



Computational prediction and validation of specific EmBR binding site on PknH

Insung Na^{a,1,2}, Huanqin Dai^{a,c,2}, Hantian Li^{c,g}, Anvita Gupta^d, David Kreda^e, Powell Zhang^h, Xiangyin Chen^f, Lixin Zhang^{f,**}, Gil Alterovitz^{b,*}

^a Computational Health Informatics Program, Boston Children's Hospital/Harvard Medical School, Boston, MA 02115, USA

^b Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA 02115, USA

^c State Key Laboratory of Mycology, Institute of Microbiology, Beijing, 100101 China

^d Department of Computer Sciences, Stanford University, Stanford CA 94305, USA

^e Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

^f State Key Laboratory of Bioreactor Engineering, East China University of Science and Technology, Shanghai, China

^g Department of Biology, The Johns Hopkins University, Baltimore, MD 21218, USA

^h Lexington High School, Boston, MA 02115, USA

ARTICLE INFO

Keywords:

Disorder-to-order transition
Protein intrinsic disorder
Binding site prediction
Drug resistance
Molecular simulation

ABSTRACT

Tuberculosis drug resistance continues to threaten global health but the underline molecular mechanisms are not clear. Ethambutol (EMB), one of the well-known first-line drugs in tuberculosis treatment is, unfortunately, not free from drug resistance problems. Genomic studies have shown that some genetic mutations in *Mycobacterium tuberculosis* (Mtb) EmBR, and EmbC/A/B genes cause EMB resistance. EmBR-PknH pair controls embC/A/B operon, which encodes EmbC/A/B genes, and EMB interacts with EmbA/B proteins. However, the EmBR binding site on PknH was unknown. We conducted molecular simulation on the EmBR-peptides binding structures and discovered phosphorylated PknH 273–280 (N'-HEALS^PDPD-C') makes β strand with the EmBR FHA domain, as β -MoRF (MoRF; molecular recognition feature) does at its binding site. Hydrogen bond number analysis also supported the peptides' β -MoRF forming activity at the EmBR FHA domain. Also, we discovered that previously known phosphorylation residues might have their chronological order according to the phosphorylation status. The discovery validated that *Mtb* PknH 273–280 (N'-HEALS^PDPD-C') has reliable EmBR binding affinity. This approach is revolutionary in the computer-aided drug discovery field, because it is the first trial to discover the protein-protein interaction site, and find binding partner in nature from this site.

1. Introduction

Tuberculosis (TB) remains a world-threatening disease, frequently listed in the top 10 causes of human death. Certainly, WHO and CDC announce 9–10 million new tuberculosis cases each year, and 1–2 million cases of tuberculosis-related death [1]. More seriously, the current COVID-19 pandemic could cause severe reductions in timely diagnosis and treatment of new TB cases, further increasing TB burden [2,3]. There are currently four types of first-line drugs (Rifampicin, Isoniazid, Pyrazinamide, and Ethambutol) to treat viral tuberculosis.

These drugs act through metabolism inhibition to interrupt the activity of the pathogen *Mycobacterium tuberculosis* (Mtb). In combatting TB today, the lack of newer and more efficacious drugs is a serious problem. Much effort has been directed towards discovering new anti-TB drugs employing target or cell-based approaches [4–7]. However, about 4% of total active TB cases continuously show drug-resistance to one or more of these first-line drugs due to genetic variation, which has become an urgent problem to solve [1]. Ethambutol was discovered in 1961 and functions through the blockage of Mtb membrane synthesis through the inhibition of arabinosyl transferases (EmbC/A/B), which are encoded in

Peer review under responsibility of KeAi Communications Co., Ltd.

* Corresponding author. Department of Medicine, Brigham and Women's Hospital/Harvard Medical School, Boston, MA 02115, USA.

** Corresponding author.

E-mail addresses: lxzhang@ecust.edu.cn (L. Zhang), ga@alum.mit.edu (G. Alterovitz).

¹ Current address: HITS Inc., 124, Teheran-ro, Gangnam-gu, Seoul 06234, South Korea.

² These authors contributed equally to this work.

<https://doi.org/10.1016/j.synbio.2021.11.006>

Received 27 June 2021; Received in revised form 9 October 2021; Accepted 9 November 2021

2405-805X/© 2021 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC

BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

the embCAB operon [8–11].

Genomic studies have shown that some genetic mutations in *Mycobacterium tuberculosis* (Mtb) EmbR, and EmbC/A/B genes cause EMB resistance [12,13]. EmbR – PknH controls the embCAB operon [12,13], and PknH phosphorylates EmbR at its C-terminal forkhead-associated (FHA) domain [14]. EmbR structure was discovered in complex form with the antibiotic peptide *Saccharomyces cerevisiae* (Yeast) Rad9 188–195 (Yeast Rad9 188–195, N'-SLEVTEAD-C', Protein Data Bank ID 2FF4), but it failed to show specific EmbR binding site on PknH [15]. PknH is one of the 11 serine/threonine protein kinases (STPKs) in Mtb [16,17]. As eukaryotic cells have phosphorylation cascade signal transduction pathways, 11 STPKs of Mtb phosphorylate substrate proteins and are involved in diverse cellular mechanisms. EmbR is one of the substrates of PknH [17]. EmbR FHA domain is the phosphorylation site, and PknH has auto-phosphorylation activity [14,18]. Nevertheless, EmbR binding site or auto-phosphorylation site on PknH is unknown. Proteomic study discovered 500 phosphorylation events in 310 proteins in Mtb [19]. However, there was no discovery of phosphopeptides originated from EmbR FHA domain or PknH auto-phosphorylation sites in the research. FHA domain is discovered from prokaryotes and eukaryotes, and this domain is well known for its specific interaction with phosphorylated threonine, such as, pTXXD, pTXXI/L motifs [16,18,20]. Recent review implies that not only the threonine, but also the surrounding residues are important to interact with FHA domain, because these surrounding residues provide additional interaction surfaces [20]. An EmbR homologous protein, EmbR2, was discovered from Mtb CDC1551 strain [21]. EmbR2 also has FHA domain, but, PknH does not phosphorylate EmbR2. Instead, EmbR2 blocks PknH auto-phosphorylation activity. Thus, it plays a role as EmbR – PknH pair control factor [21]. The Kd values of EmbR2 – threonine-containing phosphopeptides from PknH, and Rad9, and the values ranged from 12 to 21 μ M.

