Recombinant Globular Domain of TcpA Pilin from Vibrio cholerae El Tor: Recovery from Inclusion Bodies and Structural Characterization

Victor Marchenkov 1, Elena Dubovitskya 2,†, Nina Kotova 2, Igor Tuchkov 2, Nina Smirnova 2, Natalia Marchenko 1©, Alexey Surin 1,3,4, Vladimir Filimonov 1 and Gennady Semisotnov 1,*

1 Institute of Protein Research, Russian Academy of Sciences, Institutskaya Street 4, 142290 Pushchino, Russia
2 Russian Antiplague Scientific Research Institute “Microbe” of Rospotrebnadzor, Universitetskaya Street 46, 410005 Saratov, Russia
3 Branch of the Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Prospekt Nauki 6, 142290 Pushchino, Russia
4 State Research Centre for Applied Microbiology and Biotechnology, 142279 Obolensk, Russia
* Correspondence: nina@vega.protres.ru; Tel.: +7-(496)-731-8409
† Current address: Research Center “PhytoEngineering”, Moskovskaya Street 58, 141880 Rogachevo, Russia.

Abstract: The production of recombinant proteins in Escherichia coli cells is often hampered by aggregation of newly synthesized proteins and formation of inclusion bodies. Here we propose the use of transverse urea gradient electrophoresis (TUGE) in testing the capability of folding of a recombinant protein from inclusion bodies dissolved in urea. A plasmid encoding the amino acid sequence 55–224 of TcpA pilin (C-terminal globular domain: TcpA-C) from Vibrio cholerae El Tor enlarged by a His-tag on its N-terminus was expressed in E. coli cells. The major fraction (about 90%) of the target polypeptide was detected in cell debris. The polypeptide was isolated from the soluble fraction and recovered from inclusion bodies after their urea treatment. Some structural properties of the polypeptide from each sample proved identical. The refolding protocol was developed on the basis of TUGE data and successfully used for the protein large-scale recovery from inclusion bodies. Spectral, hydrodynamic, and thermodynamic characteristics of the recombinant TcpA recovered from inclusion bodies indicate the presence of a globular conformation with a pronounced secondary structure and a rigid tertiary structure, which is promising for the design of immunodiagnostics preparations aimed to assess the pilin level in different strains of V. cholerae and to develop cholera vaccines.

Keywords: inclusion bodies; protein purification; protein refolding; protein thermodynamics; Vibrio cholerae pilin

1. Introduction

Cholera, an extremely serious infectious disease, continues to pose a constant threat to the population of developing countries around the world. All pandemics have been caused by toxigenic strains of the Vibrio cholerae (V. cholerae) O1 serogroup, but different biovars, classical and El Tor, which differ in their phenotypic and molecular-genetic properties [1,2]. V. cholerae of the classical biovar was probably the causative agent of the first six pandemics from 1817 to 1923; later, it was superseded by V. cholerae biovar El Tor causing the current seventh cholera pandemic that started in 1961 [3].

At present, it has been established that the V. cholerae El Tor agent gene tcpA is represented by two main alleles: tcpA^ET and tcpA^CIRS [4,5]. The tcpA^ET allele is typical for pathogenic strains that caused cholera epidemics in 1961 to 2002, while tcpA^CIRS allele was found later in the highly pathogenic strains causing the 2003–present day epidemics in many regions, including Russia [4,5]. Unlike tcpA^ET, tcpA^CIRS carries one nucleotide substitution of adenine for guanine (A/G) at position 266 of the tcpA gene, which leads to a...
change in the TcpA aminoacid sequence at position 89 where asparagine is replaced by serine (Figure S1, alignment by BLAST [6]). The functional significance of this substitution has not been precisely determined yet. Nevertheless, the wide distribution of the pathogenic gene variants with such a modification in the tcpA structure within different regions of the world (including the territory of Russia in recent years) is likely due to the displacement of other strains of the cholera pathogen. This was the main reason for our choice of the tcpA_CIRS allele for creating a strain-producer of the encoded protein. This work is aimed to construct a recombinant strain of E. coli which is the producer of the TcpA protein of the cholera El Tor agent carrying the tcpA_CIRS gene.

To date, several recombinant strains of E. coli are known to be designed to produce the TcpA protein encoded by different alleles (but not tcpA_CIRS) for the subsequent creation of cholera vaccines on their basis [7–10]. However, in the Russian Federation, there is no such strain-producer of cholera vaccines of new generation and the immunodiagnostic test system required to assess the level of pathogenicity produced by epidemically dangerous strains.

The pathogenicity of a cholera agent, regardless of the biovar, is provided by the action of two biomolecules that are the main virulence factors, namely the toxin-coregulated pilus (TCP) and cholera toxin (CT). The major subunit of TCP is coded by the nucleotide sequence of the tcpA gene (672 nucleotide pairs), which in El Tor is only 82% identical to the classical vibrios [11].

This study is focused on TcpA pilin which plays an important role in the pathogenesis of cholera [12,13]. Toxigenic strains of V. cholerae that do not produce this protein are unable to colonize the human small intestine and to develop the subsequent infection [12]. This underlies the design of TcpA-based immunodiagnostic preparations aimed to assess the level of TCP expression in modern variants of the pathogen and to determine their ability to initiate the cholera infection by colonizing the small intestine. Moreover, it has been proven that TcpA is one of the key protective antigens whose antibodies are able to block the attachment of V. cholerae to epithelial cells of the small intestine, thus preventing the development of cholera infection [14,15]. All these facts indicate that the protein TcpA can be used to develop cholera vaccines.

