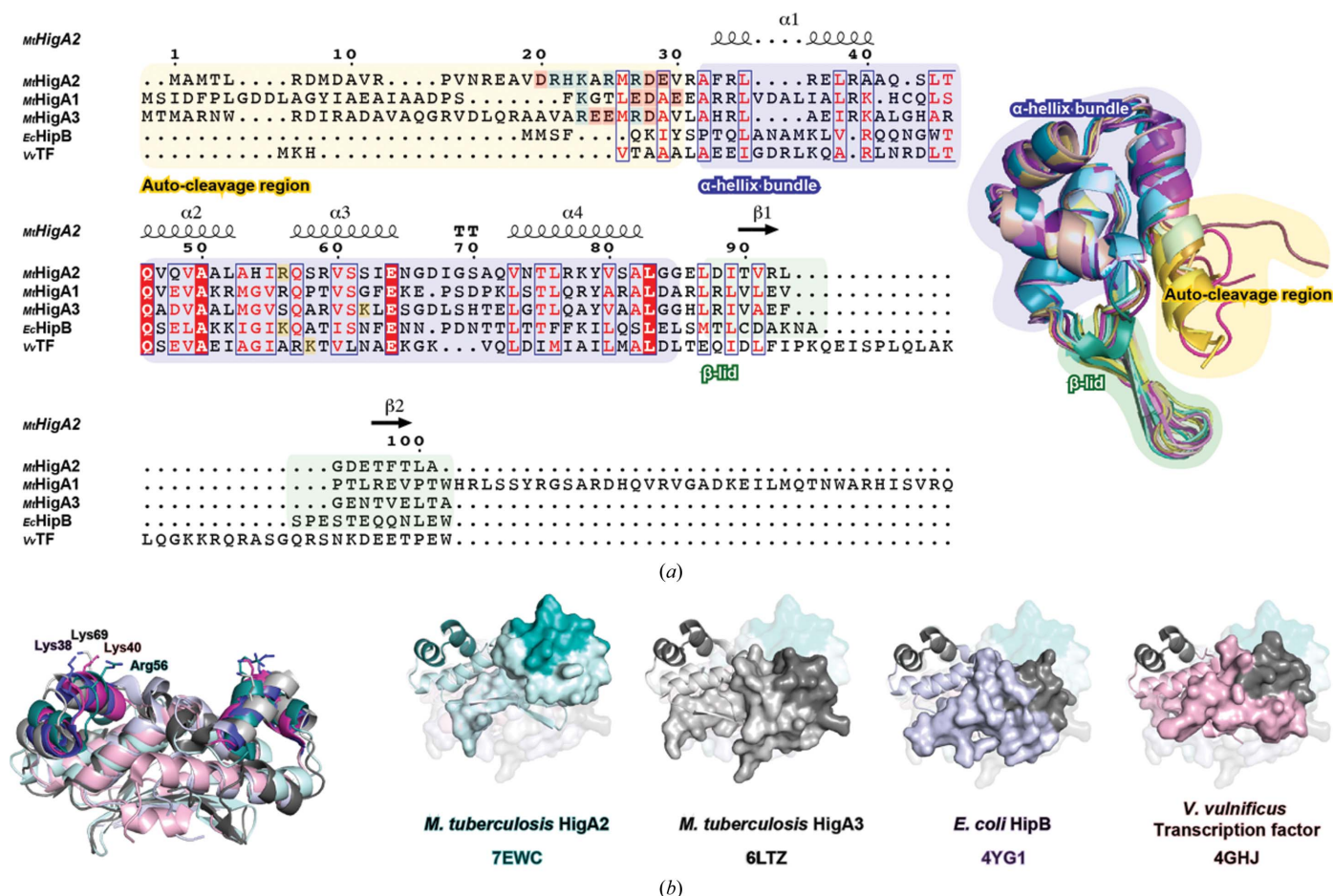


**Figure 2** Autocleavage of  $Mt$ HigA2. (a) Size-exclusion chromatography of  $Mt$ HigA2 before (top) and after (bottom) TEV cleavage. The major peak coloured yellow reduces substantially, and a small peak shown in green became the major peak after purification. (b) Schematic diagram of  $Mt$ HigA2 and the corresponding SDS-PAGE. The colours correspond to the components in the  $Mt$ HigA2 construct.