Intrinsically disordered proteins (IDP) are increasingly understood to play diverse biological roles, ranging from molecular signaling to the formation of membraneless organelles. Many classes of IDPs have been identified and their biological roles were reported. More than 30% of eukaryotic proteins have disordered regions, and these IDPs are often involved in key biological processes such as molecular recognition, cell cycle regulation, and cellular signaling [22]. Because of their flexibility, IDPs can bind to many partner proteins and some of them are hubs of protein-protein interaction networks within the cell [22]. The last two decades have seen an increase in evidence for the involvement of IDPs and IDRs in a variety of essential biological processes and molecular functions. However, their high degree of disorder makes them more challenging to study using the structural biology techniques. Nevertheless, molecular simulations and other computational researches are therefore a natural complement to study their properties. The diverse time and length scales relevant to the roles played by IDPs require flexibility in the techniques applied [23]. Intrinsically disordered regions (IDRs) of proteins are sometimes involved in molecular binding or recognition. IDRs usually compose at least 30 amino acids and some of the IDRs are involved in disorder-to-order transition upon binding, named as MoRF (molecular recognition feature) [24]. MoRFs allow diverse interactions of IDPs, and there are two possible forms; α -MoRF, and β -MoRF [24,25]. α -MoRF is defined by α -helix former IDR, and β -MoRF is defined by β -strand former IDR. MoRFs are utilizable to discover novel drugs by mimicking MoRFs to inhibit its partner target. The IDP interactions show promise as drug targets because the binding interface between an IDP and its partner, MoRF, is not large and flat, as the binding interface is between two ordered proteins [26–30]. Additionally, that binding experiment showed relatively low binding affinity between protein and peptide compared to usual bioactive bindings. It could be explained in relation to the drug lead (RG7112) discovery, which mimicked the IDRs in the binding region of p53 toward MDM2 [26–30]. p53 plays an important role as a cell cycle checkpoint protein in cell physiology, but, mutations cause more than fifty percent of

tumors [23]. The p53 α -MoRF mimicking aimed to inhibit p53 – MDM2 interaction, and expectedly greater tumor suppression with enhanced p53 activities, which is free from ubiquitin by MDM2. The lead was proven to show anti-leukemia activity through clinical trial phase I [31].

We tried to decipher EmbR binding site on PknH, focusing on its disorder-to-order transition. To achieve this goal, we analyzed biochemical similarities with already discovered Yeast Rad9 188–195 in the complex form with EmbR (PDB ID 2FF4) [15]. Applying homology modeling, molecular docking, and molecular simulation, we discovered a specific peptide for study. This strategy may be further developed as a virtual biomimicry approach to discover novel drugs, using protein structures to yield interaction predictions [32].

2. Materials & Methods

2.1. Bioinformatics discovery of specific EmbR binding IDR site of PknH

Saccharomyces cerevisiae (Yeast) Rad9 (UniProt ID P14737), and *Mycobacterium tuberculosis* (Mtb) H37Rv strain PknH cytoplasmic part (UniProt ID P9WI71 residue 1–403) protein intrinsic disorder scores were measured using PONDR® FIT, PONDR® VL3, and PONDR® VSL2 [1–4] based on their sequences from UniProt 2017_11 [5]. The Protein Model Portal linked to UniProt 2017_11 showed PknH homology model based on PDB ID 2H34 [6], which composes Mtb PknE structure of sequence identity 70% with PknH residue 14–277. To prepare new homology model of PknH cytoplasmic part (residue 1–403), PDB ID 2H34 structure was used as template. Finally, we applied MODELLER 9.17 [7] to prepare the PknH cytoplasmic part structure.

Structure of EmbR – Yeast Rad9 (PDB ID 2FF4) [8] composes EmbR C-terminal forkhead-associated (FHA) domain at its binding site, thus, already known FHA domain containing complex structure PDB ID 3OUN (Mtb 39.8 and Rv3910) [9] was used for structural alignment with each of EmbR (from PDB ID 2FF4) structure and PknH homology modeling result in corresponding order. Thus, we could prepare EmbR – PknH cytoplasmic part complex structure. Based on structural comparison with PDB ID 2FF4 Yeast Rad9 188–195 (N'-SLEVTEAD-C'), PknH 273–280 (N'-HEALSDPD-C') was taken as the potential specific EmbR binding site.

Finally, we could find similar protein intrinsic disorder pattern at the predicted site on PknH 273–280, and Yeast Rad9 188–195 using visual comparison of their plots from R 3.3.3 [10].

2.2. Molecular simulation

We obtained Mtb PknH 273–280 (N'-HEALSDPD-C') structure aligned on PDB ID 3OUN (Mtb 39.8 and Rv3910) after cytosolic PknH (residues 1–403) homology modeling (Fig. 1A). Phospho-serine was modelled using PyMol 1.7. EmbR structure from PDB ID 2FF4 was taken after structural alignment on PDB ID 3OUN. Besides, because of threonine importance at the binding site [11], we designed a new peptide based on PknH 273–280, and named NEW (N'-HEALDTPD-C') and NEWp (N'-HEAL^PDPD-C'). Thus, we could prepare 6 molecular simulation systems; EmbR-PknHp (EmbR - PknH 273–280 with phosphorylation at S277, N'-HEAL^PDPD-C'), EmbR-PknH (EmbR - PknH 273–280 without phosphorylation at S277, N'-HEALSDPD-C'), EmbR-NEWp (EmbR - N'-HEAL^PDPD-C'), EmbR-NEW (N'-HEALDTPD-C') from homology modeling, EmbR-Rad9p (EmbR - Yeast Rad9 188–195 with phosphorylation at T192, N'-SLEV^PEAD-C'), and EmbR-Rad9 (EmbR - Yeast Rad9 188–195 without phosphorylation at T192, N'-SLEVTEAD-C') from PDB ID 2FF4. We prepared these structures using PyMol 1.7.

Gromacs 2016.4 single precision installed with FFTW 3.3.7 library, and AMBER99SB-ILDN force field was used for molecular simulation [12–14]. Truncated octahedron box with a periodic boundary condition was constructed with a buffering area thickness of 1.2 nm. For explicit water models in MD simulation systems, 55.6 M water molecules were

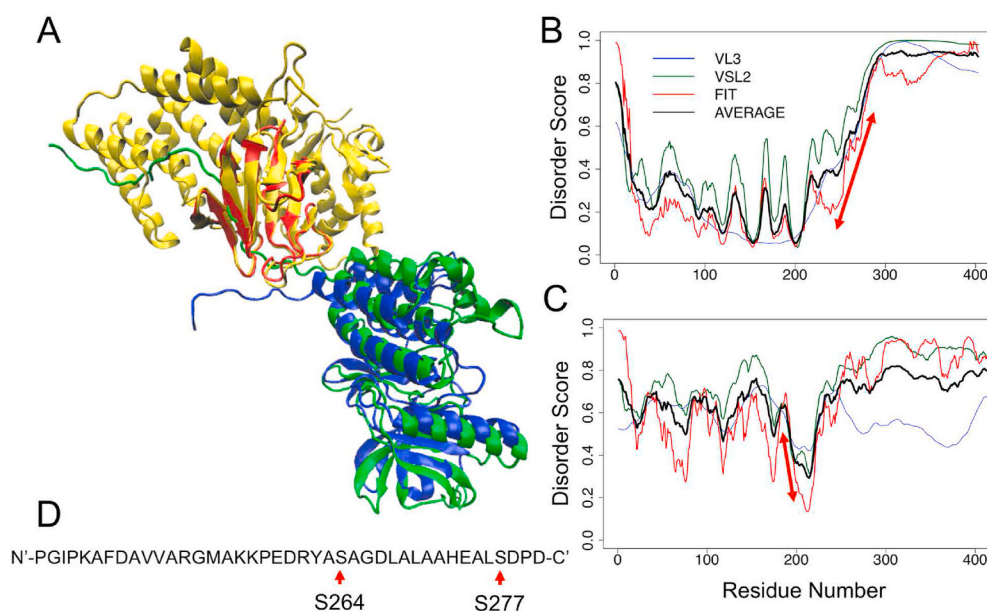


Fig. 1. Prediction of specific *M. tuberculosis* EmBR binding site of PknH. **A:** Structural alignment result. Green: Cytoplasmic PknH homology modeling result, Yellow: EmBR from PDB ID 2FF4, Blue: Rv3910 from PDB ID 3OUN, Red: 39.8 from PDB ID 3OUN. **B – C:** Protein intrinsic disorder score (PONDR predictors), **B:** PknH 1–403, **C:** Yeast Rad9 1–403, legend is same with **B**. **D:** Yeast Rad9 240–280 sequence. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

