Solution structure of *Escherichia coli* HypC

Lei Wang a,b, Bin Xia a,b,c, Changwen Jin a,b,c,*

a Beijing Nuclear Magnetic Resonance Center, Peking University, Beijing 100871, China
b College of Life Sciences, Peking University, Beijing 100871, China
c College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

Received 12 July 2007
Available online 26 July 2007

**Abstract**

*Escherichia coli* HypC plays an important role in the maturation process of the pre-maturated HycE, the large subunit of hydrogenase 3. It serves as an iron transfer as well as a chaperone protein during the maturation process of pre-HycE, and interacts with both HypD and HycE. The N-terminal cysteine residue of HypC plays a key role in the protein–protein interactions. Here, we present the three-dimensional structure of *E. coli* HypC, the first solution structure of HupF/HypC family. Our result demonstrates that *E. coli* HypC consists of a typical OB-fold β-barrel with two C-terminal helixes. Sequence alignment and structural comparison reveal that the hydrophobic region on the surface of *E. coli* HypC, as well as the highly flexible C-terminal helixes, may involve in the interactions of *E. coli* HypC with other proteins.

© 2007 Elsevier Inc. All rights reserved.

**Keywords:** HypC; Hydrogenase; NMR; Solution structure; OB-fold

As a member of the metalloenzyme family, *Escherichia coli* hydrogenase 3 harbors a dinuclear [NiFe] metal center for H₂ production under fermentative growth at low pH. As the large subunit of the hydrogenase 3, HycE forms a metal center consisting of a nickel atom, an iron atom, and four cysteine residues [1]. To form such an unusual metal center, pre-maturated HycE undergoes a complex maturation process including the incorporation of iron atom, the insertion of nickel atom, and the cleavage of a short C-terminal peptide [1].

*Escherichia coli* HypC (EcHypC) is encoded by the hypC gene of the hyp operon. It belongs to the HupF/HypC family and contains 90 amino acid residues. At the early stages of pre-HycE maturation, EcHypC interacts with *E. coli* HypD (EcHypD) to form a complex containing an iron atom. Subsequently, the HypC–HypD complex interacts with other proteins to transfer CN and CO ligands to the iron atom. In the next step, this complex dissolves and EcHypC transfers the liganded iron to pre-HycE [2].

EcHypC binds with pre-HycE and performs a chaperone-like function to facilitate the insertion of nickel atom [3,4]. Finally, after the insertion of Ni, EcHypC leaves off pre-HycE and the C-terminal peptide of pre-HycE is subsequently cleaved by *E. coli* HycI [1].

The cysteine residue of EcHypC plays an important role during the maturation process of pre-HycE. After translation, the first residue Met1 of HypC is removed, leaving an active cysteine as its N-terminal residue [4]. Site-directed mutagenesis experiments showed that this active cysteine residue is responsible for the interactions of EcHypC with both EcHypD and pre-HycE [1,4].

During the preparation of this manuscript, the crystal structure of *Thermococcus kodakaraensis* HypC (TkHypC) was reported [5]. The crystal structure of TkHypC, together with that of *T. kodakaraensis* HypD (TkHypD) and HypE (TkHypE), provides insights into the structural basis of hydrogenase maturation [5]. However, most of our present knowledge of hydrogenase maturation is based on *E. coli* hydrogenase 3. Although the first 70 residues of EcHypC share 36% identities to TkHypC, the difference of conserved cysteine residues between EcHypD and

---

* Corresponding author. Fax. +86 10 6275 3790.
E-mail address: changwen@pku.edu.cn (C. Jin).
TkHypD implies a different mechanism for diatomic ligands of the active-site Fe atom in *E. coli* [5].

Herein, we present the solution structure of EcHypC determined by NMR spectroscopy, the first solution structure of HupF/HypC family. Our result shows that EcHypC adopts an OB-fold with two C-terminal helices. Although the fold of EcHypC is similar to that of the recently reported TkHypC, EcHypC has a longer loop L34 covering on one end of the OB-fold β-barrel and an additional C-terminal α-helix. Our results provide the structural basis for further investigations on *E. coli* hydrogenase maturation.

