Supporting Information for:
“Solution NMR structure of the TatA component of the twin-arginine protein transport system of Gram-positive bacterium *Bacillus subtilis*”
by Yunfei Hu, Enwei Zhao, Hongwei Li, Bin Xia and Changwen Jin

*SI Table 1. Structural statistics of *Bacillus subtilis* TatA<sub>d</sub>*

<table>
<thead>
<tr>
<th>Experimental restraints</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total NOE</td>
<td>1878</td>
</tr>
<tr>
<td>Total unambiguous NOE</td>
<td>1388</td>
</tr>
<tr>
<td>Intra-residue</td>
<td>598</td>
</tr>
<tr>
<td>Inter-residue</td>
<td></td>
</tr>
<tr>
<td>Sequential (</td>
<td>i-j</td>
</tr>
<tr>
<td>Medium-range (1&lt;</td>
<td>i-j</td>
</tr>
<tr>
<td>Long-range (</td>
<td>i-j</td>
</tr>
<tr>
<td>Total Ambiguous NOE</td>
<td>490</td>
</tr>
<tr>
<td>Total lower distance restraints</td>
<td>33</td>
</tr>
<tr>
<td>Total dihedral angle restraints (φ + ψ)</td>
<td>67 (33 + 34)</td>
</tr>
</tbody>
</table>

**Restraint Violations**
- Distance restraint violations > 0.5 Å: 0
- Dihedral angle restraint violations > 5º: 0

**Root mean square deviation (RMSD) from mean structure (Å)**
- All heavy atoms: 1.82
- All backbone atoms: 1.45
- Secondary structure heavy atoms: 0.82
- Secondary structure backbone atoms: 0.66

**Ramachandran Statistics (%)**
- Most favored regions: 94.5
- Additional allowed regions: 4.9
- Generously allowed regions: 0.5
- Disallowed regions: 0.1

*The structural statistics reported here corresponds to the Met1-Glu52 region of BsTatA<sub>d</sub> protein, while the C-terminal unstructured tail is excluded.*
**SI Figure 1**

Backbone assignments of BsTatA$_d$. The figure shows 2D $^{15}$N-edited HSQC spectrum of BsTatA$_d$ in deuterated DPC micelles annotated with backbone assignments. The folded peaks are labeled in blue. The peaks originated from C-terminal His-tag are labeled with an asterisk. The spectrum was collected at 25 °C in a sodium phosphate buffer (pH 7.0).

**SI Figure 2**

CD spectrum of *B. subtilis* TatA$_d$ in DPC micelles. The spectrum was recorded on a Mos-450 spectropolarimeter (Bio-Logic, France) at room temperature using a 0.1 cm optical path length cell. The spectrum was collected over a wavelength range from 190 nm to 250 nm, using a step size of 1 nm and averaged from three scans.
**SI Figure 3**

Structural analyses of BsTatA<sub>d</sub> protein. The upper panel shows the total energy of each of the 200 conformers calculated using Xplor-NIH\textsuperscript{1}. The lower panel shows the Pearson’s linear correlation coefficient between input and back calculated backbone N-H RDC values for each of the 200 conformers obtained by using the program PALES\textsuperscript{2}. Red and blue dotted lines correspond to R value of 0.85 and 0.80, respectively. The majority of the conformers show correlation coefficient higher than 0.80 with measured RDC values, and the 100 conformers with lowest energy generally show correlation coefficient higher than 0.85 with measured RDC values.

**SI Figure 4**

Fitting of backbone N-H RDCs to BsTatA<sub>d</sub> structures. The figure shows four representative fitting results of backbone N-H RDC values to BsTatA<sub>d</sub> structures. The structures were
calculated based on NOE and dihedral angle restraints only. The plots were generated using the program MODULE\(^3\). The correlation coefficient R values between experimental and back calculated RDCs in (A) (B) (C) and (D) are 0.902, 0.908, 0.901 and 0.898 respectively as determined by the program PALES\(^3\).

**SI Figure 5**

SI Figure 5. Cylinder representation of the kink at the last turn of TMH. The protein backbone is shown in line. The two segments of TMH before and after residue Ile16 are shown as grey cylinders, revealing a slight kink at this position.

**SI Figure 6**
**SI Figure 6.** Multiple sequence alignments of *B. subtilis* TatA proteins with TatA (A) and TatB (B) families. The sequences used in the alignments are TatA$_d$ (TATAD_BACSU) and TatA$_y$ (TATAY_BACSU) proteins from *B. subtilis*, TatA (TATA_ECOLI) and TatB (TATB_ECOLI) from *E. coli*, TatA (TATA_HAEIN) and TatB (TATB_HAEIN) from *Haemophilus influenzae*, TatA (TATA_HELPJ) and TatB (TATB_HELPJ) from *Helicobacter pylori*, TatA (TATA_MYCTU) and TatB (TATB_MYCTU) from *Mycobacterium tuberculosis*. The alignments were drawn using ESPRIPT$^d$.

**SI Figure 7**

**SI Figure 7.** Schematic model of membrane bound BsTatA$_d$ complex. The large disc with a diameter of 100 Å represents the circular shaped particles of membrane embedded BsTatA$_d$ complex observed by electron microscopy$^3$. One copy of L-shaped BsTatA$_d$ monomer is shown in red.
**SI Methods**

**Cloning and expression of BsTatA<sub>d</sub> protein.** The *Bacillus subtilis* tat<sub>A</sub> gene was cloned into the pET21a(+) vector (Novagen) and expressed in *Escherichia coli* BL21(DE3) strain (Novagen) with a C-terminal fused His-tag. The cell culture was grown overnight in 4 mL of Luria-Bertani media with ampicillin at 35 °C, centrifuged and resuspended in 200 mL M9 minimal medium. For the production of $^{13}$C/$^{15}$N-labeled samples, $^{15}$NH$_4$Cl and $^{13}$C$_6$ glucose were used in the M9 medium. When OD$_{600}$ reached 0.6, protein expression was induced by adding isopropyl-β-D-thiogalactoside (IPTG) at a final concentration of 200 mg/L. The cells were harvested after 6 h induction at 35 °C and frozen at -80 °C.

