The Solution Structure of Escherichia coli Wzb Reveals a Novel Substrate Recognition Mechanism of Prokaryotic Low Molecular Weight Protein-tyrosine Phosphatases

Ewen Lescop, Yunfei Hu, Huimin Xu, Wei Hu, Juan Chen, Bin Xia, and Changwen Jin

From the Beijing Nuclear Magnetic Resonance Center, College of Life Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

Low molecular weight protein-tyrosine phosphatases (LMW-PTPs) are small enzymes that ubiquitously exist in various organisms and play important roles in many biological processes. In Escherichia coli, the LMW-PTP Wzb dephosphorylates the autokinase Wzc, and the Wzc/Wzb pair regulates colanic acid production. However, the substrate recognition mechanism of Wzb is still poorly understood thus far. To elucidate the molecular basis of the catalytic mechanism, we have determined the solution structure of Wzb at high resolution by NMR spectroscopy. The Wzb structure highly resembles that of the typical LMW-PTP fold, suggesting that Wzb may adopt a similar catalytic mechanism with other LMW-PTPs. Nevertheless, in comparison with eukaryotic LMW-PTPs, the absence of an aromatic amino acid at the bottom of the active site significantly alters the molecular surface and implicates Wzb may adopt a novel substrate recognition mechanism. Furthermore, a structure-based multiple sequence alignment suggests that a class of the prokaryotic LMW-PTPs may share a similar substrate recognition mechanism with Wzb. The current studies provide the structural basis for rational drug design against the pathogenic bacteria.

Low molecular weight-tyrosine phosphatases (LMW-PTPs) are small cytoplasmic enzymes (∼18 kDa) that are widely distributed in prokaryotes and eukaryotes (1, 2). In eukaryotes, LMW-PTPs specifically dephosphorylate and down-regulate many tyrosine kinase receptors, such as platelet-derived growth factor receptor (3, 4), insulin receptor (5), or ephrin receptor (6). The reaction mechanism of eukaryotic LMW-PTPs has been structurally, thermodynamically, and kinetically characterized (7). In contrast, very limited knowledge is available for prokaryotic LMW-PTPs thus far. All LMW-PTPs share a highly conserved C(X)6R motif, where X can be any amino acid. This motif adopts a loop structure, where the phosphate group of the substrate binds to, and is known as the P-loop (1). In addition, some amino acids required for the catalysis are also highly conserved, such as an aspartate residue that acts as a general acid (8). Based on the structures of mammalian LMW-PTPs in complex with exogenous substrates or serendipitous ligands, determinants for substrate specificity were proposed, including the ring stacking around the targeted phosphotyrosine provided by two aromatic side chains of the LMW-PTPs (9, 10).

Many bacterial species harbor a layer of capsular polysaccharides and surface-associated exopolysaccharides (EPS). In Escherichia coli, the group 1 capsular polysaccharides and the colanic acid EPS are assembled in a Wzy-dependent polymerization system (11). The ca gene cluster contains one tyrosine autokinase (Wzc) and one LMW-PTP (Wzb) that function as a pair of kinase/phosphatase in the regulation of colanic acid production (12, 13). As a consequence, colanic acid is only produced after the dephosphorylation of the phosphorylated Wzc by Wzb (14). Wzc was identified to regulate the activity of UDP-glucose dehydrogenase by the level of tyrosine phosphorylation and a working model for the role of Wzb and Wzc in EPS production has been proposed (15). In several Gram-negative bacteria, there exist other kinase/phosphatase pairs that are homologous to the Wzc/Wzb pair, which also involve in capsular polysaccharides or EPS transport and regulation (16–19).

The amino acid sequences of Wzb and some other bacterial LMW-PTPs contain only one of the two residues with aromatic side chains that are critical for substrate recognition in eukaryotic LMW-PTPs. This fact leads to an open question: what are the determinants for the substrate specificity of those bacterial LMW-PTPs such as Wzb? In an effort to elucidate the structural basis of Wzc/Wzb interactions and further to understand the substrate recognition mechanism of the prokaryotic LMW-PTPs, we have determined the solution structure of the LMW-PTP Wzb from E. coli by NMR spectroscopy. The overall fold of Wzb highly resembles that of the typical LMW-PTPs. However, significant diversity of the tertiary structures was observed between Wzb and the eukaryotic LMW PTPs, especially
around the active site P-loop. Our results reveal a novel mechanism of the substrate recognition for LMW-PTPs.