It is known that the entire TcpA polypeptide synthesized in V. cholerae contains 224 amino acids [11,13]. The first 25 amino acid residues belong to the signal peptide that is cleaved off upon protein transfer into the periplasm. The remaining 199 amino acid residues correspond to the “mature” pilin (mTcpA) forming pili. The first 52 residues of mTcpA fold into a long α-helix (α1) present in all pili of the IVb class. This long helix can be subdivided into two segments, α1-N and α1-C [13,16]. The N-terminal segment (α1-N, 28 amino acids long) responsible for pilin oligomerization has a highly conserved and hydrophobic primary structure. In contrast, α1-C belongs to the globular C-terminal domain of TcpA (about 171 amino acids in total) with a high content of the secondary structure and the sole disulfide bond between Cys145 and Cys211. α1-N participates in polymerization and interferes with protein crystallization; therefore, only the globular C-terminal domain of TcpA (TcpA-C) was used as a recombinant protein, purified, and crystallized. Its 3D structure was determined at a resolution of 1.3–1.6 Å [13,16].

We decided to develop our own protocol of TcpA-C (tcpA_CIRS gene) large-scale purification and to characterize its physical and chemical properties in solution for the purpose of its subsequent use in antibody preparation.

2. Materials and Methods
2.1. Bacterial Strains and Cultivation Conditions

Previously, we constructed a recombinant strain Escherichia coli BL21 Star™(DE3)/pET302tcpA (E. coli BL21 Star™ from Thermo Fisher Scientific, Waltham, MA, USA) using the plasmid expressing a part of the tcpA_CIRS gene fused with His-Tag (total length 534 bp) that was placed under the transcriptional control of the phage T7 promoter [17] (Figure S2). The donor of the tcpA_CIRS gene was the strain V. cholerae M1429 biovar El Tor that is an atypi-
cal variant of the pathogen isolated from a Russian patient in 2005 (NCBI GenBank accession number: LAEM00000000.1, https://www.ncbi.nlm.nih.gov/nuccore/LAEM00000000.1 (accessed on 5 November 2022)). The polymerase chain reaction (PCR) and calculated primers (F ccgctcgaggattcgcagaatatgactaaggctgc and R gctggatccttaactgttaccaaaagctactgt-gaat) were used to incorporate the tcpA\textsuperscript{CIRS- C} fragment into the expression plasmid pET302. The nucleotide sequence of tcpA\textsuperscript{CIRS- C} was checked and found to fit the known data on the tcpA\textsuperscript{CIRS} gene [4]; the sequencing was performed using an ABI3500xl automated sequencer (Applied Biosystems, Waltham, MA, USA) using BigDye\textsuperscript{®} Terminator v3.1 Cycle Sequencing Kit. In our vector, the corresponding N-terminal polypeptide sequence was fMHHHHHH-S55-Q56-N57-M-58- . . . (Figures S1 and S2), so the expressed polypeptide TcpA-C was 177 amino acids long for lack of N-terminal 25 a.a. of the leader peptide and 28 a.a. of the α1-N segment responsible for the pilin polymerization (Figure S1). The molecular weight of this polypeptide computed at https://web.expasy.org/protparam/ (accessed on 5 November 2022) [18] was 18.2 kDa. The bacteria were cultured in broth and agar LB medium, pH 7.4, at 30 °C or 37 °C without antibiotics or with ampicillin (50 µg/mL). Isopropyl-β-D-thiogalactoside (IPTG) (Invitrogen, Waltham, MA, USA) was used as an inductor of expression of the cloned tcpA\textsuperscript{CIRS} gene [17]. The optimal conditions for cultivating the recombinant strain were the liquid medium LB (pH 7.4), 50 µg/mL ampicillin, enhanced aeration (200 rpm), 37 °C, 2 h, and an IPTG concentration of 0.1 mM.

2.2. Preparation of Cell Lysates

E. coli cells producing the target protein were precipitated by centrifugation at 10,000× g for 10 min, washed, and resuspended in a cold buffer with 10 mM Tris-HCl, pH 8.0, 0.3 M NaCl. The precipitated cells were lysed using an ultrasonic Branson 450 Digital Sonifier homogenizer (Branson Ultrasonic Corporation, Danbury, CT, USA) with ten cycles of 30 s ON and 30 s OFF at a frequency of 60 kHz in an ice bath. The separation of soluble and insoluble fractions of TcpA-C was carried out by centrifugation of the broken cells at 10,000× g for 10 min. The presence of protein in the precipitant and the supernatant was monitored by SDS-PAGE electrophoresis.