for DNA binding. However, low sequence identity in the HTH motif is likely to influence sequence specificity. The last characteristic feature is found in the C-terminus, with *Mt*HigA1 having an additional ~40 residues. Aside from the typical toxin–antitoxin systems *Mt*HigBA2 and *Mt*HigBA3, *Mt*HigBA1 is a member of a toxin–antitoxin–chaperone system. Rv1955, Rv1956 and Rv1957 perform these functions as HigB1 (toxin), HigA1 (antitoxin) and chaperone, respectively (Sala *et al.*, 2014). The crystal structure of the chaperone, Rv1957, was determined with the C-terminal peptide of *Mt*HigA1 (Rv1956), implying that the additional C-terminal residues of *Mt*HigA1 (Rv1956) may play a key role in stable complex formation with the chaperone (Guillet *et al.*, 2019). *Mt*HigA2 and *Mt*HigA3 lack C-terminal residues and do not possess a third gene in their respective operons, highlighting that an additional chaperone protein is not required for folding (Fig. 3).

### 3.4. The unique characteristics of *M. tuberculosis* HigA2

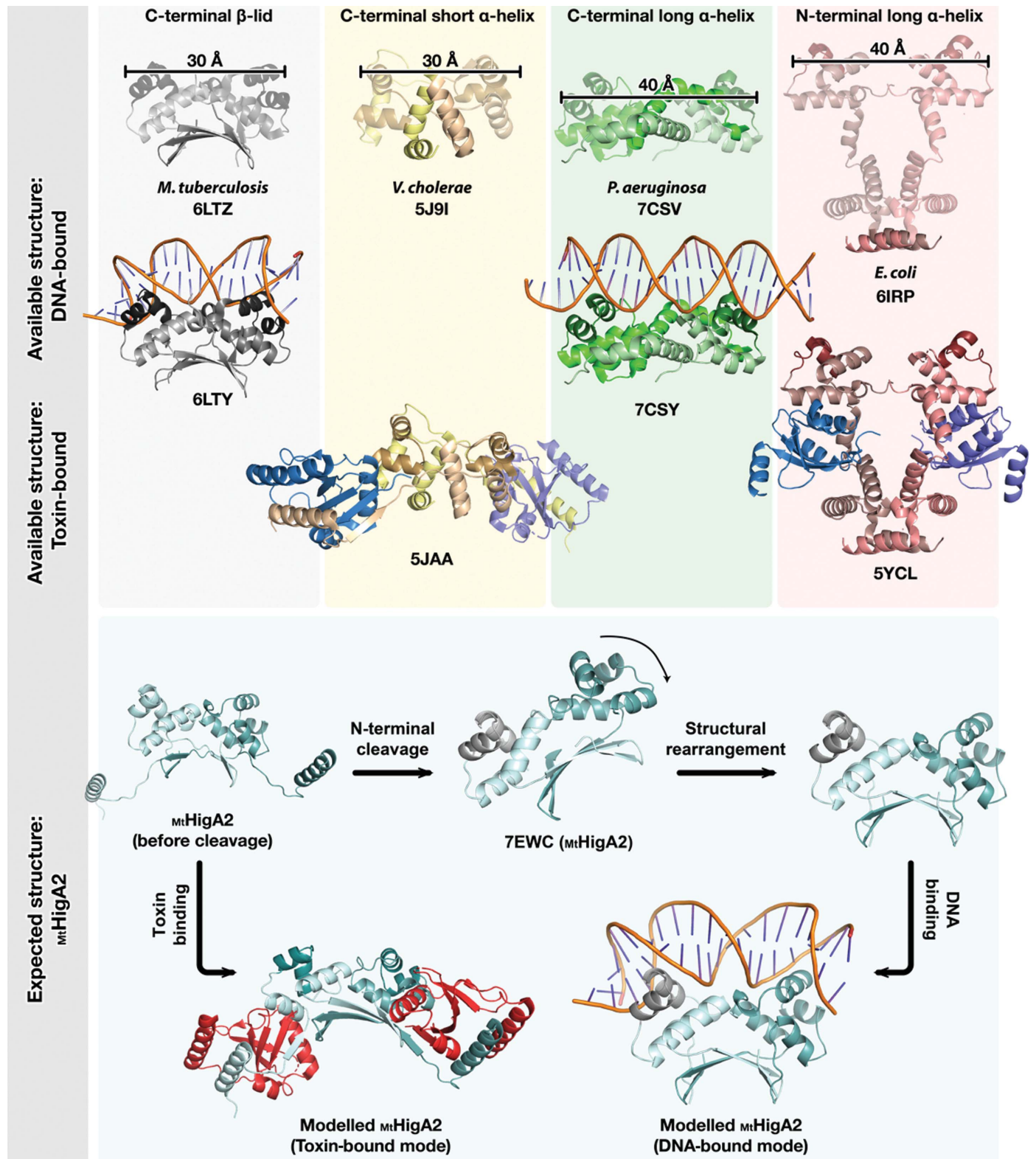
The human pathogen *M. tuberculosis* H37Rv contains 79 toxin–antitoxin pairs, whilst other mycobacteria possess fewer, implying a clinically relevant correlation. Of these 79, 38 are classified as type II, which includes the HigBA system. The canonical hierarchy for TA systems has the antitoxin gene located upstream of the toxin, likely as a regulatory control, but in the HigBA system this gene order is swapped. It is not known why the HigB toxin gene is the first gene of the operon, although the antitoxin does not possess its own promoter (Armalytė *et al.*, 2018; Park *et al.*, 2020). However, unlike the previous study on *Mt*HigBA3, the promoter region for *Mt*HigBA2 was not clearly defined. The *BPROM* tool failed to identify the  $\sigma_{70}$  binding site (Solovyev & Salamov, 2011). Therefore, four DNA sequences around the –10 box and the –35 box of *Mt*HigA2 and *Mt*HigBA2 were used in an interaction assay using EMSA, ITC and SEC, as in our previous