EXPERIMENTAL PROCEDURES

Sample Preparation—The wzb gene from E. coli was amplified by PCR and cloned into the plasmid pET28a (+) (Novagen) with an N-terminal His tag and expressed in E. coli BL21(DE3). Cells were allowed to grow overnight in 25 ml of Luria-Bertani (LB) medium at 37 °C and subsequently transferred into 1 liter of LB medium. When the A660 reached around 0.6, cells were spun down and resuspended in 250 ml of M9 minimal medium containing 15N-labeled ammonium chloride in the presence or absence of 13C-labeled glucose for 15N/13C-labeled or 15N-labeled samples, respectively (20). The resuspended cells were allowed to grow in minimal medium for 1 h before induction. The protein expression was induced by adding isopropyl β-d-thiogalactoside to a final concentration of 100 mg/liter. After 6 h of induction at 20 °C, the cells were harvested. The soluble protein was purified by the affinity chromatography (nickel-nitrilotriacetic acid-agarose column), and subsequently gel filtration (Superdex-120 column) with the AKTA FPLC system (Amersham Biosciences). Only a single peak corresponding to the monomeric molecular weight was observed during gel filtration. The purity was determined to be greater than 95% as judged by SDS-PAGE. NMR samples were prepared in a buffer containing 50 mM sodium phosphate, 50 mM NaCl, 25 mM dithiothreitol, 5 mM EDTA, 0.02% sodium azide at pH 6.0 in 15N-containing M9 medium. The protein concentration was determined by the 1H-15N residual dipolar coupling (RDC) constants were measured by dissolving Wzb protein in a dilute liquid crystal buffer containing a mixture of alkyl-poly(ethylene glycol) C12E5 and n-hexanol (25). The C12E5/water ratio was 5.5 weight %, and the molar ratio of C12E5 to n-hexanol was 0.92. The RDCs were extracted from the difference in 1H-15N splitting measured by 1H-15N IPAP-HSQC spectra between the weakly aligned and the isotropic samples at 25 °C (26). The data were analyzed using software packages PALES (27) and MODULE (28)

Structure Calculations—The initial structure calculations were performed using the automatic nuclear Overhauser effect (NOE) assignment program CANDID embedded in CYANA (29). The extended NOE contacts and chemical shift assignments were obtained by an iterative analysis of the calculated structures and NOE assignments using the program SANE (30). The NOE-derived distance restraints, in combination with the backbone dihedral angles determined by TALOS (31) and hydrogen bond information, were used in each calculation step. Two hundred structures were calculated using CYANA and the 100 structures with the lowest target function were selected for further refinement by AMBER7 (32). The 1H, 15N RDCs were used as restraints during the refinement stage. As recommended in the AMBER7 manual, explicit angle restraints were added to enforce the ideal backbone geometry for residues with RDC restraints. Finally, the 20 lowest energy structures were selected to represent Wzb. The structures were analyzed using the software packages MOLMOL (33) and PROCHECK_NMR (34).

Structure-based Multiple Sequence Alignment—The LMW-PTPs with known tertiary structures were aligned using STRAP. LMW-PTPs of unknown tertiary structures were subsequently added in the multiple sequence alignment using the program ClustalX (35).

Accession Numbers—The chemical shifts have been deposited in the BioMagResBank data base under accession number 6934. The coordinates of Wzb structures have been deposited in the Protein Data Bank with accession number 2FEK.

