



**HAL**  
open science

## Structural ensemble and biological activity of DciA intrinsically disordered region

Maud Chan-Yao-Chong, Stéphanie Marsin, Sophie Quevillon-Cheruel,  
Dominique Durand, Tâp Ha-Duong

► **To cite this version:**

Maud Chan-Yao-Chong, Stéphanie Marsin, Sophie Quevillon-Cheruel, Dominique Durand, Tâp Ha-Duong. Structural ensemble and biological activity of DciA intrinsically disordered region. *Journal of Structural Biology*, 2020, pp.107573. 10.1016/j.jsb.2020.107573 . hal-02903062

**HAL Id: hal-02903062**

**<https://hal.science/hal-02903062>**

Submitted on 22 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Structural ensemble and biological activity of DciA intrinsically disordered region

Maud Chan-Yao-Chong<sup>a</sup>, Stéphanie Marsin<sup>b</sup>, Sophie Quevillon-Cheruel<sup>b</sup>,  
Dominique Durand<sup>b,\*</sup>, Tâp Ha-Duong<sup>a,\*</sup>

<sup>a</sup>Université Paris-Saclay, CNRS, BioCIS, 92290 Châtenay-Malabry, France.

<sup>b</sup>Université Paris-Saclay, CEA, CNRS, Institute for Integrative Biology of the Cell,  
91198 Gif-sur-Yvette, France.

---

## Abstract

DciA is a newly discovered bacterial protein involved in loading the replicative helicase DnaB onto DNA at the initiation step of chromosome replication. Its three-dimensional structure is composed of a folded N-terminal domain (residues 1-111) resembling K Homology domains and a long disordered C-terminal tail (residues 112-157) which structure-activity relationship remains to be elucidated. In the present study on *Vibrio cholerae* DciA, we emphasize the importance of its disordered region to load DnaB onto DNA using surface plasmon resonance (SPR) and isothermal titration microcalorimetry (ITC). Then we characterize the conformational ensemble of the full-length protein using a combination of circular dichroism (CD), small angle X-ray scattering (SAXS), and molecular dynamics (MD) simulations. The atomic-level structural ensemble generated by MD simulations is in very good agreement with SAXS data. From initial conformations of the C-terminal tail without any secondary structure, our simulations bring to light several transient helical structures in this segment, which might be molecular recognition features (MoRFs) for the binding to DnaB and its recruitment and loading onto DNA.

**Keywords:** DNA replication initiation, Helicase DnaB loader, Intrinsically disordered protein, Conformational ensemble, Experimental and computational biophysics.

---

\*Correspondence: [dominique.durand@i2bc.paris-saclay.fr](mailto:dominique.durand@i2bc.paris-saclay.fr) or [tap.ha-duong@u-psud.fr](mailto:tap.ha-duong@u-psud.fr)

## 1. Introduction

In domain Bacteria, the replication of circular chromosomes is a tightly controlled mechanism involving a machinery of several proteins named replisome. Briefly, DNA replication initiation occurs at the replication origin *oriC*, a specific sequence recognized by the initiator protein DnaA. The binding of DnaA to *oriC* leads to the unwinding of DNA double helix and to the loading of the replicative helicase onto single-stranded DNA (ssDNA). Then the helicase is translocated on the lagging strand in the 5' to 3' direction to open the double strand of DNA. The helicase is further associated with both the primase and polymerase to perform DNA replication.

Bacterial helicase is a homo-hexamer of protein DnaB whose the loading on ssDNA is operated by protein DnaC in *Escherichia coli* or DnaI in *Bacillus subtilis*. However, a majority of bacterial species are devoid of proteins DnaC or DnaI. Actually, it was discovered in 2016 that DnaC and DnaI genes were acquired during evolution by domestication in bacteria of a phage gene, to the detriment of an ancestral gene called DciA (Dna[CI] Antecedent) [1, 2]. This gene persists in most bacteria, including many pathogens such as *Vibrio cholerae*, *Yersinia pestis*, *Mycobacterium tuberculosis* or *Pseudomonas aeruginosa*. In the latter, it was demonstrated that DciA directly and specifically interacts with DnaB and that the knock-out of DciA gene induces a blocking of the replication initiation [1, 3].

From a structural point of view, nuclear magnetic resonance (NMR) and SAXS experiments showed that *V. cholerae* DciA has a N-terminal domain which is well folded (residues 1-111) and a C-terminal tail which is presumably intrinsically disordered (segment 112-157 residues) [3]. The N-terminal domain has a fold similar to KH domains known to bind polynucleotides [6], with one long  $\alpha$ -helix (residues 10-38), two  $\beta$ -strands (segments 47-53 and 56-61) followed by a second  $\alpha$ -helix (residues 64-83) and a third  $\beta$ -strand (residues 91-97). Interestingly, the folding of DciA N-terminal domain is different from that of loading proteins DnaC and DnaI (AAA+ domains) but similar to that of the N-terminal domain of the initiator protein DnaA which also interacts with DNA and DnaB [7, 4] (Fig. 1).

Beside their folded domain, all four previous helicase binding proteins (DciA, DnaC, DnaI, and DnaA) have an intrinsically disordered region predicted by PONDR [8]. In *E. coli*, DnaC disordered segment was shown to adopt an helical conformation upon binding to helicase DnaB (PDB structures 6QEL and 6QEM [12]), suggesting that DciA disordered region might



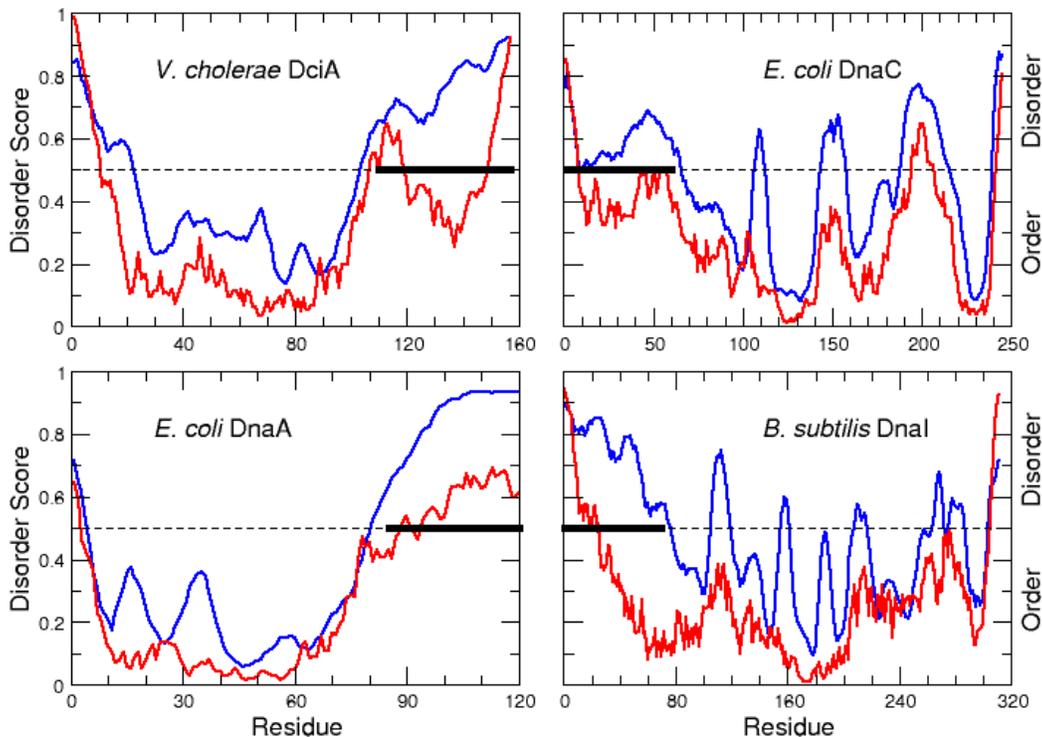


Figure 2: Prediction of disordered segments in *V. cholerae* DciA, *E. coli* DnaA and DnaC, and *B. subtilis* DnaI using web-servers PONDNR [8] and IUPred2A [9]. Blue and red lines indicate scores obtained with the predictors PONDNR VSL2 [10] and IUPred [11], respectively. Black horizontal thick bars represent disordered regions predicted by PONDNR.