2.3. Purification of the Protein from the Soluble Fraction

The bulk of TcpA-C (about 90%) was detected in inclusion bodies. The remaining ~10% of the protein population localized to the soluble fraction of cell lysate and was purified in two steps: (1) by affinity chromatography on Ni-NTA resin (Profinia\textsuperscript{TM} Affinity Chromatography Protein Purification System; Bio-Rad Laboratories, Inc., Hercules, CA, USA), followed by (2) size-exclusion chromatography on a Superdex 200 column (GE Healthcare, Chicago, IL, USA), as described elsewhere [17]. At the first stage of purification, the soluble fraction of cell lysate was applied to the Ni-NTA column. Washing was carried out using Native IMAC Wash Buffer (Bio-Rad Laboratories, Inc., Hercules, CA, USA) with stepwise increasing concentrations of imidazole—30 mM (10.0 mL), 50 mM (10.0 mL), and 100 mM (10.0 mL)—to reduce nonspecific binding of contaminating proteins to the sorbent and increase the purity of the TcpA protein. Native IMAC Elution Buffer (Bio-Rad Laboratories, Inc., Hercules, CA, USA) with 250 mM imidazole was used in elution. At the second stage, size-exclusion chromatography on a Superdex 200 column (GE Healthcare, Chicago, IL, USA) used 20 mM Tris-HCl buffer, pH 8.9, and 150 mM NaCl. The peak containing the protein of interest with M₉ of about 20 kDa was collected, concentrated up to 0.5 mg/mL, and used for the subsequent analysis.

2.4. Recovery of the Target Protein from Inclusion Bodies through the Urea-Unfolded State

The inclusion bodies were thoroughly washed two times with 20 mM Tris-HCl, pH 8.9, 1% Deoxycholate; two times with 20 mM Tris-HCl, pH 8.9, 1.0 M urea; and six times with 20 mM Tris-HCl, pH 8.9. The washing was performed by resuspension in the corresponding buffer, followed by centrifugation at 12,100× g for 5 min, and led to essential removal of the protein impurities containing in the cell debris.
To remove the few remaining protein contaminants, we used Ni-NTA chromatography in 8.0 M urea, as the first step. TcpA-C fused with the His-tag has an isoelectric point of about 8.0 (which is close to the pH optimum of the resin); therefore, we raised the buffer pH up to 8.9 to increase the protein solubility during its refolding. This resulted in the increased non-specific protein binding to Ni-NTA, making it very difficult to remove some impurities from the protein preparations. After Ni-NTA chromatography, protein-containing fractions were collected and loaded on a Mono Q 5/50GL column, as the second step of the domain purification. The fractions enriched with the target protein were collected and concentrated to 150 ml using an Amicon centrifugal concentrator MWCO 10 kDa (Millipore Ireland Ltd., Tullagreen, County Cork Carrigtwohill, Ireland). The protein renaturation was performed by 20-fold dilution on ice, with intensive stirring, using the renaturation buffer optimized with a QuickFold™ Protein Refolding Kit from AthenaES™ (Athena Environmental Sciences, Inc., Baltimore, MD, USA) and containing 100 mM Tris-HCl, pH 8.9, 200 mM NaCl, 100 mM KCl, 2 mM MgCl₂, 2 mM CaCl₂, 1 mM GSH, 0.1 mM GSSH. The sulfhydryl reagents were added to facilitate the correct S-S bonding during the domain folding. The “refolded” protein solution was dialyzed overnight against 20 mM Tris-HCl, pH 8.9, 150 mM NaCl, and concentrated again by centrifugation in an Amicon centrifugal concentrator MWCO 10 kDa (Millipore Ireland Ltd., Tullagreen, County Cork Carrigtwohill, Ireland). After centrifugation for 5 min at 12,100 × g, the protein solution was loaded on a Superdex 200 column for size-exclusion chromatography (see below). The peak containing the protein of interest with Mᵣ of about 20 kDa (see below) was collected, concentrated up to 1.5 mg/mL, and stored at −25 °C for the subsequent analysis.

The total yield of purified recombinant TcpA determined according to [19], was ~60 mg/L of the initial culture.

2.5. Mass Spectrometry Analysis of the Recombinant Protein TcpA-C

Protein bands corresponding to TcpA-C were excised from the electrophoresis gel and treated independently from each other with four proteases (Trypsin and Chymotrypsin (Sigma-Aldrich, St. Louis, MO, USA), V8 and Proteinase K (Promega Corporation, Madison, WI, USA)) at 37 °C in a Thermo Mixer thermo shaker (Eppendorf, Hamburg, Germany). 5 mM CaCl₂ was added to stabilize proteinase K. The molar ratio of enzyme/protein was 1/50. The protein hydrolysis was stopped by adding trifluoroacetic acid. The resulting peptides were mixed and analyzed by high resolution tandem mass-spectrometry, after preliminary separation. The separation was performed by reversed-phase liquid chromatography in a gradient of acetonitrile from 4% to 60% for 120 min using an Easy nLC 1000 nanoliquid chromatograph (Thermo Fisher Scientific, Waltham, MA, USA) and a column packed under laboratory conditions with Aeris Widepore XV-C18 phase 3.6 μm particle size and 300 Å pore size. Mass spectra of the sample were obtained using an Orbi Trap Elite mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA). The peptides were ionized by electrospray at the nano-liter flow rate with 1.8 kV ion spray voltage and a capillary temperature of 200 °C. The fragmentation of ionized peptides was performed by the CID method in a high-energy cell (HCD).