RESULTS

The Overall Structure of E. coli Wzb—The solution structure of E. coli Wzb was determined using the NOE-derived distance restraints in conjunction with the dihedral angle, RDC, and hydrogen bond restraints. The 20 lowest energy structures were selected to represent Wzb and are shown in Fig. 1A, together with the ribbon diagram showing the secondary structural elements (Fig. 1, B and C). The structural statistics are listed in Table 1. The root mean square deviation of the backbone heavy atoms is 0.90 ± 0.23 Å for all residues and 0.50 ± 0.09 Å for those in secondary structural elements. Ramachandran plots of all 20 structures indicate that most of the backbone dihedral angles lie within the energetically favorable regions of the (φ, ψ) space.

Overall, Wzb adopts a typical LMW-PTP fold, consisting of 4 β-strands (β1 (Asn4-Cys9), β2 (Lys11-Leu37), β3 (Leu78-Thr81)) and β4 (Lys100-His106) and five α-helices (α1 (Arg15-Tyr26), α2 (Pro47-His56), α3 (Arg70-Arg74), α4 (Lys84-Met93), and α5 (Arg121-Leu142)). The β-strands form into a central highly twisted parallel β-sheet (β4-β2-β1-β3) flanked by helices α1, α2, and α5 on one side, while α3 and α4 on the other side. Three
long loops (β2–α2, α3–α4, and β4–α5) link the secondary structural elements. In addition, a type II β-turn is formed from Leu40 to Lys43 in loop β2–α2.

The Active Site—The P-loop of Wzb contains the amino acid sequence CVGNICR, where the underlined residues are highly conserved among all identified LMW-PTPs (Fig. 2). The P-loop connects strand β4 and α-helix α1 (Fig. 1B) and is buried at the bottom of a crevice (Fig. 3), which is delineated by residues Ala39, Leu40, Ala45, and Asp46 in loop β2–α2. Glu83 and His86 in helix α2, and Asp115 to Tyr117 in loop β3–α5. Residues Asp115, Glu83, and those in the P-loop form the upper and center part of the crevice, whereas residues in loops β2–α2 and β4–α5 contribute to the lower part as viewed in Fig. 3. The side chains of Tyr117 and Leu116, together with the main chain of Ala45 and Asp46 form a ring around the side chain of Ile13 below the P-loop. To confirm the binding of the P-loop of Wzb with phosphate ions, the phosphate titration experiments were performed and monitored by a series of two-dimensional 15N-edited HSQC experiments. In the absence of phosphate ions, 13 residues (Cys9–Asn12, Cys14–Ser16, Glu20, Gly36, Leu37, Leu40, Arg66, and Tyr117) of Wzb completely disappeared in the 15N-edited HSQC spectrum, whereas the remaining cross-peaks in the spectrum were almost identical to that in the presence of phosphate ions (Fig. 4A). Upon addition of phosphate, all missing cross-peaks showed up and their intensities gradually increased with the increase of phosphate concentration. After the phosphate concentration was higher than 20 mM, the peak intensities in the HSQC spectra remained unchanged. This result confirms that the phosphate binds to Wzb and stabilizes the P-loop, and indicated that phosphate binding does not significantly alter the overall structure of Wzb except the P-loop region in which most residues are missing from the two-dimensional 15N-edited HSQC spectrum in the absence of phosphate ions. Nevertheless, we compared the P-loop region of Wzb with that of the crystal structure of phosphate-bound BPTP (Protein Data Bank code 1PHR) (36) and found that all the NH bonds of Wzb whose NMR signals were missing in the absence of phosphate point toward the position where the phosphate group is expected to bind (Fig. 4B). Based on this, the inorganic phosphate in Wzb likely occupies a similar position as that in other LMW-PTPs. In addition, the phosphate titration experiments suggest that the P-loop undergoes significant mobility (conformational exchanges) in the absence of phosphate, which causes the broadening of the NMR signals and makes them undetectable. The conformational exchanges are largely diminished after binding with phosphate ion, which is further demonstrated by the results of backbone dynamics (supplemental data).