region in loading DnaB onto DNA using surface plasmon resonance (SPR) experiments. Then, we characterized the conformational ensemble of the full-length DciA in solution using multiple molecular dynamics (MD) simulations combined with small angle X-ray scattering (SAXS) and circular dichroism (CD) data.

## 2. Material and methods

### 2.1. Bioinformatics prediction of disorder and secondary structures

We used bioinformatics tools to predict the degree of disorder of *V. cholerae* DciA, *E. coli* DnaA and DnaC, and *B. subtilis* DnaI as a function of their sequence. Predictions presented herein were obtained by using web-servers PONDNR [8] and IUPred2A [9]. We also attempted to detect

secondary structures in the C-terminal region of DciA using sequence analyses. Results reported below were provided by web-servers PsiPred [13] and NetSurfP [14].

### 2.2. Protein sample preparation

Experimental procedures for cloning, expression, and purification of proteins studied herein are described in Ref. [3].

### 2.3. Surface plasmon resonance

SPR experiments were performed using a Proteon XPR36 instrument (Bio-Rad). The measurements were done at 30 °C to avoid non-specific binding of DciA in Phosphate Buffered Saline (PBS) with 0.01 % of Tween 20 (PBST), a nonionic detergent to avoid also aspecific liaison, and with or without ATP 0.5 mM. NLC sensor chips were used to immobilize the different DNA through their biotin-tag. For immobilization, DNA was diluted in PBST and attached to the chip in such a way as to obtain 50 Resonance Unit (RU) in different orientations (3' or 5').

In each experiment, proteins were injected in PBST with 0.5 mM ATP at 50  $\mu\text{L}/\text{min}$ , during 240 s, and dissociation was run during 750 s with only PBST and 0.5 mM ATP without proteins. Then proteins were injected a second time during 240 s in PBST with 0.5 mM ATP and a second dissociation was run during 750 s. In all experiments, proteins were injected at the concentration of 1.25  $\mu\text{M}$ . After each interaction test, the chip was regenerated using 0.5 % of SDS. After correction by subtraction of the uncoated reference channel, the sensorgrams were analyzed and compared.

### 2.4. Isothermal titration microcalorimetry

Isothermal titration microcalorimetry experiments were performed with an ITC200 isothermal titration calorimeter from MicroCal (Malvern). The experiments were carried out at 20 °C. *V. cholerae* DnaB wild type concentration in the microcalorimeter cell (0.2 ml) was 33  $\mu\text{M}$ . Nineteen injections of 2  $\mu\text{l}$  of full-length DciA or DciA<sup>[1-111]</sup> at 157  $\mu\text{M}$  were performed at intervals of 180 s while stirring at 500 rpm. The experimental data were fitted to theoretical titration curves with software supplied by MicroCal (ORIGIN®). This software uses the relationship between the heat generated by each injection and  $\Delta\text{H}$  (enthalpy change in kcal/mol),  $K_a$  (the association binding constant in  $\text{M}^{-1}$ ),  $n$  (the number of binding sites), total protein concentration, and free and total ligand concentrations [15].

### 2.5. Small angle X-ray scattering

SAXS experiments were performed on the SWING beamline at the synchrotron SOLEIL, St-Aubin, France. The sample-to-detector (Eiger 4M Decris) distance was set to 1789 mm and the wavelength  $\lambda$  to 1.0 Å, allowing useful data collection over the scattering vector of  $0.005 \text{ \AA}^{-1} < q < 0.5 \text{ \AA}^{-1}$ . In order to avoid the contribution due to aggregates, SAXS data were collected directly after elution of the protein through the online size-exclusion high-performance liquid chromatography column (Superdex increase 200, 5x150) equilibrated in 20 mM Tris-HCl (pH 7.5), 200 mM NaCl. 50  $\mu\text{L}$  of protein sample was injected at 15 °C at an initial concentration close to 2 mg/mL. The protein was eluted at a flow rate of 0.3 mL/min, the frame duration was 0.99 s and the dead time between frames was 0.01 s. Scattering of the elution buffer before void volume was recorded and subtracted from all protein scattering curves. The scattered intensities were displayed on an absolute scale using water scattering.

For each frame, the protein concentration ( $\approx 0.4 \text{ mg/mL}$  at the top of elution peak) was estimated from UV absorption at 280 nm using a spectrometer located immediately upstream of the SAXS measuring cell. Data were first analyzed using Foxtrot, the Swing in-house software, and then using the US-SOMO HPLC module [16]. This program yields for each SAXS frame the value of the scattering intensity  $I(0)$  and of the radius of gyration ( $R_g$ ) by applying the Guinier analysis:  $\text{Log}[I(q)/I(0)] = -q^2 R_g^2/3$  [17]. Identical frames under the main elution peak were selected using Cormap [18] and averaged for further analysis.

In order to compare experimental and theoretical SAXS intensities of a protein, a robust and accurate method is required to back-calculate SAXS curves from its conformational ensemble. It should be reminded that the scattering intensity of a protein, obtained as the scattering curve of the protein solution minus the scattering of the buffer, not only accounts for the protein electrons but also for those of solvent molecules in excess or in deficiency at the protein surface with respect to the buffer electron density. Thus, back-calculations of the intensity scattered by a protein imply to correctly calculate the volume of solvent excluded by the protein and to accurately estimate the scattering contribution from the protein hydration layer [19, 20, 21]. In a previous study, we benchmarked different software using implicit models of solvation to back-calculate the SAXS intensity of a fully intrinsically disorder protein (N-WASP domain V) [22]. For most of these software, including CRY SOL [23] and Pepsi-SAXS [24], we found that their default parameter

for the excess electron density were too large for correctly reproducing SAXS intensities calculated with the explicit-solvent SAXS software WAXSiS [25]. Consistently with other publications [21, 26], we showed that lower values for excess electron density parameters allow to calculate SAXS intensities more accurately [22]. Regarding specifically Pepsi-SAXS, we suggested to reduce its excess electron density parameter from 5% (default value) to 2.2% ( $\Delta\rho = 0.0073 \text{ e}/\text{\AA}^3$ ). Since Pepsi-SAXS is also notably very fast, we chose to use it with the latter parameter for the present study of DciA.