**Figure 3** Sequence and structure comparison of *Mt*HigA2. (a) Sequence comparison of *Mt*HigA2 with the related proteins *Mt*HigA3 (40% sequence identity), *Mt*HigA1 (28% sequence identity), *E. coli* HipB (19% sequence identity) and *V. vulnificus* transcription factor (22% sequence identity). Three parts of *Mt*HigA2 (the N-terminal autocleavage region,  $\alpha$ -helix bundle and C-terminal  $\beta$ -lid) are coloured with yellow, blue and green backgrounds. The four monomers from form I, the four monomers from form II and the five monomers from form III are aligned in the right panel. Each monomer shows a different state of the N-terminal residues, which implies that this region is intrinsically unstable and disordered. The charged *Mt*HigA N-terminal residues are coloured with light red and light blue backgrounds. The key amino acids for DNA interaction in *E. coli* HipB and *Mt*HigA3 are highlighted with a yellow background and the corresponding positive residues are coloured in the same way. The figure was constructed using *ESPrpt* (Robert & Gouet, 2014). (b) Superposition of *Mt*HigA2 with the structurally known proteins in (a). The superposition shows a similar fold, showing a conserved orientation of positively charged amino acids, which are coloured yellow in (a). However, when they are aligned based on HTH motifs, a unique linear coordination of *Mt*HigA2 is observed. PDB codes are shown below the structures.

study on  $M_t$ HigA3. However, binding was not detected for  $M_t$ HigA2 (Supplementary Fig. S2).

Although the  $M_t$ HigA2 dimer and  $M_t$ HigA3 dimer showed similarities in three-dimensional structures (backbone r.m.s.d.



**Figure 4** Expected binding mode of  $M_t$ HigA2. HigA shows diverse dimerization modes. The major contributor to dimerization is highlighted and HTH motifs are coloured in a darker shade. The distance between two HTH motifs is shown in Å. The DNA-bound structures are available for *M. tuberculosis* and *P. aeruginosa* and toxin-antitoxin complex structures are available for *V. cholerae* and *E. coli* HigA. From the comparison, we could model the DNA-binding mode and toxin-binding mode of  $M_t$ HigA2 (coloured in cyan). PDB codes are shown below the structures.