Materials and methods

Cloning, expression, and purification of *E. coli* HypC. The hypC gene from *E. coli* was amplified by PCR, cloned into the plasmid pET21a(+) (Novagen), 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 500 ml of LB medium containing 100 µg/ml ampicillin. When the *A*ₐₒₒ reached around 0.8, cells were centrifuged and resuspended in 250 ml of M9 minimal medium containing 15N-labeled ammonium chloride in the presence or absence of 13C-labeled glucose for the assignments and generate distance restraints for structure calculations. NOESY-HSQC (mixing times 100 ms) spectra were recorded to confirm the assignments of side-chain atoms. The chemical shift assignments of side-chain atoms were further confirmed by the three-dimensional 15N-edited TOCSY-HSQC (mixing time 80 ms) spectrum [8]. The three-dimensional 15N- and 13C-edited NOESY-HSQC spectra and backbone dihedral angle restraints determined by TALOS [9]. The initial structures of EcHypC were calculated with the CANDID module of the CYANA program [10,11]. The backbone chemical shift assignments of EcHypC were obtained by the program NMRView [7]. The two-dimensional 15N-edited HSQC, three-dimensional triple-resonance spectra HN(CACB), CBCA(CO)NH, HCCH-COSY, HCCH-TOCSY, (H)CCH-COSY, and (H)CCH-TOCSY were recorded to assign the side-chain atoms. The chemical shift assignments of side-chain atoms were further confirmed by the three-dimensional 15N-edited TOCSY-HSQC (mixing time 80 ms) spectrum [8]. The three-dimensional 15N- and 13C-edited NOESY-HSQC (mixing times 100 ms) spectra were recorded to confirm the assignments and generate distance restraints for structure calculations.

Structure calculations. The solution structure of EcHypC was calculated using the inter-proton distance restraints obtained from three-dimensional 15N- and 13C-edited NOESY-HSQC spectra and backbone dihedral angle restraints determined by TALOS [9]. The initial structures of EcHypC were calculated with the CANDID module of the CYANA program [10,11]. The extended NOE restraints and chemical shift assignments were obtained by an iterative analysis of the calculated structures and NOE assignments using the program SANE [12]. At the final stage of structure calculations, two hundred structures were calculated using DYANA [10]. The 100 structures with the lowest target function were selected and further refined by AMBER [13]. Finally, the 20 lowest energy structures were selected to represent EcHypC. The structures were analyzed using the software packages MOLMOL [14] and PROCHECK_NMR [15].

Results

Chemical shift assignments

By using triple-resonance NMR experiments, the backbone chemical shift assignments were obtained for 85 out of 90 residues of EcHypC. The unassigned residues include Met1, Cys2, Pro6, Pro42, and Pro79. More than 96% of the chemical shift assignments for side-chain atoms of EcHypC were obtained. The chemical shift assignments of EcHypC have been deposited in the BioMagResBank database under Accession No. 15152.

Solution structure of EcHypC

The solution structure of EcHypC was determined based on a total of 3128 inter-proton NOE-derived distance restraints and 88 backbone dihedral angle restraints. The structural statistics is shown in Table 1. A superimposition of 20 representative structures, together with the ribbon diagram representation of the lowest energy conformer, is shown in Fig. 1A and Fig. 1B, respectively. The 20 lowest energy structures have a root mean square deviation (RMSD) of 0.92 Å for backbone atoms of residues Gly4 to Asp76, and 0.61 Å for residues in the secondary structures. The coordinates of EcHypC structures have been deposited in the Protein Data Bank (PDB entry 2OT2).