### 2.6. Molecular dynamics simulations

Initial conformations of *V. cholerae* full-length DciA were built on the basis of the NMR structure of the truncated N-terminal domain DciA<sup>[1-111]</sup> [3]. Nevertheless, there were significant uncertainties in NMR data about the presence of two  $\alpha$ -helices at C-terminal positions 98-102 and 104-110. We thus decided to build two sets of initial conformations on the basis of the folded domain DciA<sup>[1-111]</sup> with or without these two  $\alpha$ -helices in region 98-110. Two ensembles of 5 000 conformations of DciA disordered C-terminal tail were generated using the statistical coil generator software Flexible-Meccano (FM) [27], one for segment 112-157, the other one for region 98-157. Then, the NMR structures of DciA folded domain were merged with the FM conformations of the disordered tail to obtain the full-length protein. After removing all structures with steric clashes, we selected 24 conformations with Rg varying from 17 to 40 Å in the pool containing two  $\alpha$ -helices in region 98-110, and 24 other ones with Rg ranging from 19 to 42 Å in the pool without secondary structure in this segment. It should be noted that in all 48 selected conformations, there are no secondary structures in the intrinsically disordered region 112-157. Finally, the complete atomic structures were generated by adding the missing side chains of the disordered segment (generated by FM) using SCWRL4 program [28] and all missing hydrogens with GROMACS tools *pdb2gmx* [29].

Each of the 48 initial conformations of full-length DciA was solvated in a dodecahedral rhombic box of 14.0 nm edge, then neutralized by adding 234 sodium and 243 chloride ions to reach a salt concentration of 200 mM. The non-bonded interactions were treated using the smooth PME method [30] for the electrostatic terms and a cutoff distance of 1.2 nm for the van der Waals potentials. All solute and water covalent bond lengths were kept constant using the LINCS [31] and SETTLE [32] algorithms, respectively, allowing to integrate the equations of motion with a 2 fs time step. All simulations

were performed with the GROMACS software [29] in the NPT ensemble, at  $T = 310$  K and  $P = 1$  bar, using the Nose-Hoover and Parrinello-Rahman algorithms [33, 34, 35] with the time coupling constants  $\tau_T = 0.1$  ps and  $\tau_P = 0.5$  ps. Each of the 48 initial conformations was submitted to 2 ns of equilibration followed by 100 ns of production yielding an accumulated trajectory of 4.8  $\mu$ s.

As reported in various publications [36, 37, 38], several force fields for protein and water model were improved to accurately generate conformational ensembles of IDPs. This is generally achieved by using a four-site water model which better accounts for the polar properties of water, and by accentuating the depth of the solute-solvent Lennard-Jones (LJ) potentials to better solvate non-polar residues. In our previous article [22], we have tested some of these improved force fields, and found that the use of AMBER-03w and TIP4P/2005s proposed by Best *et al.* [39] can satisfactorily reproduce both NMR and SAXS data obtained on N-WASP domain V, a 67-residues intrinsically disordered protein. For this new study, we decided to use the same combination of protein and water models to perform independent MD simulations of DciA.

To minimize the possible bias induced by our selection of initial conformations, we only kept the last 80 ns of each MD simulation for subsequent analyses. We collected data every 40 ps yielding an ensemble of 96 048 structures for DciA. Most of the conformational analyses were performed using GROMACS tools and in-house python scripts. Nevertheless, it should be noted that protein radii of gyration were not calculated with the GROMACS tool *gmx gyrate*, but from the back-calculated SAXS intensities using the Guinier approximation  $\text{Log}[I(q)/I(0)] = -q^2 Rg^2/3$  when  $q \rightarrow 0$  [17]. Secondary structure propensity scores were computed by using SSP program [40], which combines chemical shifts of  $C\alpha$ ,  $C\beta$ , CO, and backbone N atoms calculated by the program SHIFTS [41]. The software STRIDE [42] was also used to assign secondary structure elements to each residue of each protein conformation, based on hydrogen bond criteria and backbone dihedral angle values.

### 2.7. Conformational sub-ensemble selection

From a large conformational ensemble, it is possible to derive a sub-ensemble that better agrees with experimental data, if available [43]. In this study, we applied the program GAJOE from the suite EOM [44, 45] to the DciA conformational ensemble generated by MD in order to select a sub-ensemble that better fit the SAXS intensities. More specifically, we

asked GAJOE to perform 100 runs of genetic algorithm optimization of the fit with experiments, and to save a subset of 50 optimal structures at the end of each run, yielding a final sub-ensemble of 5 000 selected conformations.

### 2.8. Circular Dichroism

The circular dichroism (CD) spectra of wild type DciA were measured using JASCO J-810 spectrometer. Samples of DciA were prepared at pH 7.5, in buffer including 20 mM of Tris-HCl and 50 mM sodium chloride. The initial concentration of protein was 1.39 g/L and 0.60 g/L for full-length DciA and truncated DciA<sup>[1-111]</sup>, respectively. Samples were recorded from 260-190 nm wavelengths for each scan. CD spectra were further analyzed using the web-server DICHROWEB [46] to estimate the secondary structures content of proteins. The reference data set 7 was used since it is optimized for 190-240 nm wavelengths and includes CD spectra of denatured/disordered proteins [47]. The three analysis algorithms SELCON [48], CONTIN [49], and CDSSTR [50] were employed in conjunction as recommended for a reliable analysis [51]. All three analysis programs yield percentages of residues in  $\alpha$ -helix and  $\beta$ -strand conformations, from which the number of residues in both secondary structures can be calculated. It should be noted that, in order to estimate standard deviations of these numbers, we proceeded as following: we considered that uncertainties on these numbers arise from errors in spectra intensities which fitting can be improved by applying a scaling factor [46]. Thus, for each used algorithm, we performed five CD spectra analyses with five scaling factors equal to 1.2, 1.1, 1.0, 0.9, and 0.8. Then, the output percentages and numbers of residues in  $\alpha$ -helix or  $\beta$ -strand are averaged and standard deviations can be calculated.

## 3. Results

### 3.1. DciA disordered region is important for helicase loading onto ssDNA

To investigate the role of DciA disordered segment 112-157 in loading DnaB onto DNA, we compared here SPR experiments of DnaB binding to ssDNA in presence of truncated DciA<sup>[1-111]</sup> or full-length protein. SPR results first show that binding of DnaB onto ssDNA without any DciA (pink curve of Fig. 3) or in the presence of DciA<sup>[1-111]</sup> (cyan curve) are quite similar, indicating that truncated DciA<sup>[1-111]</sup> does not particularly contribute to the loading of additional DnaB on ssDNA. In contrast, as shown by the significant increase from 100-200 to 500-600 RU of the grey, blue, and red curves