constructed. Then, ~100 mM of Na⁺ and Cl⁻ ions were added to the systems. The number of ions was adjusted to neutralize the system. Energy minimization using the steepest descent method was performed until the maximum force was less than 1 kJ mol⁻¹ nm⁻¹.

Canonical (NVT) ensemble MD simulations ($\Delta t = 1$ fs each step, and 100,000 steps for each temperature 100ps) were performed for heating from 100K to 310K, increasing 10K every 100ps. A position restraint constant of 1000 kJ mol⁻¹ nm⁻² was applied to all heavy atoms, and Nose-Hoover thermostat [15,16] was used to control temperature. The bond lengths with heavy atoms were constrained using the LINCS method [17]. For long-range electrostatics calculations, the Particle Mesh Ewald method (PME) [18] was used. The Verlet cut-off scheme [19] was used to update neighbor list every 10fs. The electrostatic cut-off, and van der Waals cutoff were set to be 1.2 nm.

Isothermal-isobaric (NPT) ensemble MD simulations of 5 ns ($\Delta t = 1$ fs each step, and 500,000 steps for each position restraint constant 500ps) were conducted for equilibration with exponentially decreasing serial position restraint constants at heavy atoms in the range between 1000 kJ mol⁻¹ nm⁻², and 1.953 kJ mol⁻¹ nm⁻². Parrinello-Rahman barostat [20,21] was applied to maintain pressure of the system, 1 bar. Other conditions were kept from heating step. NPT ensemble MD simulations of 50 ns ($\Delta t = 1$ fs each step) in three different temperatures (290K, 300K, and 310K) were continued without position restraint for production, and every 10ps structure was composed for further analysis from each set.

EmBR apo, EmBR – Rad9p, EmBR – Rad9, EmBR – NEWp, EmBR – NEW, EmBR – PknHp, EmBR – PknH 50ns trajectories in 290K, 300K, and 310K were analyzed. We measured number of hydrogen bond between EmBR, and its partner (Hydrogen bond: distance between donor and acceptor 3.5 Å, acceptor – donor – hydrogen angle 30°). We measured radius of gyration, and root-mean-squared deviation (RMSD) of heavy atoms, then plotted using the Python code obtained from Strodel lab (http://www.strodel.info/index_files/lecture/generateFES.py) to obtain ΔG value, which represents relative structural population disposition in two dimensions (radius of gyration, and RMSD) graph.

2.3. Binding validation of peptides in vitro

The binding affinities of peptides for the soluble EmBR protein were measured by biolayer interferometry on an Octet RED96 System, which monitors interference of light reflected from two sources (an internal

reflection surface and the liquid/solid interface of a fiber optic sensor) to measure the rate of molecules binding to the biosensor surface. Peptides were synthesized and biotinylated at the NH₂-terminus by Shanghai Shenggong, purified to >95% by HPLC. Biotinylated peptides were loaded onto Streptavidin (SA) biosensors at empirically determined concentrations. All affinity measurements were carried out in Kinetics Buffer (KB) (1X PBS, 0.002% Tween-20, and 0.01% BSA) at 25 °C. The biosensors were pre-equilibrated in KB for 120s, loaded with a biotinylated peptide at optimal concentrations and times, quenched with 10 µg/ml biocytin in KB for 180s, brought to baseline in KB for 120s and transferred to wells containing purified EmBR (0–400 µM). Multiple negative controls were run for each peptide including association buffer only (no EmBR) or with BSA replacing EmBR. Binding kinetics were calculated using the FortéBio Data Analysis v7.1 software.

3. Results

3.1. Prediction of EmBR binding site on PknH

According to the previous *Mtb* EmBR – PknH interaction studies [12, 14] and intrinsically disordered protein region (IDR) involvement in EmBR – Yeast Rad9 188–195 (N'-SLEVTEAD-C') binding data [15], we hypothesized that protein intrinsic disorder similarity between *Mtb* PknH and Yeast Rad9 would lead to a possible binding site on PknH. According to *Mtb* FHA domain-containing protein 39.8 - Rv3910 complex (PDB ID 3OUN) [33], EmBR (PDB ID 2FF4) was structurally aligned with 39.8, and the cytoplasmic part of PknH (residues 1–403) was aligned with Rv3910 after homology modeling. Structural alignment showed a potential EmBR binding site of PknH on its 240–280 residues with loop conformation as in Rv3910 (Fig. 1 A).

Protein intrinsic disorder score comparison between Yeast Rad9 and *Mtb* PknH showed matching opposite patterns, which have strong positive or negative slope. In detail, Yeast Rad9 188–195 (where the structure was shown in PDB ID 2FF4) was embedded in the strong negative slope between 186 and 214, and *Mtb* PknH 240–280 (Homology modeling prediction binding site) showed strong positive slope (Fig. 1B–C). Nevertheless, their sequences have similar chemical properties of hydrophobicity and charge; Yeast Rad9 188–195 (N'-SLEVTEAD-C') has a hydroxyl group containing T192 at its fifth position surrounded by three hydrophobic residues (L189, V191, A194). This region also has three other negatively charged residues (E190, E193, D195). Importantly, T192 can be phosphorylated [15]. *Mtb* PknH

240–280 doesn't contain threonine, but serine appears twice at positions 264 and 277 (Fig. 1 D). Between two serine residues, only S277 is surrounded by three hydrophobic residues (A275, L276, P279) and three other negatively charged residues (E274, D278, D280). Thus, we selected *Mtb* PknH 273–280 (N'-HEALSDPD-C') as the potential specific EmBR binding IDR site of PknH.

3.2. EmBR FHA domain two fingers hold phosphopeptides

After heating, equilibration, and production runs of molecular simulation for 6 systems as described in Materials & Methods (Rad9p, Rad9, PknHp, PknH, NEWp, NEW), we checked their continuous binding to the EmBR FHA domain. Although EmBR – NEW didn't show continuous binding at 300K, the other 5 systems showed continuous binding between EmBR and the peptides.