Table 1

<table>
<thead>
<tr>
<th>Structural statistics of <em>E. coli</em> HypC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance restraints</td>
</tr>
<tr>
<td>Intraresidue unambiguous NOEs</td>
</tr>
<tr>
<td>Sequential unambiguous NOEs</td>
</tr>
<tr>
<td>Medium range unambiguous NOEs</td>
</tr>
<tr>
<td>Long range unambiguous NOEs</td>
</tr>
<tr>
<td>Total unambiguous NOEs</td>
</tr>
<tr>
<td>Total ambiguous NOEs</td>
</tr>
<tr>
<td>Dihedral angles (φ and ψ)</td>
</tr>
<tr>
<td>Root mean square deviation from mean structure (Å)</td>
</tr>
<tr>
<td>Secondary structure backbone atoms</td>
</tr>
<tr>
<td>All backbone atoms (residues 4–76)</td>
</tr>
<tr>
<td>All heavy atoms (residues 4–76)</td>
</tr>
<tr>
<td>Energy (kcal/mol)</td>
</tr>
<tr>
<td>Mean AMBER energy</td>
</tr>
<tr>
<td>NOE distance restraint violation energy</td>
</tr>
<tr>
<td>Torsion angle restraint violation energy</td>
</tr>
<tr>
<td>Distance violations (°)</td>
</tr>
<tr>
<td>Dihedral angle (°)</td>
</tr>
</tbody>
</table>

Ramachandran statistics

| Residues in most favored regions (%)  | 82.3 |
| Residues in additional allowed regions (%) | 16.9 |
| Residues in generously allowed regions (%) | 0.3 |
| Residues in disallowed regions (%)     | 0.5 |
EcHypC consists of five β strands (β1, from Gly4 to Ile12; β2, from Gln16 to Asp20; β3, from Gln25 to Asp29; β4, from Trp47 to His51; β5, from Phe54 to Ile59) flanked by two C-terminal α-helixes (H1, from Glu61 to Asp76, H2, from Asp80 to Leu85). The five β strands are arranged with a Greek key topology in the order 3-2-1-4-5. They form an anti-parallel β-barrel with a flexible loop between β3 and β4 (L34, from Leu30 to Gln46) covering one end of the barrel. Structural analysis of this β-barrel indicates that EcHypC adopts a canonical oligonucleotide/oligosaccharide binding fold (OB-fold) [17]. In addition, the strands β1 and β5 contain β-bulges at Arg10 and Met56, respectively. H1 and H2 locate beside the β-barrel and are separated by three residues (from Val77 to Pro79). Moreover, H2 is relatively flexible reflected by the lack of NOE contacts.

**Structural comparison with TkHypC**

The overall fold of EcHypC is similar to that of TkHypC. In EcHypC, the conserved hydrophobic residues locate in strand β1, β4, β5 and the start of L34 which form a hydrophobic region on one side of the β-barrel. A similar hydrophobic region is observed in TkHypC. Moreover, there is also a conserved histidine residue (H251) near the N-terminal cysteine of EcHypC (Fig. 2B). However, significant diversities are observed, especially for the L34 and the C-terminal regions between EcHypC and TkHypC. The sequence of EcHypC has a 6-residue-insertion in L34 compared to TkHypC, resulting in a longer loop L34 in EcHypC. In addition, the longer C-terminal sequence of EcHypC leads to the formation of a second α-helix (H2) that is not present in the structure of TkHypC.

**Conformational flexibility on fast timescales**

The experimentally determined {^1H}–^{15}N NOE values versus the amino acid sequence are shown in Fig. 1C. A total of 65 residues of EcHypC were used during the analysis. The unanalyzed residues include three proline residues that have no amide protons, unassigned residues (Met1 and Cys2), and 20 residues that were either overlapped or too weak to be analyzed. The RMSD values of backbone nitrogen atoms are also shown in Fig. 1C. A good correlation is observed between the RMSD and the heteronuclear {^1H}–^{15}N NOE values.

The β-barrel of EcHypC adopts a rigid structure as reflected by the relatively high {^1H}–^{15}N NOE values for most of the residues in the five β strands. Residues in H1 and H2 show significant flexibility on pico- to nanosecond timescales, reflected by both the large RMSD and low {^1H}–^{15}N NOE values. In addition, residues Ile3 and Gly45, which locate near the hydrophobic knob, also show low NOE values, indicating the fast timescale motional flexibility at this region.