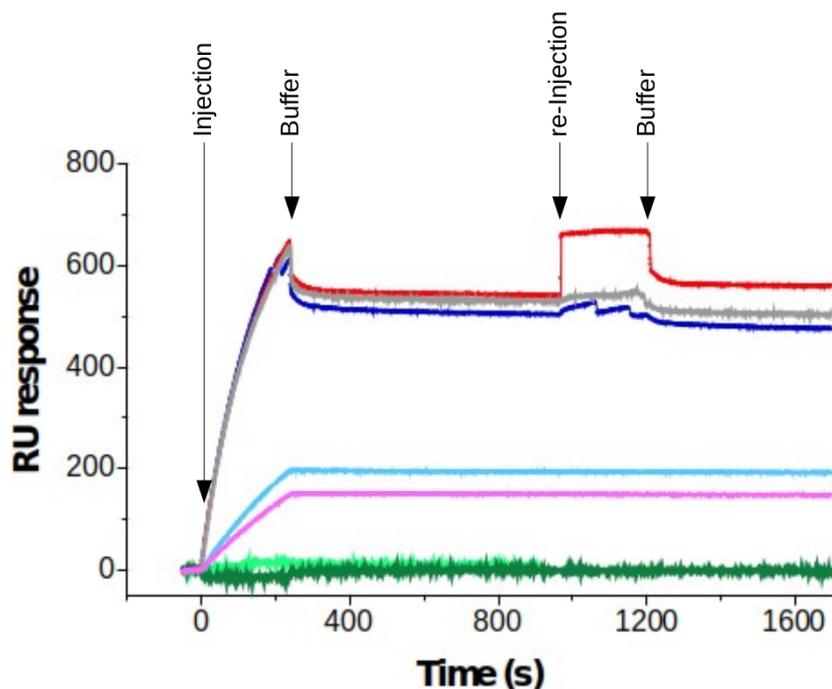


Figure 3: SPR experiments of DnaB binding to DNA, in presence of DciA<sup>[1-111]</sup> or full-length DciA. The single-strand DNA is attached to the surface by its 5' extremity. In each experiment, a first injection is made with buffer and different proteins and a buffer passage is used to clean up the excess protein (with only PBST and ATP). Then a second injection is made and a second buffer passage is used to clean up the excess protein. The results of different injections of proteins are reported here: DnaB alone and nothing for the second passage (pink curve); DnaB and DciA<sup>[1-111]</sup> then nothing for the second passage (cyan curve); DnaB with full-length DciA then nothing for the second passage (grey curve); DnaB with DciA then with DciA<sup>[1-111]</sup> for the second passage (blue curve); DnaB with DciA then with DciA for the second passage (red curve). Control experiments of DciA<sup>[1-111]</sup> (light green) and full-length DciA (dark green) binding to DNA are also reported (with a short time of buffer passage for truncated DciA).

of Fig. 3, full-length DciA clearly stimulates and increases the binding to ss-DNA of proteins which are probably complexes of DciA-DnaB. After these increases, a biphasic dissociation can be observed. First, there is a rapid signal decrease by about 100 RU in less than 100 s which would originate from the release of DciA while DnaB still hangs on DNA. Then, a slow decrease of SPR signals can be observed, which would be associated to the slow dissociation of DnaB from attached DNA (Fig. 3).

To confirm this scenario, we made a second injection of these proteins

15 min after the first injections (re-Injection arrow in Fig. 3), similarly to the work by Ioannou *et al.* on *B. subtilis* DnaI and DnaB [52]. It is observed that DciA, but not DciA<sup>[1-111]</sup>, can induce an increase in the SPR signal (red curve of Fig. 3), indicating that full-length DciA can rebind and reform a ternary complex DciA-DnaB-ssDNA. It should be noted that the signal increase amplitude (about 100 RU) is comparable to that of the rapid decrease observed after the first injection, confirming that the latter was associated to a rapid dissociation of DciA from DnaB-ssDNA. Altogether, these SPR experiments showed the importance of DciA disordered region in the loading of DnaB helicase on DNA. They suggest that DciA C-terminal tail is involved in forming the ternary complex DciA-DnaB-ssDNA.

These results were complemented by ITC experiments which evidenced an absence of interaction between DnaB and DciA<sup>[1-111]</sup> but a direct binding of full-length DciA to the helicase with an estimated  $K_d = 1.1 \pm 0.2 \mu\text{M}$  (Fig. 4). Altogether, these experimental data demonstrated that the C-terminal disordered region of DciA make direct interactions with DnaB. We thus further focused our attention on DciA conformational ensemble to gain insight into the biological role of its C-terminal tail. In particular, we attempt to detect transient secondary structures within its intrinsically disordered region which could be important for binding to DnaB. For this purpose, we generated a conformational ensemble of full-length DciA using multiple MD simulations. This ensemble was validated by using SAXS and CD experiments and further analyzed to identify possible molecular recognition features (MoRFs).

### 3.2. DciA global shape and size

Experimental SAXS intensity  $I(q)$  of full-length DciA is plotted in Fig. 5. Using the Guinier analysis of these data, the protein average radius of gyration is estimated to 26.9 Å, which is much larger than the expected value for a globular protein of 157 residues (16-17 Å). The dimensionless Kratky plot of SAXS data has a bell shape at low scattering angles followed by a continuously increasing curve for larger angles, demonstrating that DciA has a folded domain and a disordered tail [53]. Furthermore, the top of the bell-shaped profile has a position significantly offset from the peak position expected for globular proteins ( $\sqrt{3}, 1.10$ ) [53], confirming that DciA contains a compact domain and an extended region.

In order to infer from previous SAXS data a conformational ensemble of DciA, we performed multiple MD simulations of the protein starting from 48 various initial conformations. From the last 80 ns of each simulation, we

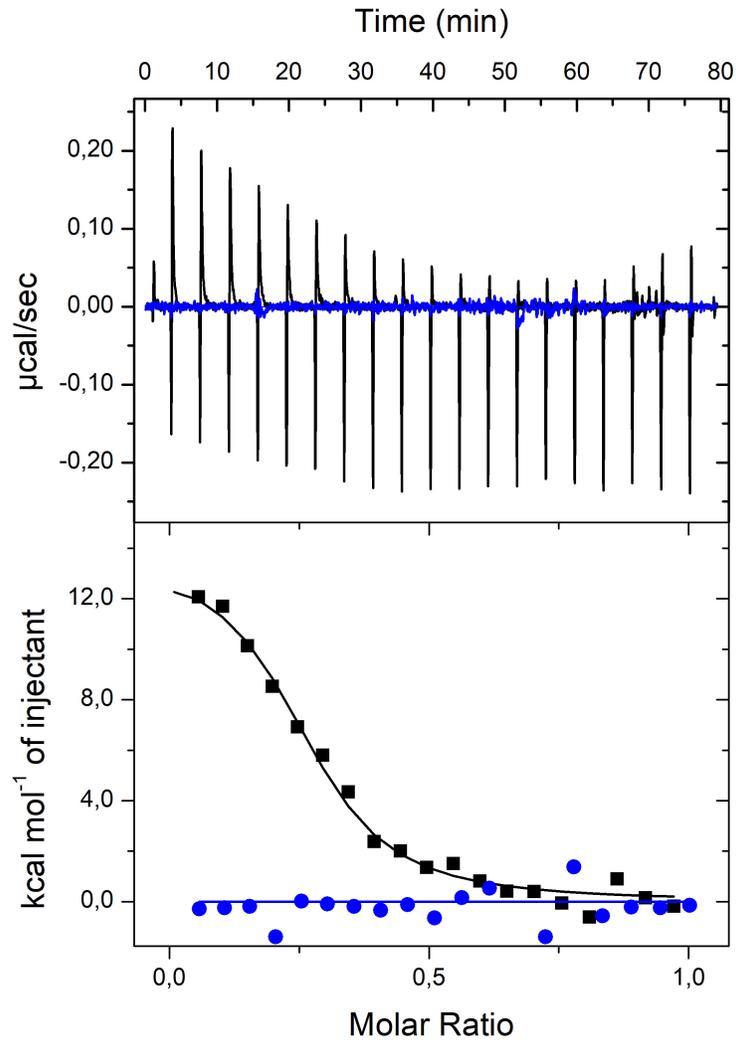


Figure 4: ITC experiments: Interaction measurements of *V. cholerae* DnaB wild type with full-length DciA (black) and truncated DciA<sup>[1-111]</sup> (blue). Top panel shows heat differences upon injection of DciA into DnaB and low panel shows integrated heats of injection with the best fit (solid line) to a single binding model using Microcal ORIGIN.