The mass spectra were processed, and the peptides were identified using the PEAKS Studio (ver. 7.5) program. 111 TcpA-C peptides were identified by mass spectrometry analysis, and 106 among them appeared unique. The tolerance errors for parent mass and fragment mass were 5.0 ppm and 0.5 Da, respectively. Four variable modifications were used (o—oxidation of Met, d—dioxidation of Met, f—formylation and c—carbamidomethylation of Cys). The protein search was performed using the Uniprot database for bacterial proteins. The database contained 982,400 aminoacid sequences. The TcpA protein proved to be the major one, while others were in trace quantities.

2.6. Structural Characterization of the Recombinant Protein TcpA-C

All measurements (except the temperature dependence detection) were taken at room temperature.
The TcpA-C absorption spectra were recorded using a Cary100Bio spectrophotometer (Varian Medical Systems, Palo Alto, CA, USA). Protein concentration was measured using the molar extinction coefficient $E_{280\text{nm}} = 3105 \text{ M}^{-1}\text{cm}^{-1}$ evaluated according to [20]. The fluorescence spectra were measured on a Cary Eclipse spectrofluorimeter (Varian Medical Systems, Palo Alto, CA, USA) with $1 \times 1 \times 4 \text{ cm}$ quartz cells.

Spectra of Far-UV circular dichroism were recorded using a Chirascan Spectropolarimeter (Applied Photophysics Ltd., London, UK) at a protein concentration of 0.2 mg/mL and a quartz cell of 0.1 mm pathway length.

The TcpA-C secondary structure content was assessed using the BeStSel web server [21], CDNN CD deconvolution software 2.1 [22], K2D3 web server [23], DSSP [24], and SELCON [25] programs.

Transverse urea gradient electrophoresis was used for protein refolding/unfolding studies, as described by Creighton [26] and Goldenberg [27]. The gels were prepared using a 0.0–8.0 M urea gradient in 20 mM Tris-HCl, pH 8.9, buffer, and the samples were running for 4 h. After electrophoresis, the gels were stained with Coomassie brilliant blue G-250 and analyzed.

The size-exclusion chromatography was performed using an HPLC chromatograph ProStar (Varian Medical Systems, Palo Alto, CA, USA), a Superdex 200 10/300 GL column, and a flow rate of 0.4 mL/min. Protein detection was performed by tyrosine intrinsic fluorescence (excitation wavelength 275 nm, emission wavelength 315 nm).

The $^1$H NMR monodimensional spectrum of TcpA-C was measured using a Bruker Avance III HD 600 MHz NMR spectrometer (Bruker, Billerica, MA, USA). The working frequency was 600 MHz, with a spectrum width of 24 ppm and a 90° pulse of 11 µs. The number of accumulations was 150 at a protein concentration of 0.8 mg/mL and a temperature of 298 K. The water signal suppression was performed using WATERGATE with a delay before 90° pulse of 1–2 s.

The temperature dependence of the protein partial heat capacity $C_p(T)$ was recorded using a SCAL-1 differential scanning microcalorimeter (Scal, Pushchino, Russia) [28] with 0.3 mL glass cells at a protein concentration of about 1 mg/mL and a heating rate of 1 K/min.

The quantity of free SH groups was defined using Ellman’s reagent [29].

3. Results

3.1. Biosynthesis of TcpA-C in E. coli Cells and Its Purification

From SDS-PAGE data on TcpA-C biosynthesis and purification (Figure 1), the following conclusions can be drawn. First, the recombinant protein with a molecular weight close to that of TcpA-C is synthesized after IPTG induction, i.e., as a product of the expression of the constructed plasmid (lane 2). Second, the bulk of TcpA-C is detected in the precipitate (cf. lanes 3 and 4) as the major protein (cf. lanes 2 and 3). Third, the protein isolated from the soluble fraction by Ni-NTA affinity chromatography seems to be sufficiently electrophoretically pure (see lanes 4 and 5). Moreover, as confirmed by mass spectrometry analysis, the bands corresponding to proteins with molecular weights close to 18 kDa include TcpA-C (see Section 2, and Figure S3).

3.2. Physico-Chemical Characterization of the Recombinant Protein

To develop the protocol for the large-scale recovery of recombinant TcpA-C from inclusion bodies, we used the results of transverse urea gradient electrophoresis (TUGE). Monitoring the cooperative (S-like) unfolding or refolding transitions is a strategy to characterize the rigidity of a protein tertiary structure packing (i.e., the proper protein folding) [30]. In contrast to spectroscopic and other physicochemical methods, TUGE allows monitoring the change in protein electrophoretic mobility reflecting the change in the protein hydrodynamic volume during unfolding or refolding using a small sample (~50 µg) for a short time (~4 h).
spectrometry analysis, the bands corresponding to proteins with molecular weights close to 18 kDa include TcpA-C (see Section 2, and Figure S3).

Figure 1. SDS-PAGE of TcpA-C purification: molecular weight markers (lane 1); cell lysates of *E. coli* after IPTG induction (lane 2); proteins in the precipitate (lane 3) and in the soluble fraction (lane 4); the soluble fraction after Ni-NTA affinity chromatography (lane 5). Lane 5 originated from another gel was corrected according to the lane of molecular weight markers.