Given the functional importance of phosphorylation, we targeted EmBR FHA domain as IDR binding partner [14]. At the beginning of production runs, we discovered structural differences between phosphopeptide bound FHA domains and non-phosphopeptide bound FHA

domains (Fig. 2A). According to surface visualizations of FHA – peptide complex forms, phosphopeptide bound FHA domains commonly showed finger-like structures. The fifth amino acid phosphoryl groups were commonly coordinated between two fingers in three systems (Fig. 2A–b, d, f). After structural investigation, it turned out that the two fingers consisted of three residues; S326, R327, and S347. Hydrogen bond formation capacity of phospho-threonine (T192 at EmBR - Yeast Rad9 188–195, N'-SLEVT^PEAD-C') has been reported in PDB ID 2FF4 (Fig. 2B) [15]. Hence, we compared the positive control (PDB ID 2FF4) structures before equilibration (Fig. 2B–a) and after the equilibration (Fig. 2B–b). Positive controls in two conditions showed FHA domain fingers, and these formed 6 hydrogen bonds with the phosphoryl group at the fifth residue of the phosphopeptides. Between the two structures, the positive control after equilibration showed shorter hydrogen bond distances and tighter finger structures (Fig. 2B–a). The hydrogen bonds at the fifth residue contributed to the stable binding between FHA domain, and corresponding peptides during molecular simulation. However, EmBR FHA domains with non-phosphopeptides didn't show finger structures (Fig. 2A–c, e).

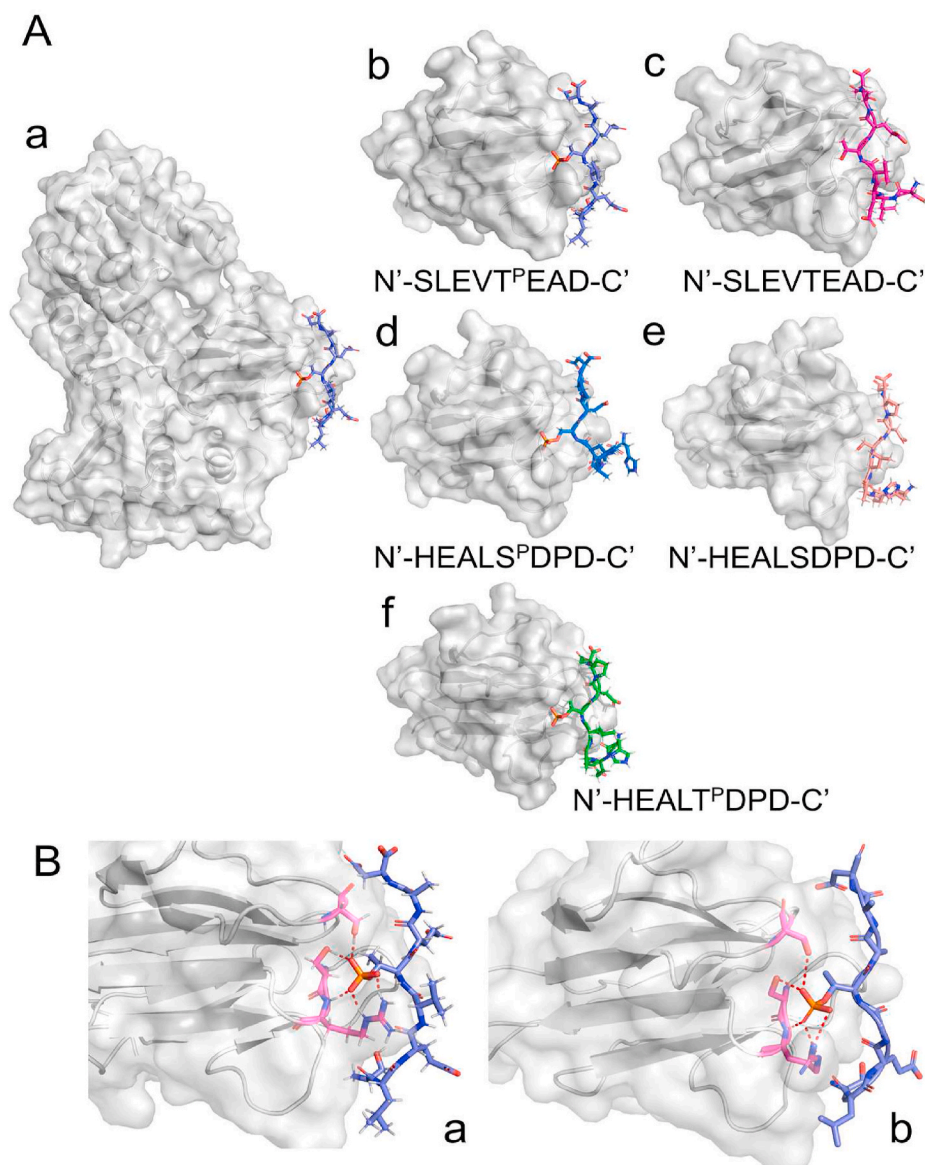


Fig. 2. EmBR FHA domain – peptide structures after equilibration. A: EmBR whole protein – Rad9p (positive control) bound form, and 5 continuous binding shown FHA domain – peptide complexes (b–f) b; Rad9p, c; Rad9, d; PknHp, e; PknH, f; NEWp, B: FHA domain – Rad9p a; (positive control) Hydrogen bonds after equilibration, b; Original position, PDB ID 2FF4 [15].

The FHA domain (EmBR 308–357 according to UniProt 2019_03) is important for the phosphorylation with PknH [14]. Mutations at three residues (R312, S326, N348) hinder phosphorylation [14]. Among these three residues, S326 only functioned as a hydrogen bond donor after equilibration of the positive control, and at the original position (PDB ID 2FF4) [15].

3.3. β -strand range of backbone torsional angles (φ , and ψ) of phosphopeptides in complexes

We focused on the backbone structural changes of peptides in 5 systems (Fig. 3). Due to the disorder-to-order transition characteristic of molecular recognition feature (MoRF) in contact region (α -MoRF, and β -MoRF) [24], we hypothesized that the backbone torsional angles (φ , and ψ) of the peptides would be distributed inside the α -helix or β -strand range in Ramachandran plot [34].

Before the torsional angle measurement, we confirmed the structural stability of the peptides in complex forms. When we compared their energy minimized structures, the heavy atom root-mean-squared deviation (RMSD) ranged between 0 and 5 Å, and the radius of gyration (RofG) ranged between 5 and 10 Å. PknHp, and PknH showed relatively more dispersed patterns than other peptides. However, these values are still incorporated in the RMSD range from 0 to 5 Å and the RofG 5–10 Å as shown in Fig. 3A.