built an ensemble of 96 048 different structures, which the 10 most populated clusters representative structures are shown in Fig. 6. The average SAXS intensity calculated from this conformational ensemble is directly compared to experimental data in Fig. 5. The theoretical and experimental scattering curves are in fairly good agreement, the  $\chi^2$  value between the two sets of

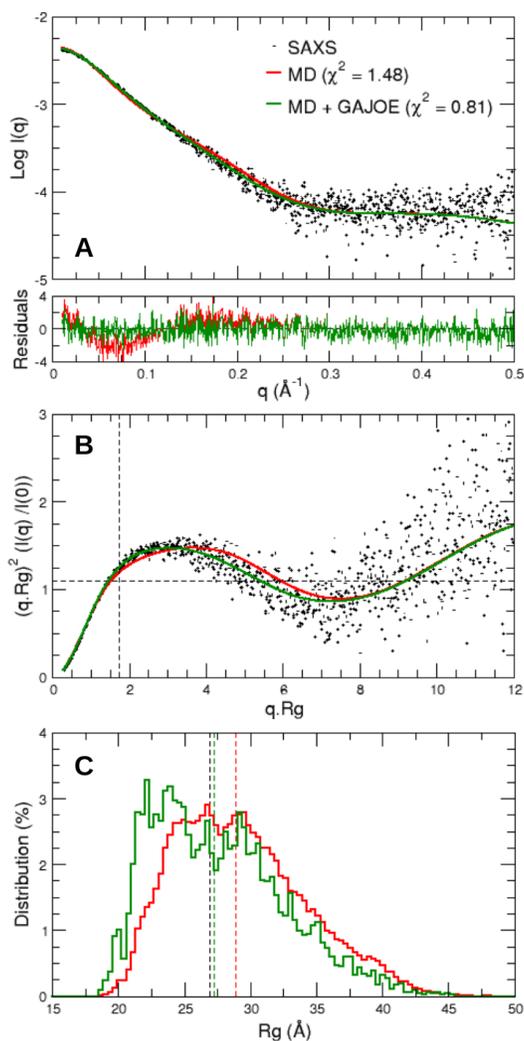


Figure 5: (A) Comparison of DciA SAXS experimental data (black dots) with back-calculated intensities from conformational ensemble generated by MD (red line) and from sub-ensemble selected by GAJOE (green line). Reduced residuals  $\Delta/\sigma = [I_{calc}(q) - I_{exp}(q)]/\sigma_{exp}(q)$  are shown just below. (B) Dimensionless Kratky plots. The point where the two dashed lines cross indicates the position  $(\sqrt{3}, 1.10)$  of the top of bell-shaped profiles typical of globular proteins. (C) Distribution of DciA radius of gyration computed over conformational ensemble generated by MD (red line) and over sub-ensemble selected by GAJOE (green line). Vertical dashed lines indicate average radius of gyration from experiments and simulations.

data being slightly larger than 1. Nevertheless, as seen in the residual plot, the agreement is not perfect for small angles, and in the Kratky plot, the top of the calculated bell-shaped curve seems to be slightly shifted toward higher angles when compared to experimental data, indicating that extended conformations are slightly more frequent in MD simulations than in experiments. Indeed, the average radius of gyration computed over DciA conformational ensemble generated by MD has a value of 28.9 Å slightly higher than the experimental one (26.9 Å) (Fig. 5).

In order to find a DciA conformational sub-ensemble that better fit the SAXS data, we applied GAJOE to the MD-derived ensemble. This generated a sub-ensemble of 5 000 conformations whose the back-calculated SAXS intensity is now in very good agreement with experiments (Fig. 5). The distribution of radius of gyration computed over this sub-ensemble is distorted toward lower values when compared to the MD ensemble, and its average radius of gyration (27.2 Å) is now very close to the experimental value (Fig. 5). Altogether, the conformational sub-ensemble selected by GAJOE is a good representation of DciA in solution, at least at the global level.

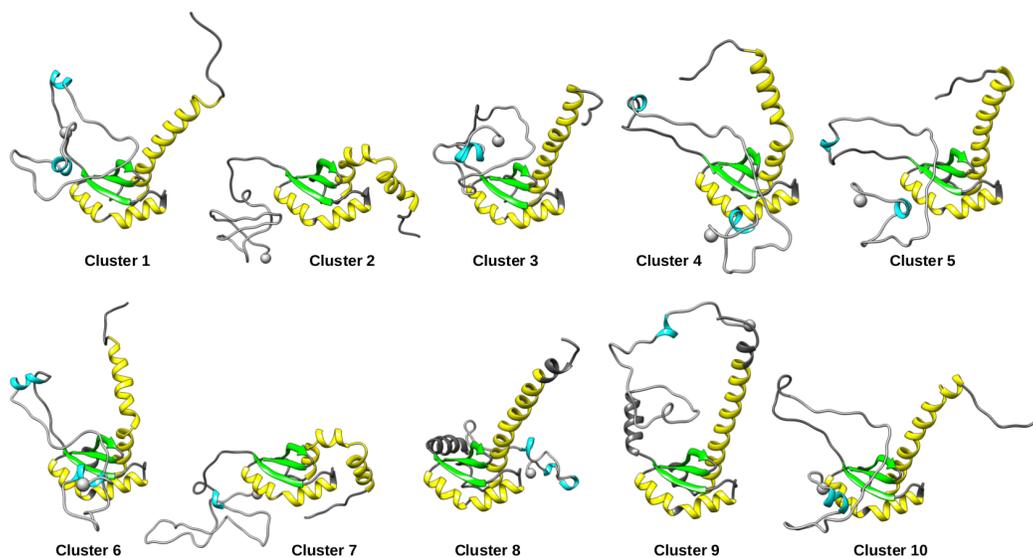


Figure 6: Representative structures of the 10 most populated clusters of the DciA conformational ensemble generated by MD simulations.

### 3.3. DciA secondary structures

Secondary structure content of DciA disordered region can be estimated by comparing circular dichroism (CD) spectra of truncated DciA<sup>[1-111]</sup> and full-length DciA<sup>[1-157]</sup>. Nevertheless, it should be mentioned that CD data presented here were obtained on His-tagged proteins. Both spectra of 6xHis-DciA<sup>[1-111]</sup> and 6xHis-DciA<sup>[1-157]</sup> are typical of  $\alpha$ -helix-rich proteins with a maximum absorption band close to 190 nm and two minimum  $\Delta\epsilon$  around 210 and 220 nm (Fig. 7). Regarding full-length DciA, the first minimum is slightly shifted toward low wavelengths and deeper than the second one, consistently with an increase in random coil residues relative to truncated protein 6xHis-DciA<sup>[1-111]</sup>.