3.2. Physico-Chemical Characterization of the Recombinant Protein

To develop the protocol for the large-scale recovery of recombinant TcpA-C from inclusion bodies, we used the results of transverse urea gradient electrophoresis (TUGE). Monitoring the cooperative (S-like) unfolding or refolding transitions is a strategy to characterize the rigidity of a protein tertiary structure packing (i.e., the proper protein folding) [30]. In contrast to spectroscopic and other physicochemical methods, TUGE allows monitoring the change in protein electrophoretic mobility reflecting the change in the protein hydrodynamic volume during unfolding or refolding using a small sample (~50 µg) for a short time (~4 h).

Figure 2 shows the electrophoretic pattern of renaturation of a sample extracted from cell debris after its dissolution in a buffer containing 20 mM Tris-HCl, pH 8.9, 8.0 M urea, followed by centrifugation (Figure 2a), and that of denaturation of a sample isolated from the soluble fraction by Ni-NTA chromatography (Figure 2b). It should be emphasized that the recombinant TcpA-C, the major component of inclusion bodies, undergoes a relatively abrupt refolding in 3.0–1.0 M urea. Importantly, the refolding of TcpA-C from inclusion bodies is very similar to unfolding of this protein isolated from the soluble fraction after Ni-NTA chromatography (Figure 2b).

This means that the urea-induced folded-to-unfolded transition of the recombinant protein is a spontaneous and reversible process following the two-state model:

\[
N \rightleftharpoons U
\]

We derived the thermodynamic information from electrophoretic data by applying the simplest linear extrapolation method (LEM) to the S-shaped transition curves [31]. The electrophoresis data were digitalized (Figure 2) and fitted to the equation

\[
\Delta G = \Delta G_w - mC_{ur}
\]

where

\[
\Delta G = RT \ln K
\]

\[
C_{ur,m} = \Delta G_w / m
\]

\(\Delta G\) is the Gibbs energy change as a function of temperature T (298K in our case) and urea molar concentration \(C_{ur}\), while \(\Delta G_w\) is the Gibbs energy change in water (kJ/mol).
\( m \) is the proportionality coefficient (in kJ/M mol), and \( R = 8.314 \text{ kJ/K mol} \) is the gas constant. \( C_{ur,m} \) is the denaturant concentration at the transition midpoint, which visually corresponds to the inflection point of the sigmoid curve. The average thermodynamic parameters obtained from the fitting analysis of TUGE data are \( \Delta G_w = 12.6 \pm 0.6 \text{ kJ/mol} \); \( m = 6.2 \pm 0.2 \text{ kJ/M mol} \), and \( C_{ur,m} = 2.0 \pm 0.1 \text{ M} \), which correspond to those for 20 kDa globular proteins [32].

\[
\Delta G = \Delta G_w - m \cdot C_{ur,m}
\]

\[
\Delta G = R \cdot T \ln K
\]

**Figure 2.** Transverse urea gradient electrophoresis of (a) protein components of the cell precipitate dissolved in 20 mM Tris-HCl, pH 8.9, 8.0 M urea, followed by centrifugation (refolding); (b) unfolding of TcpA-C isolated from the soluble fraction after Ni-NTA and Superdex 200 chromatography. The dashed lines show the points used in the evaluation of thermodynamic parameters.

The TUGE data (Figure 2) also hint at a possibility of refolding of TcpA-C from 8.0 M urea without essential aggregation using simple dilution with the native buffer to a final urea concentration below 1.0 M. Figure 3a presents a size-exclusion chromatography profile of TcpA-C purified from inclusion bodies and refolded in the buffer 20 mM Tris, pH 8.9, 150 mM NaCl (see Section 2). As seen, the bulk of the refolded protein elutes like a globular protein with a molecular weight of about 20 kDa (Figure 3c) and an elution volume very
close to that of the native protein isolated from the soluble fraction (Figure 3b), though minor aggregation is present (Figure 3a, arrow).

Figure 3. Size-exclusion chromatography on Superdex 200 (20 mM Tris-HCl, pH 8.9, 150 mM NaCl): (a) TcpA-C purified from inclusion bodies and refolded; the concentration of the loaded protein was 1 mg/mL; an arrow indicates the peak corresponding to aggregates; (b) TcpA-C from the soluble fraction after Ni-NTA chromatography; the concentration of the loaded protein was 0.05 mg/mL; (c) the dependence of the elution time on the protein molecular weight (the chromatographic column calibration); proteins used for the calibration are aldolase (158 kDa), bovine serum albumin (66 kDa), bovine carbonic anhydrase B (29 kDa), and cytochrome c (12.3 kDa), the positions of which are shown by open circles, while the position of monomeric TcpA-C is indicated by a closed circle.

Figure 4 presents some physicochemical properties of the recombinant TcpA-C after its recovery from inclusion bodies (see Section 2). Figure 4a shows SDS-PAGE of inclusion bodies at different stages of purification. Lane 2 in Figure 4a represents washed inclusion bodies (see Section 2); it has one major protein band corresponding to TcpA-C by molecular weight, but the protein purity is not high enough. Further purification of TcpA-C involves Ni-NTA chromatography in 8.0 M urea (Figure 4a, lane 3) and MonoQ chromatography, followed by protein refolding and size-exclusion chromatography (Figure 4a, lane 4). The additional purification by MonoQ chromatography removes some impurities that remained after Ni-NTA chromatography at pH 8.9 within the region of 37 kDa (Figure 4a, lane 2–3). The UV-absorbance and fluorescence spectra (Figure 4b) of TcpA-C recovered from inclusion bodies (Figure 4a, lane 4) are typical for the protein chain with a high content of phenylalanine residues, only a few tyrosines, and no tryptophanes, which is in agreement with the amino acid composition of TcpA-C (Figure S1, [11,17]) and confirms the absence of highly absorbing impurities like nucleotides.