DISCUSSION

Structural Basis of the Catalytic Mechanism—We have solved the solution structure of E. coli Wzb, the first LMW-PTP structure from the Gram-negative bacteria. To date, structures of LMW-PTPs from eukaryotic organisms, such as human HCPTPA and HCPTPB (9, 37), bovine BPTP (36, 38, 39), and yeast Saccharomyces cerevisiae LTP1 (10) are available. The structures of the prokaryotic LMW-PTPs MPtpA from Mycobacterium tuberculosis (40) and YwlE from the Gram-positive Bacillus subtilis (41) have been reported recently. Although Wzb shares a low sequence identity (less than 30%) with these LMW-PTPs, it shows a high structural similarity with them, especially the highly conserved regular secondary structures. Residues Cys8–Ser16 of the catalytic site fold into a typical P-loop conformation, and the root mean square deviation values of the backbone atoms of these residues with those of the equivalent residues (Cys12–Ser19 in BPTP and HCPTPA, Cys13–Ser20 in LTP1) in eukaryotic LMW-PTPs range from
Table 1: Structural statistics of E. coli Wzb

| Statistics                                      | 
|-------------------------------------------------|
| **Restraints**                                  | 
| Intra-residue unambiguous NOEs                   | 797 |
| Sequential unambiguous NOEs                     | 606 |
| Medium range unambiguous NOEs                   | 494 |
| Long range unambiguous NOEs                     | 743 |
| Total ambiguous NOEs                            | 2640 |
| Total dihedral angles (backbone + side chain)   | 1630 |
| Hydrogen bond distance restraints              | 100 |
| \(^1\)H–\(^15\)N RDC restraints                | 97  |
| **Restraint violations**                        | 
| NOE-derived distance (>0.2 Å)                  | 0   |
| Torsion angle, >5°                              | 3   |
| Torsion angle, >10°                             | 0   |
| **Energy (kcal/mol)**                           | 
| Mean AMBER energy                               | \(-7887 \pm 13\) |
| NMR restraints violation energy                 | \(32.3 \pm 7.3\) |
| **Root mean square deviation from mean structure (Å)** | 
| Backbone heavy atoms                            | 0.90 ± 0.23 |
| Secondary structures                            | 0.50 ± 0.09 |
| All heavy atoms                                 | 1.58 ± 0.23 |
| Secondary structures                            | 1.29 ± 0.16 |
| **Ramachandran statistics**                    | 
| Residues in most favored regions (%)            | 86.0 |
| Residues in additionally allowed regions (%)    | 12.8 |
| Residues in generously allowed regions (%)      | 0.4  |
| Residues in disallowed regions (%)              | 0.8  |

\(^a\) \(\bar{f} = \bar{j} = 1\), \(\bar{g} = \bar{j} = 5\).
\(^b\) Two upper limits were used for each hydrogen bond restraint (d(HN, O) < 2.3 Å and d(N, O) < 3.3 Å).
\(^c\) Include NOE, dihedral angle, and RDC violation energy.
\(^d\) The backbone heavy atoms including \(^{\text{C}}\) amide nitrogen, and carbonyl carbon atoms.
\(^e\) All heavy atoms including all the non-hydrogen atoms of the protein.