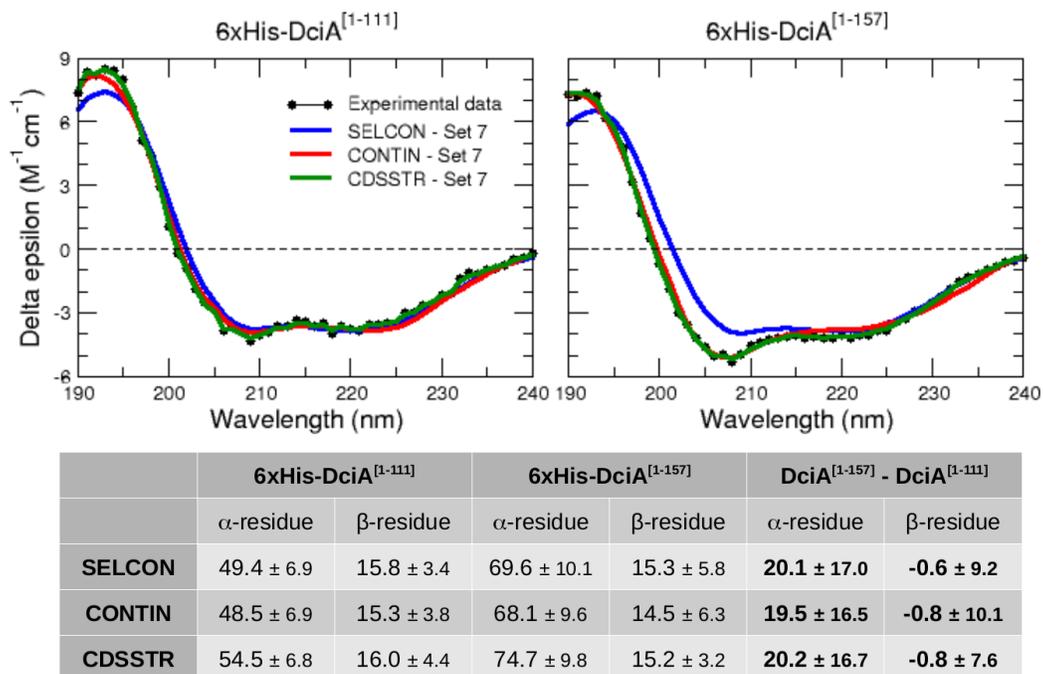


Figure 7: Top: Experimental and DICHROWEB CD spectra of His-tagged truncated DciA<sup>[1-111]</sup> and full-length DciA<sup>[1-157]</sup>. Bottom: Number of residues in helical or extended conformation for each protein as provided by DICHROWEB. The two last columns indicate differences in  $\alpha$ -residues and  $\beta$ -residues between 6xHis-DciA<sup>[1-157]</sup> and 6xHis-DciA<sup>[1-111]</sup>.

Quantitative analyses of DciA secondary structures can be inferred from CD spectra by using the web-server DICHROWEB which performs deconvolutions of experimental data into basis spectra of known proteins [46].

We used here the reference data set 7 optimized for 190-240 nm wavelengths which includes CD spectra of denatured/disordered proteins [47] and is therefore the most suitable set to study DciA. Reconstructed spectra output from DICHROWEB are shown and compared to experimental data in Fig. 7. SELCON provides the less accurate fits with NRMSD parameters [54] equal to 0.130 and 0.249 for 6xHis-DciA<sup>[1-111]</sup> and 6xHis-DciA<sup>[1-157]</sup>, respectively. CONTIN yields NRMSD values of 0.073 and 0.068, and CDSSTR produces the best fits with NRMSD equal to 0.021 and 0.0020 for truncated and full-length DciA, respectively.

From CD spectra decomposition of DciA<sup>[1-111]</sup>, the numbers of residues in helical and strand conformations are estimated by all three analysis algorithms in the ranges of 49-55 and 15-16, respectively. These numbers can help to remove the uncertainty about the presence or not of two  $\alpha$ -helices at positions 98-102 and 104-110 in the NMR structure of truncated DciA<sup>[1-111]</sup>. Indeed, in presence of these two helices, the number of  $\alpha$ -residues in truncated DciA would be 60 instead of 48 without them (Fig. 1). Thus, CD data seem to suggest that DciA<sup>[1-111]</sup> might partially lack these two  $\alpha$ -helices in region 98-110.

Despite differences in method and accuracy for fitting experimental CD data, all three analysis programs SELCON, CONTIN, and CDSSTR reveal very similar observations about variations in secondary structure numbers of full-length DciA with respect to truncated one (Fig. 7). First, no significant variation in the number of  $\beta$ -residues was detected. Secondly, the number of  $\alpha$ -residues increases of about 19-20 in 6xHis-DciA<sup>[1-157]</sup> relative to 6xHis-DciA<sup>[1-111]</sup>. This strongly suggests that DciA disordered C-terminal region 98-157 has significant propensity to form  $\alpha$ -helical structures. However, it should be noted that uncertainties on additional numbers of  $\alpha$ -residues remain very large due to the low resolution of CD technique, indicating that these numbers should be viewed with precaution. Moreover, CD cannot say where are located the  $\alpha$ -residues in the disordered tail of full-length DciA and how transient they are. To address these questions, we examine the propensity of the DciA residues to adopt transient helical structures in the protein conformational ensemble generated by MD.

First, we computed the residue-specific secondary structure propensity score (SSP) [40] from  $C\alpha$ ,  $C\beta$ , CO, and backbone N chemical shifts, either measured by NMR experiments (for DciA<sup>[1-111]</sup>) or calculated using the program SHIFTS [41] over MD conformational ensemble. The latter calculations were already used in a previous study on the N-WASP disordered domain