The Far-UV CD spectrum of TcpA-C recovered from inclusion bodies (Figure 4c) is typical for an $\alpha + \beta$ globular protein and is very close to that of TcpA-C purified from the soluble fraction (Figure S4). These data hint at the correct folding of TcpA-C recovered from inclusion bodies. Moreover, we compared the protein secondary structure contents in solution with that in a crystal [16] using several programs of deconvolution (BeStSel web server, CDNN CD deconvolution software 2.1, K2D3 web server, see Section 2). The result is presented in Table S1. As seen, in the crystal and solution, $\alpha$-helix and $\beta$-structure contents differ depending on the program of deconvolution, likely due to the imperfection of deconvolution of this sort of protein secondary structure. Nevertheless, we can conclude that almost half of the protein sequence is involved in the secondary structure both in solution and in the crystal.

The rigid packing of TcpA-C recovered from inclusion bodies was confirmed by differential scanning calorimetry (DSC) and $^1$H-NMR. The former is presented in Figure 5. The DSC curve fitting based on the two-state unfolding model [33] gave a very good result showing the following parameters: the transition midpoint $T_m = 315.9 \pm 0.2$ K; the heat
The effect $\Delta H_m = 339 \pm 10 \text{ kJ/mol}$; the heat capacity increment $\Delta C_{p,m} = 8.5 \pm 0.4 \text{ kJ/K mol}$ at the midpoint; and Gibbs energy change at 298.15 K $\Delta G_{298.15K} = 14.0 \pm 0.7 \text{ kJ/mol}$. The Gibbs energy changes revealed by two different methods (cf. $\Delta G$ values for isothermal unfolding by urea and those of temperature-induced unfolding) are very similar and typical of the middle-size globular protein unfolding [32].

Figure 4. Physico-chemical properties of the recombinant TcpA-C. (a) SDS-PAGE of TcpA-C from inclusion bodies (IB): molecular weight markers (lane 1); washed IB (lane 2); TcpA-C after purification from washed IB by Ni-NTA chromatography in 8.0 M urea (lane 3), followed by MonoQ chromatography and refolding (lane 4); the position of lane 4 originated from another gel was corrected according to the lane of molecular weight markers; (b) Far-UV absorption (solid line) and intrinsic fluorescence (dashed line) spectra of TcpA-C purified and refolded from inclusion bodies (20 mM Tris-HCl, pH 8.9, 150 mM NaCl), and intrinsic fluorescence (dotted line) spectrum of TcpA-C unfolded in 8.0 M urea; (c) Far-UV CD spectra of TcpA-C purified and refolded from inclusion bodies (20 mM Tris-HCl, pH 8.9, 150 mM NaCl) (circles), and of TcpA-C unfolded in 8.0 M urea (squares).

Figure 6 represents the high field region of $^1$H-NMR spectrum of TcpA-C recovered from inclusion bodies. The presence of the high-field signals of methyl protons caused by persistent interactions between methyl groups in aliphatic residues and aromatic side chains is indicative of rigid packing of the protein tertiary structure [34].
The rigid packing of TcpA-C recovered from inclusion bodies was mostly localized to intracellular aggregates in aliphatic residues and aromatic side chains is indicative of rigid packing of the protein tertiary structure [34].

4. Discussion

The main aim of the present work is to obtain the recombinant toxin-coregulated pilus A protein encoded by $tcpA^{CIRS}$ (TcpA) that is a promising tool for the development of new cholera vaccines [7,8]. Regretfully, the recombinant TcpA produced by expression of the appropriate plasmid in $E. \text{coli}$ cells is mostly localized to intracellular aggregates called inclusion bodies [35,36] (Figure 1).