To further investigate the structural changes of peptides in molecular simulation trajectories, we measured φ and ψ angles of each amino acid of the peptides. Characteristically, it turned out that the backbone torsional angle values were distributed in the range of the β -strand characteristic backbone torsional angles (Fig. 3B) [34]. Among 5 peptides, all trajectories showed residue 4–7 φ and ψ angles in β -strand range. Residue 2–3 φ and ψ angles were inside β -strand range in native phosphopeptides (Fig. 3B–a, c). However, non-phosphopeptides and artificially designed NEWp did not have their distributions in β -strand range (Fig. 3B–b, d, e). Moreover, error bar ranges decreased from N-terminal residue (Residue 2) toward C-terminal residue (Residue 7). This pattern implies that the backbone stability increased along the amino acid order from N-terminal to C-terminal, and thus strengthens the β -MoRF formation.

3.4. Phosphorylation residues with greater chance of hydrogen bond formation with peptides

Our molecular simulation indicated that the torsional angles of 5 peptides backbone are distributed in the β -strand range (Fig. 3). β -strand amino acids have hydrogen bonds with neighbors in β -strand secondary structures [35]. Besides, previous discoveries of β -strand formation in protein – protein contact regions [36] and IDR β -MoRF [24] confirmed this characteristic. Thus, we investigated hydrogen bond numbers in contact regions between EmBR FHA domain and 5 peptides (Fig. 4).

Some amino acids have shown hydrogen bond formations in selected loops (Fig. 4A; Loop 1 G311 – H329 (red), Loop 2 D343 - A360 (green), Loop 3 R370 – Q375 (blue)) of EmBR FHA domain. Loop 1 residues R312, A323, N3234, S326, and R327 showed hydrogen bond forming patterns. As investigated after the equilibration of the positive control (Rad9p, N'-SLEVT^PEAD-C', Fig. 2B–a, b), S326, and R327 still contain the hydrogen bonds. We also found R312 to be a strong hydrogen bond former (donor/acceptor). Between two sorts of peptides (phosphopeptides, and non-phosphopeptides), S326, and R327 formed hydrogen bonds more likely with phosphopeptides than non-phosphopeptides (Fig. 4B). Whereas, R312 formed a greater number of hydrogen bonds with non-phosphopeptides than phosphopeptides (Fig. 4B).

In Loop 2 and 3, S347, N348, R356, and R370 showed greater number of hydrogen bonds. S347, N348, and R370 appeared as the major hydrogen bond formers in Loop 2 and 3. Among these residues, N348, and R370 formed more hydrogen bonds with non-phosphopeptides than phosphopeptides as shown in Loop 1 R312. However, S347 formed more hydrogen bonds with phosphopeptides than non-phosphopeptides as patterns shown in Loop 1 S326, and R327. R356 formed hydrogen bond only with PknHp (phosphorylated PknH 273–280, N'-HEALSPDPD-C'). In Fig. 4B, the red arrows indicate phosphorylation related residues R312, S326 and N348 according to previous research [14]. Their substitution with alanine caused reduced phosphorylation by PknH. Thus, hydrogen bond formation with these residues, as well as with R356, is important in EmBR – PknH interaction.

We also measured the number of hydrogen bonds formed on the peptide (Fig. 4C) amino acid residues. As expected and discovered from equilibration structures (Fig. 2A), phosphorylated residues (fifth position, Residue 5) in phosphopeptides (Fig. 4C 1, 3, 5) showed a greater

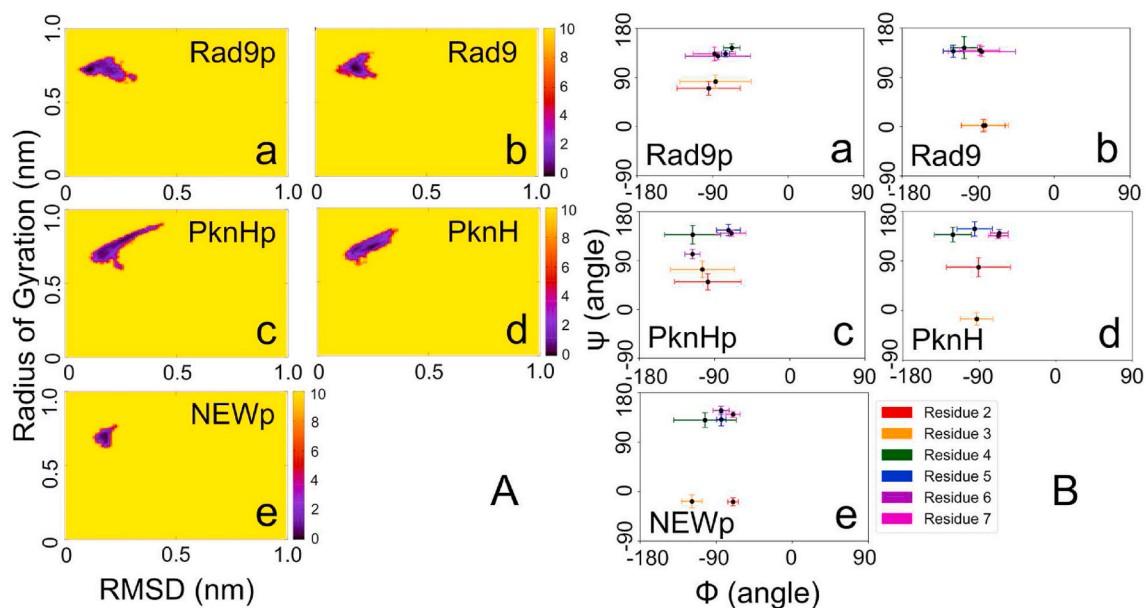


Fig. 3. Peptide structural stability and backbone torsional angle measurement. A: Heavy atoms radius of gyration (RofG), and root-mean-squared deviation (RMSD) of peptides (a–e) a; Rad9p, b; Rad9, c; PknHp, d; PknH, e; NEWp. $\Delta G = 0$ is corresponding to the greatest number of structures. B: Torsional angles (φ , and ψ) of the peptide backbones in ranges between -180° – 90° for φ , and -90° – 180° for ψ (a–e) a; Rad9p, b; Rad9, c; PknHp, d; PknH, e; NEWp.