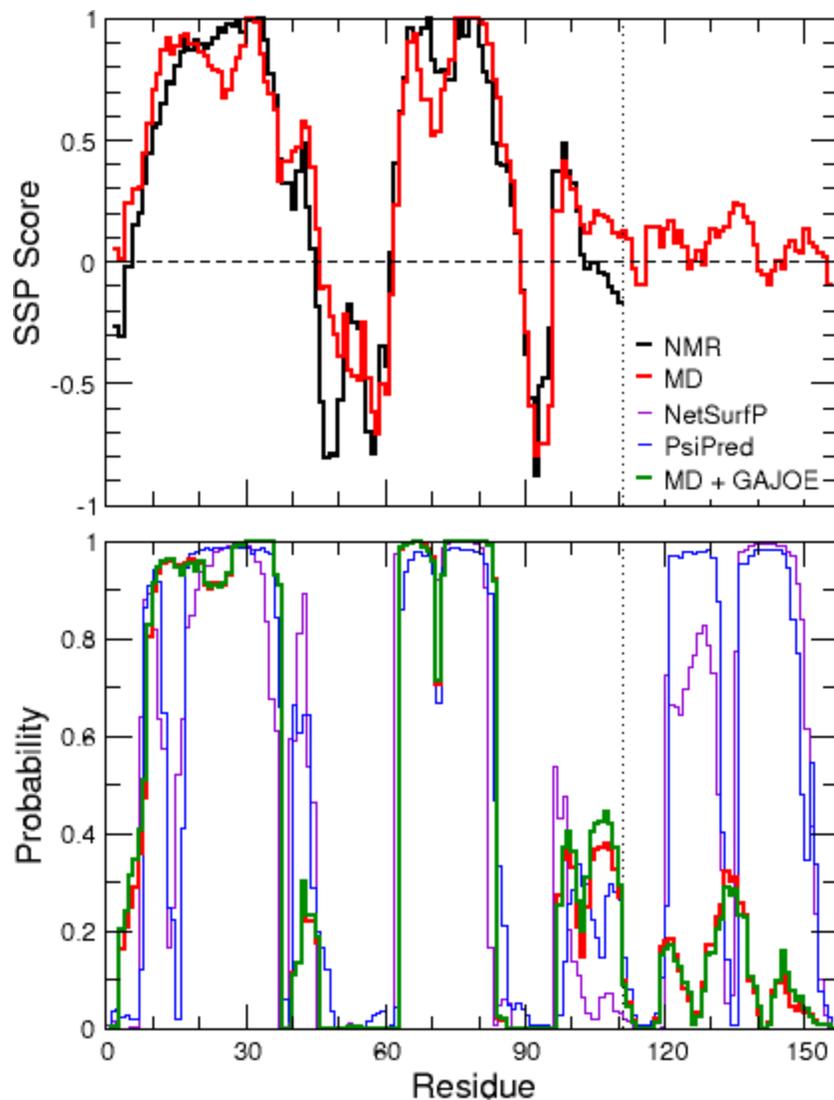


Figure 8: Top: Propensity of DciA residues to form secondary structure estimated with SSP scores from NMR experiments on truncated DciA (black) or averaged over MD ensemble of full-length protein (red). Bottom: Probability of DciA residues to form  $\alpha$ -helix computed from MD ensemble (red), with NetSurfP [14] (violet), PsiPred [13] (blue), or from sub-ensemble selected by GAJOE (green).

V and were able to satisfactorily retrieve the SSP scores from NMR measurements [22]. For a given residue, positive and negative values of the SSP score indicate the proportion of helix and strand structures in the protein

conformational ensemble, respectively. As shown in Fig. 8, both measurements and calculations provide similar SSP scores for residues of DciA folded N-terminal domain. Furthermore, propensities of region 98-102 to form an  $\alpha$ -helix are similar in NMR experiments of truncated DciA<sup>[1-111]</sup> and in MD ensemble of full-length DciA. However, region 104-110 has different behaviors in truncated and full-length proteins: it is rather unstructured in DciA<sup>[1-111]</sup>, whereas it has a propensity to form an  $\alpha$ -helix in full-length protein. Regarding DciA region 112-157, for which no NMR information was available, it is interesting to note that, starting from entirely random coil conformations, simulations generated non negligible  $\alpha$ -helical structures in this disordered C-terminal tail.

In order to closely locate these secondary structures, we directly computed the residue probabilities to be in  $\alpha$ -helix using the program STRIDE [42] which assigns secondary structure elements on the basis of hydrogen bond criteria and backbone dihedral angle values. Since no experimental information was available for DciA disordered region, we also estimated these probabilities from the protein primary sequence using the predictors PsiPred [13] and NetSurfP [14]. Overall, both bioinformatics and biophysics approaches yielded similar  $\alpha$ -helix probabilities for the folded N-terminal domain 1-111 (Fig. 8). In the intermediate segment 98-111, MD simulations and PsiPred revealed two transient short helices whereas NetSurfP only detects a single one.

However, regarding the disordered C-terminal tail, bioinformatics tools predict two stable  $\alpha$ -helices at regions 122-134 and 137-148, whereas physics-based simulations generated three transient helical segments at positions 119-124, 130-138, and 144-148. This discrepancy is not fully understood and is still under investigation, but it is worthy to note that the length of the two helical segments predicted by bioinformatics methods is in agreement with CD spectra analyses which indicate that full-length DciA has about 20 additional  $\alpha$ -residues with respect to truncated DciA<sup>[1-111]</sup>. However these two stable helices are conflicting with disorder predictors PONDR and IUPRed2A as well as SAXS data which strongly support that DciA C-terminal tail is intrinsically disordered. On the other hand, the three transient helices generated by MD simulations are in line with this disorder but the number of additional  $\alpha$ -helical residues relative to truncated DciA is less consistent with CD analyses. Altogether, among the five techniques used to characterized the structural ensemble of DciA, two of them (CD and secondary structure predictors) seem to indicate that its C-terminal region is structured into two

stable  $\alpha$ -helices. The three other approaches (disorder predictors, SAXS, and MD simulations) tend to prove that this region is mainly disordered with transient  $\alpha$ -helical structures.

Interestingly, in the sub-ensemble of conformations selected by GAJOE, the residue probabilities to be  $\alpha$ -helical are similar to those calculated from the ensemble generated by MD. Notably in the C-terminal disordered region, apart from residues 104-110 which have a slightly greater probability to be helical in the selected sub-ensemble than in the MD ensemble, the three transient helical segments at positions 119-124, 130-138, and 144-148 are retrieved to the same extent in the GAJOE sub-ensemble which also satisfactorily describes the global shape and size of DciA in solution.

#### 4. Discussion and conclusions

DciA is a recently discovered protein ancestral to the DnaC in *E. coli* and DnaI in *B. subtilis* which were acquired during evolution by horizontal gene transfer from phages [2]. All three proteins are involved in the loading of helicase DnaB onto bacterial ssDNA to initiate chromosome replication. However, DciA has none sequence and structure similarities with DnaC and DnaI and little is known about the relationship between DciA structural ensemble and its biological activity as DnaB loader. NMR experiments revealed that *V. cholerae* DciA N-terminal domain (residues 1-111) is folded like KH domains which are known to bind polynucleotides [6]. This N-terminal domain is also very similar to the folded domain of the initiation factor protein DnaA in *E. coli* (Fig. 1). Altogether, this suggests that DciA folded N-terminal domain is involved in binding bacterial DNA.

We report here results of SPR experiments which show that *V. cholerae* DciA C-terminal region (residues 112-157) is important for loading helicase DnaB onto ssDNA (Fig. 3). Complementary ITC experiments evidenced a direct binding of DnaB to full-length DciA but not to truncated DciA<sup>[1-111]</sup> (Fig. 4), indicating that DciA C-terminal region is crucially involved in binding helicase DnaB. In a parallel study, colleagues from the Institute for Integrative Biology of the Cell (I2BC) succeeded in crystallizing DnaB in complex with DciA. The crystal structure reveals three DciA folded N-terminal domains at the surface of the DnaB hexamer, but also six  $\alpha$ -hairpins which sequence was not unambiguously attributed but which probably belong to the C-terminal tail of DciA [3].

On the other hand, SAXS experiments reported here clearly demonstrated that DciA C-terminal tail is intrinsically disordered in the unbound state (Fig. 5), consistently with disorder predictions of PONDER and IUPRed2A (Fig. 2). Multiple MD simulations of the protein allowed to further characterize not only the ensemble average extension of the disordered tail in fair agreement with SAXS data (Fig. 5), but also the secondary structures that transiently appear in DciA C-terminal segment (Fig. 8). We notably identified three transient short  $\alpha$ -helices in region 112-157 but which do not necessarily appear simultaneously (Fig. 6). The length and persistence of these helices are not fully consistent with CD spectra analyses (Fig. 7). However, CD remains a low resolution technique and the secondary structure assignment can significantly vary with errors in spectra intensities, reference sets of spectra, and analysis programs, especially for disordered proteins.