The recovery of recombinant proteins from inclusion bodies is a challenge in their large-scale production by cell biotechnology [37]. The difficulty arises from the absence of...
prediction of the protein ability to refold after denaturant-induced unfolding [38,39]. Here, we used a simple and efficient approach to determining this ability after the solubilization of inclusion bodies in urea-containing buffers. This approach, the so-called TUGE, was previously developed by T. Creighton and D. Goldenberg [26,27] for visualization of unfolding/refolding transitions of globular proteins. TUGE allows tracking changes in mobility of the major protein component of cell lysate precipitate with decreasing urea concentration and requires neither high purity nor large amount of the substance (Figure 2a). Additionally, this method allows defining the urea concentrations for the native and unfolded states of proteins and for possible aggregation of intermediate conformations. Moreover, it provides some thermodynamic parameters of the protein stability and the information on rigidity of the refolded protein tertiary structure. The presence of even a small amount of the recombinant protein in the soluble fraction of *E. coli* lysate (Figure 1) shows that this protein can be folded in cytoplasmic conditions (pH, ionic strength, temperature), and allows comparison of some structural parameters of an in vivo folded protein with those of the protein in vitro recovered from inclusion bodies (Figures 2, 3 and S4). We demonstrate that the renaturation curve of TcpA whose aggregates are dissolved in urea (Figure 2a) and the denaturation curve of this protein isolated from the soluble fraction (Figure 2b) are very similar, and their processing gives almost identical thermodynamic parameters. Furthermore, the S-like change in protein electrophoretic mobility is typical for the urea gradient-induced folding/unfolding transition of a rigid globular protein [26]. The decrease in urea concentration resulting from 20-fold dilution with the native buffer, followed by size-exclusion chromatography, leads to refolding of TcpA into a globular conformation similar in hydrodynamic dimension to that of the protein isolated from the soluble fraction of the cell lysate (Figure 3). Moreover, Far-UV CD spectra of these proteins are very similar to each other (Figure S4), thus indicating the similarity of their secondary structure contents. Although, in our case, the TcpA-C production was hampered by formation of inclusion bodies, the yield of pure protein was evaluated as 60 mg/L of the initial culture. This value is intermediate between previously reported yields of TcpA from other strains, e.g., 8 mg/L [7] and 2.1 mg/mL [8]. The large-scale recovery of TcpA from inclusion bodies allows its characterization by several physico-chemical methods requiring a significant amount of protein. Thus, the circular dichroism spectra (Figure 4c) indicate a high extent of recovery of the protein secondary structure. The presence of a rigid tertiary structure in TcpA recovered from inclusion bodies is clearly demonstrated by the presence of a pronounced peak on the DSC heat absorption curve (Figure 5). Moreover, Gibbs free energy determined by isothermic unfolding/refolding of TcpA (Figure 2), coincides with the value determined by DSC (Figure 5). Note that the rigidity of the tertiary structure of recombinant TcpA-C recovered from inclusion bodies according to the proposed protocol is confirmed by the presence of high field resonances of methyl protons in the 1H-NMR spectrum (Figure 6). Unfortunately, the physico-chemical parameters of the recombinant TcpA-Cs from other strains are underexplored, which prevents comparison of different TcpA-C preparations [7,8,10]. The exception is crystallographic studies of the TcpAET-C globular domain (Figure S1) [16], which give a detailed information about the protein spatial structure. One interesting piece of data is the presence of the single intramolecular disulfide bond between Cys145 and Cys211 (Figure S1) [16]. This feature hints at the possibility of forming S-S bonds in *E. coli* cytoplasm and highlights an important point indicating the correct protein folding. Using Ellman’s reagent [29], we ascertain that our preparation of TcpA-C recovered from inclusion bodies contains an S-S bond. Moreover, the mobility of TcpA-C in SDS-PAGE is visibly reduced after protein treatment with dithiothreitol (DTT) (Figure S5), likely due to an increase in the protein hydrodynamic volume caused by disruption of the disulfide bond.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/life12111802/s1, Figure S1: The sequences of TcpA proteins; Figure S2: Schematic presentation of the expression plasmid pET302 NT-His tcpACIRS; Figure S3: Mass spectrometry analysis of recombinant TcpA-C; Figure S4: Far-UV CD spectra of TcpA-C purified.
and refolded from inclusion bodies and purified from the soluble fraction; Figure S5: SDS-PAGE of TcpA-C purified from inclusion bodies; Table S1: The TcpA-C secondary structure contents.