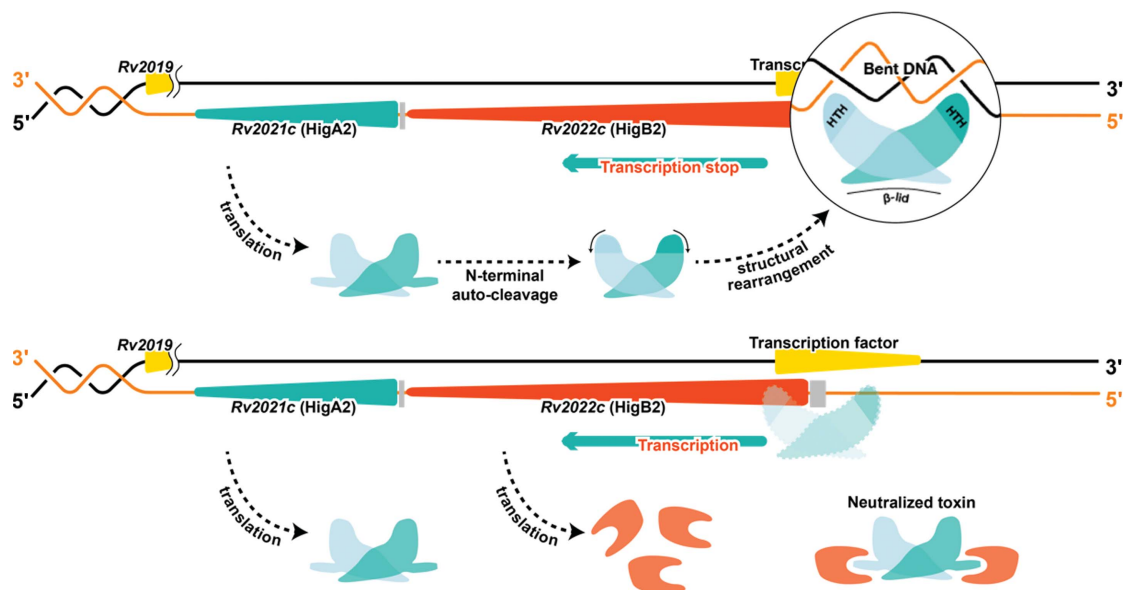


Figure 5

$M_t$ HigA2 operon and action of  $M_t$ HigA2. Based on our results, a schematic of the  $M_t$ HigBA system is suggested.  $M_t$ HigA2 is translated downstream of  $M_t$ HigB2. Upon translation, the N-terminus of  $M_t$ HigA2 is autocleaved and structural rearrangement occurs to bind DNA, thereby blocking additional expression of the HigBA operon. When the  $M_t$ HigB toxin is expressed,  $M_t$ HigA2 with an intact N-terminus binds to the toxin and neutralizes it.

of 2.6 Å),  $M_t$ HigA2 shows a major difference in the arrangement of the two HTH motifs from each monomer.  $M_t$ HigA3 showed an arched formation of HTH motifs, while  $M_t$ HigA2 has a linear arrangement (Park *et al.*, 2020). This linear configuration is unique compared with similar transcription factors [Fig. 3(b)]. From our study, the linear arrangement and negative result in DNA binding suggest that a rearrangement of the HTH motif is required for interaction with promoter DNA.