**Discussion**

**Sequence alignment of proteins of the HupF/HypC family**

Multiple sequence alignment (Fig. 2A) shows that proteins of the HupF/HypC family contain many conserved hydrophobic residues, in addition to the conserved N-terminal motif MC(L/I/V)(G/A)(L/I/V)P[4]. In both EcHypC and TkHypC, these conserved residues form a hydrophobic region on one side of the β-barrel (Fig. 2B), in which the conserved N-terminal cysteine is located. Some of these conserved residues are involved in the interaction surface in the docking models of *T. kodakaraensis* HypC–HypD complex [5]. Thereby, the hydrophobic region formed by conserved residues may be a common feature of the HupF/HypC family proteins that facilitate the protein interactions.

There are five highly conserved residues in HupF/HypC family (Met1, Cys2, Pro6, Gly45, and Ala64) in addition to those hydrophobic residues. Among these residues, Gly45 locates before strand β4 in EcHypC, which is a common feature of the OB-fold proteins [18]. Two other residues, Pro6 and Ala64, are imbedded in the cliff formed by...
Fig. 2. (A) Sequence alignment of 19 proteins from HupF/HypC family proteins. The proteins were selected by protein database Pfam [19]. The secondary structural elements of EcHypC are shown on the top. The conserved residues are highlighted in boxes (red characters on a white background) and residues that are strictly conserved in the column are shown in white characters on a red background. The 6-residue-insertion in the middle of the top four protein sequence is enveloped by a red rectangle. The alignment was performed using the program ClustalW [20] and ESPript [21]. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)

(B) The electrostatic surface of EcHypC. The conserved N-terminal cysteine, Cys2, the histidine, His51, and the valine, Val52 are annotated with the one-letter amino acid code with the residue number. The graphics were generated using MOLMOL [14]. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)
H1 and the β-barrel, stabilizing the conformation of the protein. This may also explain the previous observation that the activity of EcHypC does not abolish after mutating the conserved proline to residues with short side-chain residues [4].

The sequence alignment also reveals that four HupF/HypC proteins, including EcHypC, contain a 6-residue-insertion in their sequences (Fig. 2A). A close inspection shows that this insertion mainly exists in bacteria such as Enterobacteriales and Pasteurellales. Moreover, the HypD proteins in these bacteria harbor four conserved cysteine residues in their C-terminal part. As indicated by the crystal structure of TkHypD, the lack of the two conserved cysteine residues in the C-terminal part may result in a different mechanism of diatomic ligands of the active-site Fe atom [5]. Thus, this insertion may lead to the diversity of the mechanisms for diatomic ligands of the active-site Fe atom. In EcHypC, this insertion locates in the long loop L34, which forms a negative charged region. However, the role of this long loop remains unclear so far.

Functional implications

The C-terminal region including the two helices of EcHypC are quite flexible as reflected by the large RMSD and low \[^1H\]–\[^15N\] NOE values, indicating the conformational flexibility on the picosecond to nanosecond timescales. This observation is in good agreement with that of TkHypC [5]. Therefore, the conformational flexibility of C-terminal region may be a common feature of the HupF/HypC proteins. In addition, EcHypC shares 50% sequence identity with E. coli HybG (EcHybG); however, they show distinct affinity to their target protein EcHypD. The major difference between the two proteins is that EcHybG has a shorter C-terminal part than EcHypC, which may result in shorter helix(es). Moreover, the docking model of T. kodakaraensis HypC–HypD complex shows that the C-terminal helix of TkHypC locates near the interface of TkHypC and TkHypD [5]. Therefore, the C-terminal region of EcHypC plays an important role in the protein interactions, and its conformational flexibility may facilitate the association with their target proteins.

Acknowledgments

All NMR experiments were carried out at the Beijing Nuclear Magnetic Resonance Center (BNMRC), Peking University. We thank You Li, Drs. Xinxin Zhang, and Xianrong Guo (BNMRC) for helpful discussions. This research was supported by Grant 2006AA02A323 from the National High Technology Research and Development Program of China to C.J.

References