Author Contributions: Conceptualization, N.K., N.S. and G.S.; Data curation, G.S.; Formal analysis, V.F. and G.S.; Investigation, V.M., E.D., N.K. and A.S.; Methodology, V.M., A.S. and V.F.; Project administration, G.S.; Resources, V.M., E.D. and N.K.; Software, V.M. and V.F.; Supervision, G.S.; Validation, N.K., I.T. and G.S.; Visualization, V.M. and N.M.; Writing—original draft, V.F.; Writing—review and editing, N.M. and G.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to Kutyshenko V.P. (Institute of Theoretical and Experimental Biophysics of Russian Academy of Sciences, Pushchino, Moscow region, Russia) for 1H-NMR measurements, and to Serebrova E.V. (Institute of Protein Research of Russian Academy of Sciences, Pushchino, Moscow region, Russia) for English language editing.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Kaper, J.B.; Morris, J.G.; Levine, M.M. Cholera. Clin. Microbiol. Rev. 1995, 8, 48–86. [CrossRef]
2. Smirnova, N.I.; Goryaev, A.A.; Kutyrev, V.V. The evolution of the Vibrio cholerae genome during the modern period. Mol. Genet. Microbiol. Virol. 2010, 25, 148–157. [CrossRef]
3. Hu, D.; Liu, B.; Feng, L.; Ding, P.; Guo, X.; Wang, M.; Cao, B.; Reeves, P.R.; Wang, L. Origins of the current seventh cholera pandemic. Proc. Natl. Acad. Sci. USA. 2016, 113, E7730–E7739. [CrossRef] [PubMed]
4. Grim, C.J.; Hasan, N.A.; Taviani, E.; Haley, B.; Chun, J.; Brettin, T.S.; Bruce, D.C.; Detter, J.C.; Han, C.S.; Chertkov, O.; et al. Genome sequence of hybrid Vibrio cholerae O1 MJ-1236, B-33, and CIRS101 and comparative genomics with V. cholerae. J. Bacteriol. 2010, 192, 3524–3533. [CrossRef] [PubMed]
5. Bhandari, M.; Jennison, A.V.; Rathnayake, I.U.; Huygens, F. Evolution, distribution and genetics of atypical Vibrio cholerae—A review. Infect. Genet. Evol. 2021, 89, 104726. [CrossRef]
6. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. J. Mol. Biol. 1990, 215, 403–410. [CrossRef]
7. Kiaie, S.; Abtahi, H.; Mosayebi, G.; Alikhani, M.; Pakzad, I. Recombinant Toxin-Coregulated Pilus A (TcpA) as a candidate subunit cholera vaccine. Iran J. Microbiol. 2014, 6, 68–73.
8. Molaei, N.; Mosayebi, G.; Atozande-Nobaveh, A.; Soleymen, M.R.; Abtahi, H. Evolution of the Immune response against recombinant proteins (TcpA, TcpB, and FlaA) as a candidate subunit cholera vaccine. J. Immunol. Res. 2017, 2017, 2412747. [CrossRef]
9. Price, G.A.; Holmes, R.K. Immunizing adult female mice with a TcpA-A2-CTB chimera provides a high level of protection for their pups in the infant mouse model of cholera. PLoS Negl. Trop. Dis. 2014, 8, e3356. [CrossRef] [PubMed]
10. Rollenhagen, J.E.; Kalsy, A.; Cerda, F.; John, M.; Harris, J.B.; LaRocque, R.C.; Qadri, F.; Calderwood, S.B.; Taylor, R.K.; Ryan, E.T. Transcutaneous Immunization with toxin-coregulated pilin a induces protective immunity against Vibrio cholerae O1 El tor challenge in mice. Infect. Immun. 2006, 74, 5834–5839. [CrossRef]
11. Rhine, J.A.; Taylor, R.K. TcpA pilin sequences and colonization requirements for O1 and O139 Vibrio cholerae. Mol. Microbiol. 1994, 13, 1013–1020. [CrossRef]
12. Herrington, D.A.; Hall, R.H.; Losonsky, G.; Mekalanos, J.J.; Taylor, R.K.; Levine, M.M. Toxin, toxin-coregulated pilus, and the toxR regulon are essential for vibrio cholerae pathogenesis in humans. J. Exp. Med. 1988, 168, 1487–1492. [CrossRef] [PubMed]
13. Craig, L.; Pique, M.E.; Tamer, J.A. Type IV pilus structure and bacterial pathogenicity. Nat. Rev. Microbiol. 2004, 2, 363–378. [CrossRef]
14. Krebs, S.J.; Taylor, R.K. Protection and attachment of vibrio cholerae mediated by the toxin-coregulated pilus in the infant mouse model. J. Bacteriol. 2011, 193, 533–540. [CrossRef] [PubMed]
15. Hauke, C.A.; Taylor, R.K. Production of putative enhanced oral cholera vaccine strains that express toxin-coregulated pilus. PLoS ONE 2017, 12, e0175170. [CrossRef]
16. Lim, M.S.; Ng, D.; Zong, Z.; Arvai, A.S.; Taylor, R.K.; Tainer, J.A.; Craig, L. Vibrio cholerae El Tor TcpA crystal structure and mechanism for pilus-mediated microcolony formation. Mol. Microbiol. 2010, 77, 755–770. [CrossRef]
17. Klavdienko, E.V.; Tuchkov, I.V.; Polunina, T.A.; Guseva, N.P.; Smirnova, N.I. Construction of recombinant escherichia coli strain—Producer of basic subunit of Toxin-Coregulated Pilus of Adhesion (TCPA) of Vibrio cholerae biovar El Tor. Probl. Part. Danger. Infect. 2017, 4, 32–37. [CrossRef]

18. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M.R.; Appel, R.D.; Bairoch, A. Protein identification and analysis tools on the ExPASy server. In The Proteomics Protocols Handbook; Walker, J.M., Ed.; Humana Press: Totowa, NJ, USA, 2005; pp. 571–607, ISBN 978-1-59259-890-8.

19. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 1976, 72, 248–254. [CrossRef]

20. Pace, C.N.; Vajdos, F.; Fee, L.; Grimsley, G.; Gray, T. How to measure and predict the molar absorption coefficient of a protein. Protein Sci. 1995, 4, 2411–2423. [CrossRef]

21. Micsonai, A.; Moussong, É.; Wien, F.; Boros, E.; Vadászi, H.; Murvai, N.; Lee, Y.-H.; Molnár, T.; Réfrégiers, M.; Goto, Y.; et al. BeStSel: Webserver for secondary structure and fold prediction for protein CD spectroscopy. Nucleic Acids Res. 2022, 50, W90–W98. [CrossRef]

22. Böhm, G. CDNN: CD Spectra Deconvolution Software, Version 2.1; University of Halle-Wittenberg: Halle, Germany, 1997.

23. Louis-Jeune, C.; Andrade-Navarro, M.A.; Perez-Iratxeta, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. Proteins Struct. Funct. Bioinforma. 2012, 80, 374–381. [CrossRef]

24. Crescenzi, P.; Saneja, A.; Srichandan, S.; Panda, A.K. Bacterial inclusion bodies: A treasure trove of bioactive proteins. Trends Biotechnol. 2020, 38, 474–486. [CrossRef]

25. Singhvi, P.; Upadhyay, V.; Upadhyay, A.K.; Singh, S.M.; Panda, A.K. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. Microb. Cell Fact. 2015, 14, 41. [CrossRef]