0.30 to 0.35 Å. In addition, most of the main chain NH bonds in the P-loop point toward the expected position of the phosphate group. Furthermore, the side chain of the strictly conserved asparagine residue Asn12 is spatially closed to that of His63 (the distance between N\(^2\) of Asn12 and N\(^2\) of His63 is d(N\(^2\)Asn12, N\(^2\)His63 = 3.4 ± 0.4 Å), Ser\(^{16}\) (d(N\(^2\)Ser16, S16O\(^{\gamma}\) = 3.7 ± 0.8 Å), and Ser\(^{43}\) (d(N\(^2\)Ser43, S43O\(^{\gamma}\) = 4.9 ± 0.5 Å). A similar geometry is observed in BPTP, HCPTPA, and LTP1 (residues Asn\(^{15}\), His\(^{72}\), Ser\(^{15}\), and Ser\(^{43}\), respectively in BPTP). These distances suggest the presence of a hydrogen-bonding network involving Asn12 in Wzb, which may stabilize the backbone conformation of the P-loop, as previously observed in BPTP (38). On the upper side of the P-loop as viewed in Fig. 3, the guanidinium moiety of the conserved Arg\(^{15}\) is in a favorable distance with the acidic groups of Asp\(^{43}\) and Glu\(^{83}\) to form hydrogen bonds or salt bridges. Moreover, the H\(^\alpha\) atom of Arg\(^{15}\) in Wzb is superimposable with that of the equivalent arginine in BPTP (Arg\(^{18}\), HCPTPA (Arg\(^{19}\), and LTP1 (Arg\(^{17}\)). In those LWM-PTPs, the H\(^\alpha\) atom of the arginine likely forms a hydrogen bond with an oxygen atom of the phosphate substrate. As a consequence, residue Arg\(^{15}\) in Wzb most likely contributes to the recognition and position of the substrate in a similar way as eukaryotic LWM-PTPs. The acidic group of residue Asp\(^{15}\) in Wzb is close to the guanidinium moiety of Arg\(^{15}\) and is in a favorable position to provide a proton to the leaving phenolic group. This aspartic residue is believed to serve as a general acid in the reaction (8). In Wzb, the Cys\(^9\) side chain is directed toward the phosphate binding site and the proximity of the hydroxyl group of Ser\(^{16}\) to Cys\(^9\) side chain (d(C9\(^\gamma\), S16\(^{\gamma}\) = 3.9 ± 0.8 Å) suggests hydrogen bonding. In BPTP, Cys\(^{12}\) (equivalent to Cys\(^9\) in Wzb) forms a covalent phosphocysteine intermediate during the reaction (42) and is likely activated by hydrogen bonding with Ser\(^{19}\) (equivalent to Ser\(^{16}\) in Wzb) (43). As a consequence, Cys\(^9\) may be activated by Ser\(^{16}\) in Wzb. The fact that Wzb shares high structural similarity with the eukaryotic LWM-PTPs strongly suggests that the dephosphorylation mechanism of Wzb is similar to that of other LWM-PTPs (7). Briefly, the phosphoryl group of the phosphorylated Wzc undergoes a nucleophilic attack by the active site Cys\(^9\) side chain of Wzb. The guanidinium moiety of Arg\(^{15}\) stabilizes the phosphoenzyme intermediate state and the invariant aspartate residue Asp\(^{12}\) protonates the ester oxygen of the leaving group. In the following step, the phosphoenzyme is hydrolized by the nucleophilic attack of a water molecule, possibly assisted by Asp\(^{15}\). Finally, one cycle of the dephosphorylation is finished with the release of the inorganic phosphate (7).

Substrate Recognition—Previous structural studies on eukaryotic LWM-PTPs in complex with exogenous substrates identified two regions around the catalytic center that are important for the enzyme activity (9, 10). In loop B\(_{2} – \alpha_3\) of the eukaryotic LWM-PTPs, an aromatic ring (Tyr\(^{131}\) in HCPTPA and BPTP, Trp\(^{34}\) in LTP1) was found to stabilize the substrate by interacting with the aromatic ring of the phosphotyrosine from one side. In Wzb, Tyr\(^{117}\) is structurally equivalent to Tyr\(^{131}\) in BPTP (Fig. 5A) and likely plays a similar role in substrate recognition and stabilization by providing \(\pi – \pi\) stacking contacts with the phenolic ring of the substrate. Notably, an aromatic residue is conserved at this position among all bacterial LWM-PTPs with a preference for tyrosine (Fig. 2). This fact suggests that this kind of ring-ring contact is essential for the function of LWM-PTPs in both prokaryotes and eukaryotes. On the other side of the substrate, the loop B\(_{2} – \alpha_3\) of eukaryotic LWM-PTPs contains another aromatic residue (Tyr\(^{49}\) in HCPTPA, Trp\(^{49}\) in BPTP, and Tyr\(^{41}\) in LTP1) that provides contacts with the substrate and closes the claw around it (9, 10). However, the loop B\(_{2} – \alpha_3\) of Wzb does not contain an aromatic residue that is equivalent to Trp\(^{49}\) in BPTP. Consequently, loop B\(_{2} – \alpha_3\) is shorter by 1 residue in Wzb compared with the eukaryotic LWM-PTPs (Fig. 2). As a result, the C\(^\beta\) traces diverge significantly between Wzb and BPTP in this region (Fig. 5A). Furthermore, residues Glu\(^{50}\) and Asn\(^{53}\) in human and rat LWM-PTPs have been demonstrated to play a critical role in substrate specificity (9, 44). In Wzb, this position is occupied by hydrophobic residue Leu\(^{40}\). The side chain of Leu\(^{40}\) in Wzb delineates one side (Fig. 3) of the substrate entry site and is superimposable with the side chains of Asn\(^{50}\) in BPTP (Fig. 5A) and Glu\(^{50}\) in HCPTPA. As a consequence, Leu\(^{50}\) may play a role in recognizing a hydrophobic patch on the substrate. The structural differences are more obvious between segments Gly\(^{38}\)–Ala\(^{39}\) of Wzb and Ser\(^{12}\)–Trp\(^{49}\) of BPTP. In eukaryotic LWM-PTPs, such as BPTP, this segment loops out and exposes the side chains of Asp\(^{48}\) and Trp\(^{49}\) to the solvent. In contrast, the methyl group of Ala\(^{39}\) in Wzb points.
toward the interior of the structure and is in close contact with
the side chain of Asn12 in the P-loop and the backbone atoms of
Arg66. This structural arrangement has two consequences. (i)