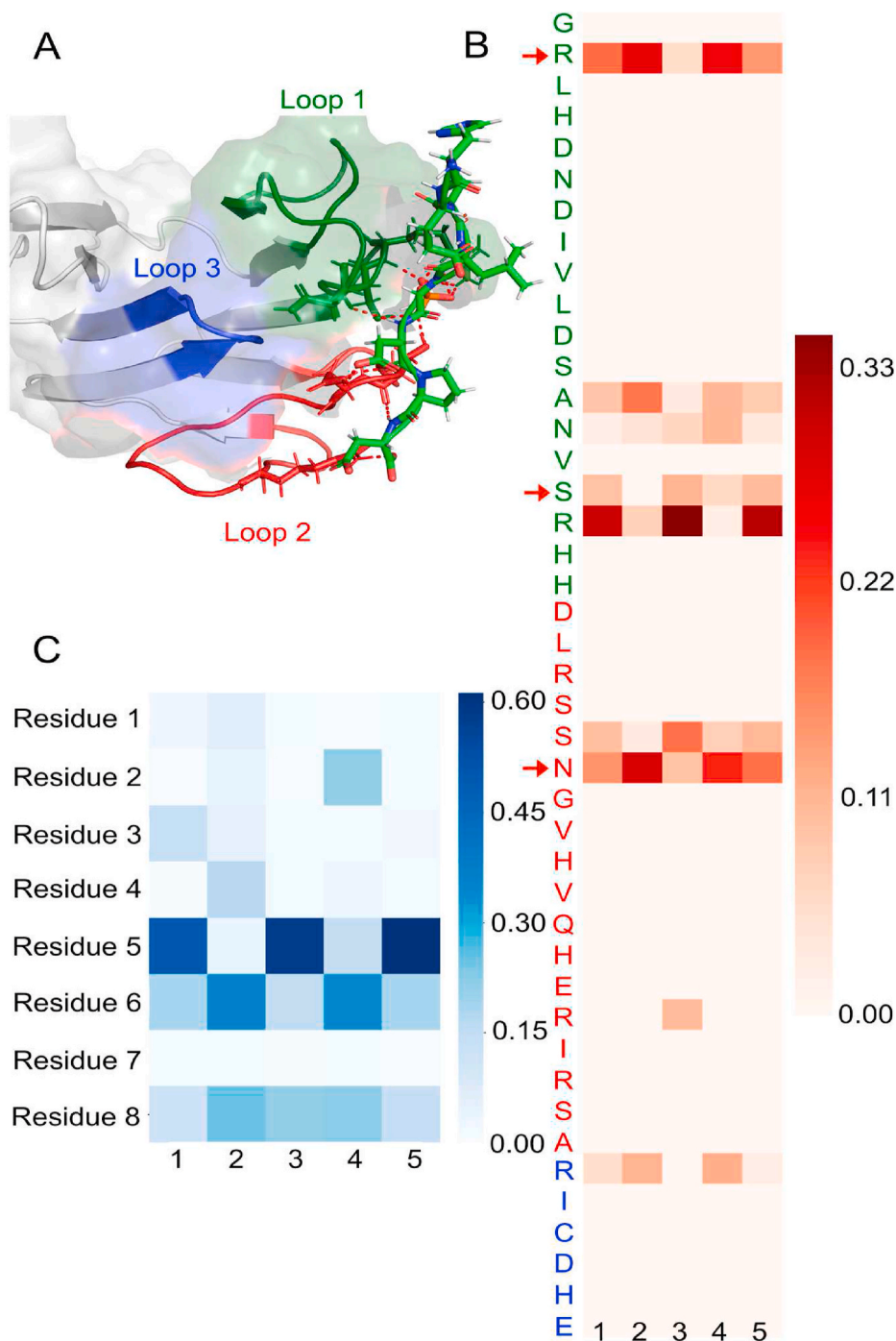


Fig. 4. Hydrogen bond patterns in EmBR FHA domain – 5 peptides. **A:** EmBR FHA domain loops. Loop 1 G311 – H329 (green), Loop 2 D343 - A360 (red), Loop 3 R370 – Q375 (blue) **B:** FHA domain hydrogen bond formers. 1; Rad9p, 2; Rad9, 3; PknHp, 4; PknH, 5; NEWp. Red arrows indicate important residues for phosphorylation. **C:** 5 peptides hydrogen bond formers. 1; Rad9p, 2; Rad9, 3; PknHp, 4; PknH, 5; NEWp. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

number of hydrogen bonds. As shown in Fig. 4C, Residues 5–8 showed greater density than Residues 1–4. Thus, we could conclude that phosphopeptides bind the EmBR FHA domain through the residues 5–8 with hydrogen bond formation capacity.

Binding validation of peptides with EmBR protein by biolayer interferometry.

The IDR-peptides and their phosphorylated ones were synthesized, biotinylated, and loaded onto Streptavidin biosensors for the measurement of their association and dissociation rate constants by biolayer interferometry (see Supporting Information, Materials & Methods). The association and the dissociation rate constants k_{on} and k_{off} were obtained by fitting the data to a 1:1 model, and the dissociation constant K_d values were calculated as the ratio k_{off}/k_{on} as shown in Table S1. For

each peptide, multiple negative controls were run (with association buffer only and with BSA replacing EmBR).

Notably, the binding profile generated by the biotinylated phosphopeptide corresponding to the predicted binding site on PknH (phosphorylated PknH 273–280, N'-HEALS^PDPD-C' with K_d 0.950 μ M) showed the strongest affinity with EmBR, which is distinguished from other peptides including IDR-peptide derived from the positive control (phosphorylated Rad9p 188–195, N'-SLEVT^PEAD-C' with K_d 1.52 μ M). This optimal binding affinity was found to be 0.950 μ M, corresponding to a moderate biologically relevant and specific binding. Three clinical TB drugs (ethambutol, rifampicin, and isoniazid) were also tested as controls and showed no significant binding to EmBR under the same conditions.

4. Discussion

From protein intrinsic disorder score measurement, we discovered *Mtb* EmBR binding region on *Mtb* PknH. The previous protein structure study (PDB ID 2FF4) of EmBR in complex with Yeast Rad9 188–195 (N'-SLEVTEAD-C') [15] did not determine the exact binding site, but served to provide the background logic for our discovery. We found a region within *Mtb* PknH (273–280 N'-HEALSDPD-C') containing similar protein intrinsic disorder, charge, and hydrophobicity with the region from PDB ID 2FF4. From the discovery, we investigated its binding pattern, applying molecular simulation in the anticipation of binding with EmBR FHA domain based on the position of the positive control (PDB ID 2FF4, Yeast Rad9 188–195, N'-SLEVTEAD-C') [15]. Through structural investigation of the molecular simulation trajectories of the 5 forms in stable complexes, we observed finger-like structures holding the phosphorylated residues where a less tight binding was observed in the structure study (PDB ID 2FF4) [15]. Additionally, we observed the C-terminal backbone torsional angles (ϕ and ψ) of the peptides' amino acids distributed in β -strand range on Ramachandran plot [34]. According to previous research on β -strand formation in protein-protein contact regions [36] and molecular recognition features (MoRF) [24], this once again confirmed the peptides' bindings toward EmBR. Further, hydrogen bond analysis of the simulations showed significant hydrogen bond forming potential of the residues in the FHA domain of EmBR. A previous mutagenesis study discovered some of these have important phosphorylation roles [14], and our simulation trajectories also showed their direct involvement in hydrogen bond formation with the FHA domain.

Protein intrinsic disorder oriented interaction predictors are being developed, and progressive. There are diverse forms of MoRF prediction algorithms [37], and efforts begun from α -MoRF predictor development in 2007 [38]. Most of the predictor developers applied machine learning to find correlations between protein intrinsic disorder score and binding [37]. Our approach to find the specific binding region directly relied on protein intrinsic disorder score and the comparison of biochemical properties with the positive control (PDB ID 2FF4) [15]. We also confirmed binding availability through biochemical property comparisons. Protein intrinsic disorder prediction algorithms already compose correlations between amino acid properties and their unfolding contributions as the first generation algorithms were developed [39]. Our discovery indicates that, if the binding site is short and has similar biochemical properties, protein intrinsic disorder score itself can be enough to predict the specific binding site.