### 3.5. Suggested model for interaction of *M. tuberculosis* HigA2 with DNA or HigB2

The common structural characteristic of dimeric HigA antitoxins is that they bind to a specific promoter DNA using HTH motifs from each monomer. However, the HigA HTH motifs show various arrangements as a consequence of dimerization. While  $M_t$ HigA2 and  $M_t$ HigA3 use a C-terminal  $\beta$ -lid for dimerization, *P. aeruginosa* HigA and *P. vulgaris* HigA utilize a long  $\alpha$ -helix and *V. cholerae* HigA uses a short  $\alpha$ -helix at the C-terminus for dimerization (Song *et al.*, 2021; Schureck *et al.*, 2019; Hadži *et al.*, 2017). Another dimerization mode is found in *E. coli* HigA and *Shigella flexneri* HigA, showing close contact through the N-terminal  $\alpha$ -helix (Fig. 4). The dimerization determines the position of the HTH motifs, which recognize DNA major grooves as pairs. When the dimer is formed by a C-terminal long  $\alpha$ -helix or an N-terminal  $\alpha$ -helix, the distance between each HTH motif is  $\sim 40$  Å, whilst it is  $\sim 30$  Å for dimers that use a  $\beta$ -lid or a C-terminal short  $\alpha$ -helix. Although the number of antitoxin–DNA complexes in the PDB is low, the DNA-bending function is predictable from the distances between the two HTH motifs.

To identify the binding mode of  $M_t$ HigA2 to DNA, we performed docking simulations to the corresponding DNA

with  $M_t$ HigA2 in linear and bent conformations using *ClusPro* (Desta *et al.*, 2020). The linear  $M_t$ HigA2 model was chosen from our highest resolution structure (PDB entry 7ewc) and the bent conformation was modelled using  $M_t$ HigA3. For docking, the DNA was defined as the receptor, with  $M_t$ HigA2 as the ligand. Among 7000 rotations, the top ten lowest scoring results were identified and visually inspected in *PyMOL*. Docking trials failed when the linear  $M_t$ HigA2 was docked to linear DNA, which agrees with our experimental observations (EMSA, ITC and SEC). Docking highlights that  $M_t$ HigA2 binds to both bent and linear DNA upon structural rearrangement. Without the rearrangement of  $M_t$ HigA2, docking shows a disfavoured interaction with DNA as the HTH motifs remain exposed to the solvent area. When bent  $M_t$ HigA2 interacts with linear DNA, the distance between the HTH motifs is widened by a  $\beta$ -lid distortion (Supplementary Fig. S3). The docking experiments clearly indicate that structural rearrangement is necessary for interaction with DNA (Supplementary Fig. S3). We can also predict the  $M_t$ HigA2– $M_t$ HigB2 complex structure from *V. cholerae* HigA, because a cleavable N-terminus is surmised to be involved in the formation of a toxin–antitoxin complex. The structural diversity found in  $M_t$ HigA2 requires flexible dimerization, which is closely related to interaction with the DNA or toxin.

## 4. Discussion

Structural studies of  $M_t$ HigA2 confirmed that it has three regions: (i) a disordered N-terminus that is liable to cleavage, (ii) an  $\alpha$ -helix bundle containing an HTH motif for DNA binding and (iii) a  $\beta$ -lid for dimerization. The N-terminal part is known to be responsible for interaction with the toxin, but when the antitoxin is solely expressed this part spontaneously autocleaves, showing TA-system regulation at the protein



level. In the absence of the N-terminal part,  $Mt$ HigA2 forms a stable dimer both in solution and in the crystal structure on account of the extensive hydrogen bonds observed in the  $\beta$ -lid. The three different crystal structures in our study reveal the flexibility in dimerization mode while anchored by the  $\beta$ -lid.

Antitoxins function to neutralize toxin activity either by direct toxin binding or by repression of expression (negative autoregulation). The antitoxin  $Mt$ HigA2 has two binding partners: DNA and the toxin  $Mt$ HigB2. Although  $Mt$ HigA2 was expected to bind DNA using HTH motifs, we obtained negative results in interaction studies. To fit two HTH motifs to the DNA major grooves, the structure should cover the length between two major grooves, which is 34 Å. However, the distance between two HTH motifs in  $Mt$ HigA2 is 30 Å, meaning that DNA bending is indispensable. Despite this, the HTH motifs in  $Mt$ HigA2 arrange linearly, implying that structural rearrangement is required for DNA binding. Upon DNA binding, the expression of the toxin–antitoxin operon is downregulated. Another binding partner is the  $Mt$ HigB2 toxin. The N-terminal antitoxin has an autocleavable ~25 residues and this part is predicted to bind the C-terminus of the toxin. When  $Mt$ HigB2 is translated because of weak  $Mt$ HigA2–DNA binding, the toxin action can still be blocked by the formation of an  $Mt$ HigBA complex with successively translated  $Mt$ HigA2. The toxin  $Mt$ HigB2 inhibits protein synthesis by mRNA cleavage and causes cell-growth arrest and cell death. Therefore, the antitoxin  $Mt$ HigA2 has a double protection system to neutralize the toxin (Fig. 5). Further studies of DNA- or toxin-bound structures remain to be performed. However, this study contributes various methods by which the antitoxin may control its function, which include autocleavage or structural flexibility. This will provide useful information for future studies of gene-regulating proteins.