No aromatic residue in Wzb is equivalent to Trp49 in BPTP,
which provides the ring-ring interactions on one side of
the substrate ring. This factor may also contribute to the relatively
lower substrate affinity of Wzb ($K_m$ of 1 mM) (13) compared
with BPTP (0.38 mM) and LTP1 (0.017 mM) (7) using
p-nitrophenyl phosphate as the substrate. (ii) The molecular surface of
the lower part of the crevice of Wzb is significantly different
from that of BPTP (Fig. 5B). In Wzb, the backbone carbonyl
atom of Gly38 delineates one part of the substrate entry. This
group is directed toward the expected position of the phospho-
tyrosine substrate and exposes a polar patch to the substrate-
binding site. The side chains of Leu40 and Tyr117, together with
the polar carbonyl atom of Gly38 are potential candidates to
determine the substrate specificity of Wzb. Overall, the sub-
strate entry is slightly bigger in Wzb than that in BPTP. How-
ever, the catalytic center of Wzb is only accessible for long side
chain atoms. This may explain why Wzb is inactive on phos-
phoserine or phosphothreonine (13). In addition, six charged
residues (Lys43, Glu83, Arg85, Arg89, Arg118, and His86) are spa-
tially close to the catalytic center and are within reasonable
distances to interact with residues adjacent to the target phos-
photyrosine in the endogenous substrate, and will be discussed
below.

The autokinase Wzc is the endogenous substrate of Wzb and
the Wzb/Wzc pair regulates the production of colanic acid
in vivo (14). Wzb was shown to exclusively and extensively
dephosphorylate the phosphotyrosines of the C-terminal tyro-
sine-rich cluster in Wzc (Tyr708-Gly-Tyr710) and Tyr711.

FIGURE 2. The multiple sequence alignment. The plot shows the structure-based multiple sequence alignment of the LMW-PTPs Wzb (E. coli), Ptp (Acinetobacter johnsonii), Etp (E. coli), Aspi (E. amylovora), Yor5 (K. pneumoniae), EpsP (Pseudomonas solanacearum), YePh (B. subtilis), PtpA and PtpB (S. aureus), PtpA (S. coelicolor), MptPA (M. tuberculosis), Stp1 (S. pneumoniae), LTP1 (S. cerevisiae), HCFP and BPTP (bovine). References for sequences not cited in the text can be found in Ref. 2. Strictly conserved residues are labeled on a dark background and the residues with high similarity are
highlighted by boxes. Stretches of residues containing the hallmarks of class I (blue) and class II (red) LMW-PTPs are colored. The secondary structures of Wzb are shown at the top. The figure was prepared with ESPript by applying the BLOSUM 62 scoring matrix.