Through molecular simulation, we also discovered finger-like structures at FHA domain, composed of three amino acids; S326, R327, and S347. Two fingers hold the phosphorylated fifth residue of phosphopeptides (Fig. 2A). The importance of phosphothreonine was already highlighted among FHA domain binding partners [18]. *Mtb* signal transduction is heavily dependent on Ser/Thr protein kinases (STPKs) [16], and their partners have FHA domains for corresponding phosphorylation [20]. EmBR FHA – Yeast Rad9 188–195 (N'-SLEVTEAD-C') is a representative FHA domain structure, although it didn't show STPK partner binding [15,18]. Our fingers at this FHA domain with phosphorylated *Mtb* PknH 273–280 (N'-HEALSDPD-C') strengthened recent viewpoint [20], which is challenging previous viewpoint of importance on phosphothreonine [18]. The recent article reviewed that the phosphorylation does not only always rely on threonine, but it also needs surrounding residues [20]. We designed a new peptide through the replacement of the fifth serine with threonine (NEWp; N'-HEALSDPD-C', NEW; N'-HEALSDPD-C') according to the previous knowledge [18]. But, NEW at 300K flew off from original binding position during production run, although other native peptides without phosphorylation were bound there during the production run. It implies that native non-phosphopeptides, as well as their phosphopeptides can have interactions at the contact region.

Nevertheless, hydrogen bond analysis distinguished FHA – peptide

binding patterns. Phosphopeptides and non-phosphopeptides have shown different hydrogen bonding patterns (Fig. 4B and C). The finger residues, S326, R327, and S347 were major hydrogen bond formers in binding with phosphopeptides. However, R312, A326, N348, and R370 appeared as major hydrogen bond formers with non-phosphopeptides. A previous study of mutagenesis at R312, S326, and N348 showed no phosphorylation [14]. Among these three residues, R326 was the only major hydrogen bond former with phosphopeptides. Thus, we infer the binding and phosphorylation orders. R312, and N348 may participate in the binding with non-phosphorylated substrate (in this case, PknH). After the binding, S326 may induce phosphorylation, in coordination with other finger residues (R327 and S347). Then these finger residues may hold the phosphorylated residue tighter than a naïve binding partner. Certainly, the peptides with phosphorylated fifth residue showed more hydrogen bond formations toward the FHA domain (Fig. 4C).

Peptides in molecular simulation trajectories showed β -strand former ϕ and ψ angles in their backbone according to our discovery (Fig. 3B). Although the backbone torsional angles of residues 2–3 didn't show meaningful β -strand former ranges, residues 4–7 ϕ and ψ angles obviously showed their β -strand former capacity according to Ramachandran plot [34]. Peptide residue 5–8 also confirmed their hydrogen bond formation with FHA domain (Fig. 4C). Protein – protein interaction includes structural change. Especially, this case belongs to the β -strand formation [36]. As we discovered the EmBR specific binding site on PknH, it is a part of β -MoRF, a disorder-to-order transition region [24]. Importantly, the EmBR – peptide binding form all showed peptides stable binding. Heavy atoms root-mean-squared deviation range between 0 and 5 Å, and radius of gyration range between 5 and 10 Å. These results all support our hypothesis of *Mtb* PknH 273–280 binding toward EmBR FHA domain.

EmBR FHA domain is well known for its complex structure with Yeast Rad9 188–195 (N'-SLEVTEAD-C') as a representative example of FHA domain [15,18]. FHA domain only contacts threonine, and this dogma didn't change due to lack of counterpart discovery [16,18,19]. However, recent viewpoint describes that not only the threonine, but also other surrounding residues are important for molecular recognition [20]. We discovered, for the first time, that the serine embedded peptide part was the EmBR binding responsible site on PknH. As described in the previous review paper [11], EmBR is phosphorylated by PknH, and this reaction predictively causes structural change. PknH binding, and phosphorylation toward EmBR may cause structural change of EmBR itself, and it may allow better binding toward EmBR/A/B operon promoter region.

We synthesized the peptides in their various forms and performed validation *in vitro*. The binding assay results confirmed the overall better binding from the phosphopeptides compared to others and our binding site prediction PknH 273–280 (N'-HEALSDPD-C') outperformed the positive control Rad9 188–195 (N'-SLEVTEAD-C'), even after additional phosphorylation.

Unfortunately, none of the peptides showed significant inhibitory ability in terms of stopping the virulent reference strain H37Rv (see Table S2). We also probed for synergistic efficacy with clinical drugs ethambutol, rifampicin, and isoniazid, but did not pick up meaningful interaction.

From a translational research viewpoint, our discovery provides drug resistance overcoming background logic. The EmBR – PknH pair controls the embCAB operon [12,13], and genetic mutations on EmBR/A/B/R cause Ethambutol resistance [40–42]. Because the EmBR – PknH pair formation is upstream of embCAB operon control, EmBR – PknH inhibition is a potential Ethambutol resistance overcoming strategy, although EmBR inhibition has yet to be considered as a drug discovery target. PknH control could be considered as another strategy, however, previous PknH knockout experiment showed increased virulence *in vivo* [43].

Our approach pioneered protein intrinsic disorder utilization for

target's binding partner mimicking to inhibit the target. As described in the Introduction section, p53 α -MoRF utilization [24] confirmed its anti-leukemia efficacy through the example of RG-7112 [26,31]. Although that discovery is still a representative example of intrinsically disordered region (IDR) utilization, there was no approach to find a target's binding partner from protein intrinsic disorder prediction for drug discovery. We suggest that this approach is revolutionary in the computer-aided drug discovery field and can be applied in cases beyond the development of tuberculosis treatments.

CRedit authorship contribution statement

Insung Na: designed the experiments, analyzed the data, performed the experiments, wrote the manuscript with input from the other authors. **Huanqin Dai:** designed the experiments and analyzed the data, performed the experiments, wrote the manuscript with input from the other authors. **Hantian Li:** designed the experiments and analyzed the data, performed the experiments. **Anvita Gupta:** designed the experiments and analyzed the data, performed the experiments. **David Kreda:** designed the experiments and analyzed the data. **Powell Zhang:** performed the experiments. **Xiangyin Chen:** performed the experiments. **Lixin Zhang:** conceived the project. **Gil Alterovitz:** conceived the project.

Declaration of competing interest

This manuscript has not been published and is not under consideration for publication elsewhere. We have no conflicts of interest to disclose, all authors read the manuscript and agreed to submit to Signal Transduction and Targeted Therapy.

Acknowledgements

This work was supported by the National Institutes of Health Grant No. 7R01GM118467-05 and the National Natural Science Foundation of China (31720103901). We would like to thank Mr. James Jones in Boston Children's Hospital and Ms. Ning Xie in Brigham and Women's Hospital for critical reading and helpful discussions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.synbio.2021.11.006>.