**Purification of BsTatA<sub>d</sub> protein and sample preparation.** The cells containing over-expressed BsTatA<sub>d</sub> protein were sonicated and treated with lysozyme for 1 h at 4 °C. After centrifugation (17,000 g for 20 min at 4 °C), the soluble components were resolved in buffer A (20 mM Tris, 200 mM NaCl, pH 8.0) containing 0.5 % AZ314 (Anatrace, Affymetrix) for 1 h. The AZ314 solubilized sample was mixed with 2 mL Ni-NTA resin (Qiagen), and sequentially washed with 20 mL buffer A containing 0.1% AZ314, 15 mL buffer A containing 0.1% AZ314 and 20 mM imidazole, 10 mL buffer A containing 0.2% DPC and 20 mM imidazole. Finally, BsTatA<sub>d</sub> protein was eluted using 10 mL buffer A containing 0.5% DPC (or perdeuterated DPC) and 200 mM imidazole. Purified BsTatA<sub>d</sub> in DPC was exchanged into 50 mM sodium phosphate buffer (pH 7.0). 10% of D$_2$O was added for preparations of the NMR samples, and 2,2-dimethyl-2-silapentanesulfonic acid was added as the internal chemical shift reference.

**NMR spectroscopy.** The triple resonance NMR experiments were performed at 25 °C on Bruker Avance 500 and 800 MHz spectrometers equipped with four RF channels and
triple-resonance cryo-probes with pulsed-field gradients. The chemical-shift assignments of backbone and side-chain atoms were obtained by 2D $^{15}$N- and $^{13}$C-edited HSQC experiments and 3D HNCA, HNCACB, HNCO, HN(CA)CO, CBCACONH, HBHACONH, (H)CC(CO)NH, (H)CCH-COSY, and (H)CCH-TOCSY experiments\(^6\) using $^{13}$C/$^{15}$N labeled BsTat\(A_d\) sample in deuterated DPC. 3D $^{15}$N- and $^{13}$C-edited NOESY-HSQC spectra as well as the $^{13}$C-edited NOESY-HSQC spectrum optimized for aromatic resonances (mixing times of 75 ms) for BsTat\(A_d\) in deuterated DPC micelles were collected to confirm the chemical-shift assignments and generate distance restraints for structure calculations. The NMR spectra were processed using NMRPipe\(^7\) and analyzed by NMRView\(^8\).

**RDC measurements.** The backbone N-H RDC measurements of BsTatAd was performed using the liquid crystalline phase of G-tetrad DNA\(^9\). The RDCs were extracted from the difference in $^1$H-$^{15}$N splitting measured by $^1$H-$^{15}$N IPAP-HSQC spectra between the weakly aligned and the isotropic samples at 35 °C\(^10\). The data were analyzed using software packages PALES\(^2\) and MODULE\(^3\).

**Backbone $^1$H-$^{15}$N heteronuclear NOE measurements.** The backbone steady-state heteronuclear $^1$H-$^{15}$N NOE values of BsTat\(A_d\) protein were measured on a Bruker Avance 800 MHz NMR spectrometer at 35 °C\(^11\). The experiments were performed in the presence and absence of a 3-s proton presaturation period prior to the $^{15}$N excitation pulse.

**Structure calculations.** The structure calculation of BsTat\(A_d\) (segment Met1-Glu52) was carried out based on exclusively inter-proton NOE-derived distance restraints and dihedral angle restraints. The C-terminal tail (K53-G70) is largely unstructured based on analysis of the NOESY spectra, and thus excluded in the structure calculation. The 3D
NOESY-HSQC spectra were manually analyzed to generate inter-proton distance restraints. In addition, 33 lower distance restraints were added for residues at the hinge region based on the absence of NOE cross-peaks. The program TALOS was used to predict dihedral angles $\psi$ and $\varphi$ restraints\textsuperscript{12}.

The initial structure calculations were iteratively performed using the program DYANA\textsuperscript{13} for several rounds. In each round, one hundred structures were calculated and the 20 conformers with the lowest target function were selected for structural and violation analysis. We also calculated the structures using the simulated annealing approach in torsion angle space with the program Xplor-NIH\textsuperscript{1}. The results from these two programs were generally similar. However, the structures generated by Xplor-NIH sampled larger conformational space than those by DYANA. Therefore, subsequent calculations were performed using Xplor-NIH. In each calculation cycle, 200 structures were generated based on NOE and dihedral angle restraints. One hundred conformers with lowest total energy were selected for fitting with experimentally measured backbone N-H RDCs. The fitting of RDC values with calculated structures were performed using the single value decomposition (SVD) method by the program PALES\textsuperscript{2}. Conformers with correlation coefficient R value higher than 0.85 were subjected to further structural and violation analysis. In the final round of calculation, 70 out of the 100 lowest-energy conformers showed R values higher than 0.85. Therefore, 40 conformers among these with the lowest backbone RMSD to mean structure were selected as representative structures for the BsTatA\textsubscript{d} protein. Finally, the structures were analyzed by PROCHECK\textsubscript{NMR}\textsuperscript{14} and MOLMOL\textsuperscript{15}.
REFERENCES