FIGURE 3. Ribbon representation of Wzb showing the catalytic P-loop. Residues nearby the active center are represented by sticks and labeled. The figure was generated using MOLMOL (33).
Tyr^713\text{-}Glu-Tyr^715\text{-}Lys-Ser-Asp-Ala-Lys) (14, 15). The *E. coli* Etp/Etk and the *Acinetobacter johnsonii* Ptp/Ptk are pairs of LMW-PTP/kinase and are the homologs of Wzb/Wzc (16, 45). Interestingly, Etp and Ptp could also dephosphorylate the phosphorylated Wzc in vitro to the same extent as Wzb (13, 14).

Although the kinetic data of the dephosphorylation reactions are not available, these experiments provide a basis for understanding the determinants on Wzc dephosphorylation by LMW-PTPs. Etp and Ptp share 56 and 35% sequence identity and 73 and 54% sequence similarity with Wzb, respectively. Regions with the highest sequence similarity among the three LMW-PTPs include the catalytically essential P-loop, the DPY (Asp^115\text{-}Pro^116\text{-}Tyr^117 in Wzb) segment in loop 4–5 and buried residues (Fig. 2). In addition, both Etp and Ptp also possess a hydrophobic side chain (leucine or methionine) like Leu^40 in Wzb. This observation supports the role of Leu^40 in Wzb on substrate recognition. In contrast, residues Lys^43, Glu^83, Arg^85, Arg^89, Arg^118, and His^86 that surround the active site of Wzb are neither conserved nor replaced by similar amino acids in Etp and Ptp (Fig. 2). This fact suggests that electrostatic interactions are not essential for the recognition of residues adjacent to the phosphotyrosines in Wzb. Notably, Asp^53 was proposed to drive the substrate specificity of HCPTPA (9) and this position is occupied by Lys^43 and Tyr^44 in Wzb and Ptp, respectively. This may suggest that Lys^43 in Wzb plays a dispensable role in substrate recognition, which is different from residue Asn^53 in HCPTPA.

In addition, the sequences containing the phosphotyrosine residues recognized by prokaryotic and eukaryotic LMW-PTPs are distinct. Wzb has been shown to target the C-terminal tyrosine-rich tail in Wzc (YGYYEYEYKSDAK) (14, 15). Interestingly, the homologous proteins of Wzc in prokaryotes, including Etk from *E. coli* and other predicted protein-tyrosine kinases (AmsA from plant pathogen *Erwinia amylovora* and Orf6 from human pathogen *Klebsiella pneumoniae*) all share a C-terminal tyrosine-rich tail with similar sequences (YSYGNYGYSYSEKE for Etk, YGGYNYYDYDSYSDKK for Orf6, and YGGYDYYDYDYSYQQGEKS for AmsA, respectively) (16). However, these tyrosine-rich sequences are generally not found in tyrosine kinase receptors targeted by the eukaryotic LMW-PTPs. The sequences recognized by the eukaryotic LMW-PTPs usually contain a single tyrosine residue flanked by charged amino acids. In contrast, the sequences with multiple tyrosine residues recognized by prokaryotic LMW-PTPs (i.e. Wzb, Etp, and Ptp) contain less charged amino acids and are more hydrophobic. This may explain the fact that hydrophobic residues like Leu^40 in Wzb are engaged in substrate specificity of prokaryotic LMW-PTPs, and the charged residues around the active site are dispensable in substrate recognition.