References

- [1] Pai M, et al. Tuberc Nat Rev Dis Primers 2016;2:16076.
- [2] Adepoju P. Tuberculosis and HIV responses threatened by COVID-19. Lancet HIV 2020;7(5):e319–20.
- [3] Alene KA, Wangdi K, Clements AC. Impact of the COVID-19 pandemic on tuberculosis control: an overview. Trop Med Infect Dis 2020;5(3):123.
- [4] Wang Q, et al. Abyssomicins from the South China Sea deep-sea sediment *Verrucosipora* sp.: natural thioether Michael addition adducts as antitubercular prodrugs. Angew Chem Int Ed Engl 2013;52(4):1231–4.
- [5] Wu F, et al. Chrysoomycin A derivatives for the treatment of multi-drug-resistant tuberculosis. ACS Cent Sci 2020;6(6):928–38.
- [6] Ashforth EJ, et al. Bioprospecting for antituberculosis leads from microbial metabolites. Nat Prod Rep 2010;27(11):1709–19.
- [7] Chen C, et al. A marine-derived *Streptomyces* sp. MS449 produces high yield of actinomycin X2 and actinomycin D with potent anti-tuberculosis activity. Appl Microbiol Biotechnol 2012;95(4):919–27.
- [8] Thomas JP, et al. A new synthetic compound with antituberculous activity in mice: ethambutol (dextro-2,2'-(ethylenediimino)-di-1-butanol). Am Rev Respir Dis 1961; 83:891–3.
- [9] Belanger AE, et al. The embAB genes of *Mycobacterium avium* encode an arabinosyl transferase involved in cell wall arabinan biosynthesis that is the target for the antimycobacterial drug ethambutol. Proc Natl Acad Sci U S A 1996;93(21): 11919–24.
- [10] Goude R, et al. The arabinosyltransferase EmbC is inhibited by ethambutol in *Mycobacterium tuberculosis*. Antimicrob Agents Chemother 2009;53(10): 4138–46.
- [11] Kieser KJ, Rubin EJ. How sisters grow apart: mycobacterial growth and division. Nat Rev Microbiol 2014;12(8):550–62.
- [12] Sharma K, et al. Transcriptional control of the mycobacterial embCAB operon by PknH through a regulatory protein, EmbR, in vivo. J Bacteriol 2006;188(8): 2936–44.
- [13] Angala SK, et al. The cell envelope glycoconjugates of *Mycobacterium tuberculosis*. Crit Rev Biochem Mol Biol 2014;49(5):361–99.
- [14] Molle V, et al. An FHA phosphoprotein recognition domain mediates protein EmbR phosphorylation by PknH, a Ser/Thr protein kinase from *Mycobacterium tuberculosis*. Biochem 2003;42(51):15300–9.
- [15] Alderwick LJ, et al. Molecular structure of EmbR, a response element of Ser/Thr kinase signaling in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 2006;103 (8):2558–63.
- [16] Sherman DR, Grundner C. Agents of change - concepts in *Mycobacterium tuberculosis* Ser/Thr/Tyr phosphosignalling. Mol Microbiol 2014;94(2):231–41.
- [17] Richard-Greenblatt M, Av-Gay Y. Epigenetic phosphorylation control of *Mycobacterium tuberculosis* infection and persistence. Microbiol Spectr 2017;5(2).
- [18] Mahajan A, et al. Structure and function of the phosphothreonine-specific FHA domain. Sci Signal 2008;1(51):re12.
- [19] Prisc S, et al. Extensive phosphorylation with overlapping specificity by *Mycobacterium tuberculosis* serine/threonine protein kinases. Proc Natl Acad Sci U S A 2010;107(16):7521–6.
- [20] Almawi AW, Matthews LA, Guarne A. FHA domains: phosphopeptide binding and beyond. Prog Biophys Mol Biol 2017;127:105–10.
- [21] Molle V, et al. EmbR2, a structural homologue of EmbR, inhibits the *Mycobacterium tuberculosis* kinase/substrate pair PknH/EmbR. Biochem J 2008; 410(2):309–17.
- [22] Babu MM, et al. Intrinsically disordered proteins: regulation and disease. Curr Opin Struct Biol 2011;21(3):432–40.
- [23] Sigal AA. Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. Cancer Res 2000;60:6788–93.
- [24] Mohan A, et al. Analysis of molecular recognition features (MoRFs). J Mol Biol 2006;362(5):1043–59.
- [25] Hsu WL, et al. Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. Protein Sci 2013;22(3):258–73.
- [26] Kussie PH, et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. Science 1996;274(5289):948–53.
- [27] Vassilev LT, et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science 2004;303(5659):844–8.
- [28] Uversky VN, Oldfield CJ, Dunker AK. Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu Rev Biophys 2008;37:215–46.
- [29] Graves B, et al. Activation of the p53 pathway by small-molecule-induced MDM2 and MDMX dimerization. Proc Natl Acad Sci U S A 2012;109(29):11788–93.
- [30] Vu B, et al. Discovery of RG7112: a small-molecule MDM2 inhibitor in clinical development. ACS Med Chem Lett 2013;4(5):466–9.
- [31] Andreeff M, et al. Results of the phase I trial of RG7112, a small-molecule MDM2 antagonist in leukemia. Clin Cancer Res 2016;22(4):868–76.
- [32] Ross NT, Katt WP, Hamilton AD. Synthetic mimetics of protein secondary structure domains. Philos Trans A Math Phys Eng Sci 2010;368:989–1008. 1914.
- [33] Gee CL, et al. A phosphorylated pseudokinase complex controls cell wall synthesis in mycobacteria. Sci Signal 2012;5(208):ra7.
- [34] Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1963;7:95–9.
- [35] Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem 1981;34:167–339.
- [36] Remaut H, Waksman G. Protein-protein interaction through beta-strand addition. Trends Biochem Sci 2006;31(8):436–44.
- [37] Katuwawala A, et al. Computational prediction of MoRFs, short disorder-to-order transitioning protein binding regions. Comput Struct Biotechnol J 2019;17:454–62.
- [38] Cheng Y, et al. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. Biochem 2007;46(47):13468–77.
- [39] Romero P, et al. Sequence complexity of disordered protein. Proteins 2001;42(1): 38–48.
- [40] Brossier F, et al. Molecular analysis of the embCAB locus and embR gene involved in ethambutol resistance in clinical isolates of *Mycobacterium tuberculosis* in France. Antimicrob Agents Chemother 2015;59(8):4800–8.
- [41] Xu Y, et al. Mutations found in embCAB, embR, and ubiA genes of ethambutol-sensitive and -resistant *Mycobacterium tuberculosis* clinical isolates from China. Biomed Res Int 2015;951706. 2015.
- [42] Zhao LL, et al. Analysis of embCAB mutations associated with ethambutol resistance in multidrug-resistant mycobacterium tuberculosis isolates from China. Antimicrob Agents Chemother 2015;59(4):2045–50.
- [43] Papavinasandaram KG, et al. Deletion of the *Mycobacterium tuberculosis* pknH gene confers a higher bacillary load during the chronic phase of infection in BALB/c mice. J Bacteriol 2005;187(16):5751–60.