## Acknowledgements

We thank the beamline staff members at Pohang Light Source (BL-5C), Korea and Diamond Light Source (I04), UK.

## Funding information

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1G1A1100821) and ‘Leaders in Industry-university Cooperation +’ Project supported by the Ministry of Education and NRF.

## References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.  
 Armalytė, J., Jurėnas, D., Krasauskas, R., Čepauskas, A. & Sužiedėlienė, E. (2018). *Front. Microbiol.* **9**, 732.  
 Aslanidis, C. & de Jong, P. J. (1990). *Nucleic Acids Res.* **18**, 6069–6074.  
 Brennan, R. G. & Matthews, B. W. (1989). *J. Biol. Chem.* **264**, 1903–1906.  
 Desta, I. T., Porter, K. A., Xia, B., Kozakov, D. & Vajda, S. (2020). *Structure*, **28**, 1071–1081.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K. & Fox, P. T. (2009). *Hum. Brain Mapp.* **30**, 2907–2926.  
 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.  
 Eschenfeldt, W. H., Lucy, S., Millard, C. S., Joachimiak, A. & Mark, I. D. (2009). *Methods Mol. Biol.* **498**, 105–115.  
 Fraikin, N., Goormaghtigh, F. & Van Melderen, L. (2020). *J. Bacteriol.* **202**, e00763-19.  
 Fukunaga, R., Glaziou, P., Harris, J. B., Date, A., Floyd, K. & Kasaeva, T. (2021). *MMWR Morb. Mortal. Wkly Rep.* **70**, 427–430.  
 Ghafourian, S., Raftari, M., Sadeghifard, N. & Sekawi, Z. (2014). *Curr. Issues Mol. Biol.* **16**, 9–14.  
 Guillet, V., Bordes, P., Bon, C., Marcoux, J., Gervais, V., Sala, A. J., Dos Reis, S., Slama, N., Mares-Mejía, I., Cirinesi, A. M., Maveyraud, L., Genevaux, P. & Mourey, L. (2019). *Nat. Commun.* **10**, 782.  
 Gupta, A. (2009). *FEMS Microbiol. Lett.* **290**, 45–53.  
 Hadži, S., Garcia-Pino, A., Haesaerts, S., Jurėnas, D., Gerdes, K., Lah, J. & Loris, R. (2017). *Nucleic Acids Res.* **45**, 4972–4983.  
 Jain, S. K., Hernandez-Abanto, S. M., Cheng, Q. J., Singh, P., Ly, L. H., Klinkenberg, L. G., Morrison, N. E., Converse, P. J., Nuermberger, E., Grosset, J., McMurray, D. N., Karakousis, P. C., Lamichhane, G. & Bishai, W. R. (2007). *J. Infect. Dis.* **195**, 1634–1642.  
 Jeong, J.-Y., Yim, H.-S., Ryu, J.-Y., Lee, H. S., Lee, J.-H., Seen, D.-S. & Kang, S. G. (2012). *Appl. Environ. Microbiol.* **78**, 5440–5443.  
 Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.  
 Kang, S.-M., Jin, C., Kim, D.-H., Park, S. J., Han, S.-W. & Lee, B.-J. (2020). *FEBS J.* **288**, 1546–1564.  
 Kędzierska, B. & Hayes, F. (2016). *Molecules*, **21**, 790.  
 Keegan, R. M., McNicholas, S. J., Thomas, J. M. H., Simpkin, A. J., Simkovic, F., Uski, V., Ballard, C. C., Winn, M. D., Wilson, K. S. & Rigden, D. J. (2018). *Acta Cryst.* **D74**, 167–182.  
 Kim, H. J. (2020). *Biochem. Biophys. Res. Commun.* **533**, 118–124.  
 Kim, H. J., Kwon, A.-R. & Lee, B.-J. (2018). *Biosci. Rep.* **38**, BSR20180768.  
 Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nat. Protoc.* **3**, 1171–1179.  
 Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.  
 Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. & Russell, R. B. (2003). *Structure*, **11**, 1453–1459.  
 Liu, Y., Zheng, W., Zhang, W., Chen, N., Liu, Y., Chen, L., Zhou, X., Chen, X., Zheng, H. & Li, X. (2015). *Chem. Sci.* **6**, 745–751.  
 Luscombe, N. M., Austin, S. E., Berman, H. M. & Thornton, J. M. (2000). *Genome Biol.* **1**, REVIEWS001.  
 Mabhula, A. & Singh, V. (2019). *Med. Chem. Commun.* **10**, 1342–1360.  
 Maisonneuve, E. & Gerdes, K. (2014). *Cell*, **157**, 539–548.  
 Matthews, B. W., Ohlendorf, D. H., Anderson, W. F. & Takeda, Y. (1982). *Proc. Natl Acad. Sci. USA*, **79**, 1428–1432.  
 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.  
 McGeehan, J. E., Streeter, S. D., Papapanagiotou, I., Fox, G. C. & Kneale, G. G. (2005). *J. Mol. Biol.* **346**, 689–701.  
 Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* **D67**, 355–367.  
 Park, J.-Y., Kim, H. J., Pathak, C., Yoon, H.-J., Kim, D.-H., Park, S. J. & Lee, B.-J. (2020). *IUCrJ*, **7**, 748–760.  
 Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. (1994). *Anal. Biochem.* **216**, 413–417.

- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Ramage, H. R., Connolly, L. E. & Cox, J. S. (2009). *PLoS Genet.* **5**, e1000767.
- Ren, J., Sainsbury, S., Nettleship, J. E., Saunders, N. J. & Owens, R. J. (2010). *Proteins*, **78**, 1798–1802.
- Robert, X. & Gouet, P. (2014). *Nucleic Acids Res.* **42**, W320–W324.
- Sala, A., Bordes, P. & Genevoux, P. (2014). *Toxins*, **6**, 1002–1020.
- Schumacher, M. A., Piro, K. M., Xu, W., Hansen, S., Lewis, K. & Brennan, R. G. (2009). *Science*, **323**, 396–401.
- Schureck, M. A., Meisner, J., Hoffer, E. D., Wang, D., Onuoha, N., Ei Cho, S., Lollar, P. & Dunham, C. M. (2019). *Mol. Microbiol.* **111**, 1449–1462.
- Schureck, M. A., Repack, A., Miles, S. J., Marquez, J. & Dunham, C. M. (2016). *Nucleic Acids Res.* **44**, 7944–7953.
- Solovyev, V. & Salamov, A. (2011). *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, edited by R. W. Li, pp. 61–78. Hauppauge: Nova Science Publishers.
- Song, Y., Luo, G., Zhu, Y., Li, T., Li, C., He, L., Zhao, N., Zhao, C., Yang, J., Huang, Q., Mu, X., Tang, X., Kang, M., Wu, S., He, Y. & Bao, R. (2021). *Environ. Microbiol.* **23**, 1541–1558.
- Stewart, G. R., Patel, J., Robertson, B. D., Rae, A. & Young, D. B. (2005). *PLoS Pathog.* **1**, e33.
- Villard, J. (2004). *Swiss Med. Wkly*, **134**, 571–579.
- Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. (2011). *Acta Cryst. D* **67**, 235–242.
- Wintjens, R. & Rooman, M. (1996). *J. Mol. Biol.* **262**, 294–313.