**A Novel Substrate Recognition Mechanism in Prokaryotic LMW-PTPs—** *E. coli* Wzb and *B. subtilis* YwIE share high structural similarity near the active site. Similar to Wzb, no aromatic residue equivalent to Trp^49 in BPTP is present in YwIE (41). In addition, the Phe^40 backbone carbonyl atom and the side chains of Ala^41 and Ser^42 of YwIE are superimposable with the carbonyl atom of Gly^38 and the side chains of Ala^39 and Leu^40 of Wzb, respectively. In eukaryotic LMW-PTPs, two aromatic rings sandwich the phosphotyrosine. In contrast, in both Wzb and YwIE, a polar group exposed by a backbone segment provides a potential contact with the substrate. Taken together, the comparison of the structures suggests that Wzb and YwIE likely adopt a novel mechanism on substrate recognition that is different from the eukaryotic LMW-PTPs. Further sequence
alignment and structural analysis of other LMW-PTPs suggest that some LMW-PTPs from prokaryotes may share the same substrate recognition mechanism as Wzb. The absence of an aromatic residue equivalent to Trp in BPTP, and therefore a shorter loop in BPTP, distinguishes the substrate recognition mechanisms between the prokaryotic-like (Wzb and YwlE) and eukaryotic-like LMW-PTPs (BPTP). Based on the above results, we suggest that the LMW-PTPs can be further classified into two groups with distinct substrate recognition mechanisms: those sharing the similar recognition mechanism with eukaryotic LMW-PTPs such as BPTP can be referred as class I LMW-PTPs, and those sharing the similar recognition mechanism with prokaryotic Wzb can be referred as class II LMW-PTPs. An unrooted phylogenetic tree inferred from multiple sequence alignment (see supplemental data) shows that the two classes of LMW-PTPs diverged during evolution and hence support the proposed classification.

A closer examination of the sequence alignment provides additional information regarding the differences between class I and class II LMW-PTPs. In fact, all the class I LMW-PTPs possess a tyrosine at the position equivalent to Tyr in the BPTP sequence, whereas this position may accommodate various amino acids in the class II LMW-PTPs. The phosphorylation state of Tyr plays an important physiological role in regulating the activity of the protein BPTP and controlling the strength of NIH-3T3 cell substrate adhesion (46, 47). The absence of a tyrosine at this position in class II LMW-PTPs suggests that their enzyme activities might be regulated by other processes (48).

Furthermore, the structure-based multiple sequence alignment suggests that both classes of LMW-PTPs exist in the prokaryotic cells. The determinants for substrate specificity of Class II LMW-PTP (Wzb-like) are shared by other prokaryotic LMW-PTPs such as Etp, Ptp, Yor5, EpsP, Ansm, and Staphylococcus aureus PtpB (Fig. 2). In contrast, the prokaryotic M. tuberculosis MPtpA, Streptomyces coelicolor PtpA, and S. aureus PtpA belong to Class I LMW-PTPs and share a similar substrate recognition mechanism with the eukaryotic LMW-PTPs. Notably, MPtpA is secreted into the extracellular medium and was shown to inhibit phagocytosis of mycobacteria (49, 50). Because MPtpA targets components of the eukaryotic host signaling pathways, its similarities with eukaryotic LMW-PTPs may be of biological significance. Further investigations are expected to elucidate the biological relevance of the two classes of LMW-PTPs with different substrate recognition mechanisms in prokaryotes.

Concluding Remarks—Our current studies demonstrate that a class of prokaryotic LMW-PTP adopts a substrate recognition mechanism different from that of the eukaryotic LMW-PTPs. Although all LMW-PTPs known so far may share many determinants in substrate recognition, class II LMW-PTPs probably partly recognize their endogenous substrates by hydrophobic side chains and a backbone carbonyl group, instead of an aromatic ring in class I LMW-PTPs. Many class II LMW-PTPs such as Wzb, involve in the positive regulation of capsules and/or exopolysaccharide that are important for bacterial pathogenesis (51). Therefore, the Wzb structure provides a clue for the rational design of drugs that are highly specific to bacterial virulence factors without leading to side effects in the eukaryotic organisms.

Acknowledgments—All NMR experiments were carried out at the Beijing Nuclear Magnetic Resonance Center (BNMRC), Peking University.

REFERENCES
1. Kennelly, P. J., and Potts, M. (1999) Front. Biosci. 4, D372–D385
2. Cozzone, A. J., Grangeasse, C., Doublet, P., and Duclos, B. (2004) Arch. Microbiol. 181, 171–181
3. Chiarugi, P., Cirri, P., Marra, F., Raugei, G., Fiaschi, T., Camici, G., Manao, G., Romanelli, R. G., and Ramponi, G. (1998) J. Biol. Chem. 273, 6776–6785