More intriguingly, MD-derived probabilities of  $\alpha$ -helices in DciA disordered tail are much lower than those estimated by bioinformatics tools such as PsiPred [13] or NetSurfP [14]. These latter predicted two stable  $\alpha$ -helices which might correspond to the  $\alpha$ -hairpin observed in the crystallographic DnaB-DciA complex [3]. Thus, discrepancies between MD probabilities and bioinformatics analyses might come from the fact that these latter rely on training data sets of protein crystallographic structures, including protein-protein complexes, and might not be suitable for detecting transient secondary structures in intrinsically disordered regions of unbound proteins.

Nevertheless, these discrepancies give some clues or hypotheses about the mechanism of DnaB recruitment by DciA: First, DnaB is caught by DciA C-terminal tail, possibly through the recognition of one of its short transient  $\alpha$ -helices or  $\alpha$ -MoRFs. Then, following an "induced fit" mechanism upon binding, DciA disordered region folds into an  $\alpha$ -hairpin at the surface of DnaB, and the latter is brought closer to DciA N-terminal domain as observed in the crystallographic structure of DnaB-DciA complex.

## Acknowledgements

We are grateful to Magali Aumont-Niçaise for her expertise in isothermal titration microcalorimetry (ITC) and to H el ene Walbott for fruitful discussions about DnaB-DciA complex crystallographic structure. This work has benefited from the I2BC Macromolecular interactions measurements Platform supported by the French Infrastructure for Integrated Structural Biology (FRISBI) (grant ANR-10-INSB-05-01). We thank the staff of the Swing

beamline at the SOLEIL synchrotron for assistance during SAXS experiments. MD simulations were performed using HPC resources from GENCI-CINES (grant A0040710415). This research was supported by the “IDI 2016” project funded by the IDEX Paris-Saclay (grant ANR-11-IDEX-0003-02).

## References

- [1] P. Brézellec, I. Vallet-Gely, C. Possoz, S. Quevillon-Cheruel, J.-L. Ferat, DciA is an ancestral replicative helicase operator essential for bacterial replication initiation, *Nat Commun* 7 (2016) 13271.
- [2] P. Brézellec, M.-A. Petit, S. Pasek, I. Vallet-Gely, C. Possoz, J.-L. Ferat, Domestication of Lambda Phage Genes into a Putative Third Type of Replicative Helicase Matchmaker, *Genome Biol Evol* 9 (2017) 1561–1566.
- [3] S. Marsin, A. Yazid, J. Andreani, S. Bacconnais, P. Legrand, I. Gallay-Li de la Sierra, C. Cargemel, H. Walbott, C. Possoz, A. Humbert, M. Aumont-Niçaise, C. Velours, F. Oschenbein, D. Durand, E. Le Cam, S. Quevillon-Cheruel, J.-L. Ferat, Dcia-chaperones helicases self-load on DNA, Under submission (2020).
- [4] Y. Abe, T. Jo, Y. Matsuda, C. Matsunaga, T. Katayama, T. Ueda, Structure and Function of DnaA N-terminal Domains, *J. Biol. Chem.* 282 (2007) 17816–17827.
- [5] P. P. Datta, D. N. Wilson, M. Kawazoe, N. K. Swami, T. Kaminishi, M. R. Sharma, T. M. Booth, C. Takemoto, P. Fucini, S. Yokoyama, R. K. Agrawal, Structural Aspects of RbfA Action during Small Ribosomal Subunit Assembly, *Molecular Cell* 28 (2007) 434–445.
- [6] R. Valverde, L. Edwards, L. Regan, Structure and function of KH domains: Structure and function of KH domains, *FEBS Journal* 275 (2008) 2712–2726.
- [7] H. Seitz, C. Weigel, W. Messer, The interaction domains of the DnaA and DnaB replication proteins of *Escherichia coli*, *Molecular Microbiology* 37 (2000) 1270–1279.

- [8] n. Romero, n. Obradovic, n. Dunker, Sequence Data Analysis for Long Disordered Regions Prediction in the Calcineurin Family, *Genome Informatics. Workshop on Genome Informatics* 8 (1997) 110–124.
- [9] B. Mészáros, G. Erdős, Z. Dosztányi, IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding, *Nucleic Acids Research* 46 (2018) W329–W337.
- [10] K. Peng, S. Vucetic, P. Radivojac, C. J. Brown, A. K. Dunker, Z. Obradovic, Optimizing long intrinsic disorder predictors with protein evolutionary information, *J Bioinform Comput Biol* 3 (2005) 35–60.
- [11] Z. Dosztányi, V. Csizmók, P. Tompa, I. Simon, The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins, *Journal of Molecular Biology* 347 (2005) 827–839.
- [12] E. Arias-Palomo, N. Puri, V. L. O’Shea Murray, Q. Yan, J. M. Berger, Physical Basis for the Loading of a Bacterial Replicative Helicase onto DNA, *Molecular Cell* 74 (2019) 1–12.
- [13] L. J. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server, *Bioinformatics* 16 (2000) 404–405.
- [14] M. S. Klausen, M. C. Jespersen, H. Nielsen, K. K. Jensen, V. I. Jurtz, C. K. Sønderby, M. O. A. Sommer, O. Winther, M. Nielsen, B. Petersen, P. Marcatili, NetSurfP-2.0: Improved prediction of protein structural features by integrated deep learning, *Proteins: Structure, Function, and Bioinformatics* 87 (2019) 520–527.
- [15] T. Wiseman, S. Williston, J. F. Brandts, L.-N. Lin, Rapid measurement of binding constants and heats of binding using a new titration calorimeter, *Analytical Biochemistry* 179 (1989) 131–137.
- [16] E. Brookes, P. Vachette, M. Rocco, J. Pérez, US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data, *J Appl Cryst* 49 (2016) 1827–1841.

- [17] A. Guinier, La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques, *Ann. Phys.* 11 (1939) 161–237.
- [18] D. Franke, C. M. Jeffries, D. I. Svergun, Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra, *Nat. Methods* 12 (2015) 419–422.
- [19] J. Trewella, A. Duff, D. Durand, F. Gabel, J. Guss, W. Hendrickson, G. Hura, D. Jacques, N. Kirby, A. Kwan, J. Pérez, L. Pollack, T. Ryan, A. Sali, D. Schneidman-Duhovny, T. Schwede, D. Svergun, M. Sugiyama, J. Tainer, P. Vachette, J. Westbrook, A. Whitten, 2017 publication guidelines for structural modelling of small-Angle scattering data from biomolecules in solution: an update, *Acta Crystallogr. D Struct. Biol.* 73 (2017) 710–728.
- [20] J. Hub, Interpreting solution X-ray scattering data using molecular simulations, *Curr. Opin. Struct. Biol.* 49 (2018) 18–26.
- [21] J. Henriques, L. Arleth, K. Lindorff-Larsen, M. Skepö, On the Calculation of SAXS Profiles of Folded and Intrinsically Disordered Proteins from Computer Simulations, *J. Mol. Biol.* 430 (2018) 2521–2539.
- [22] M. Chan-Yao-Chong, C. Deville, L. Pinet, C. van Heijenoort, D. Durand, T. Ha-Duong, Structural Characterization of N-WASP Domain V Using MD Simulations with NMR and SAXS Data, *Biophysical Journal* 116 (2019) 1216–1227.
- [23] D. Franke, M. V. Petoukhov, P. V. Konarev, A. Panjkovich, A. Tuukkanen, H. D. T. Mertens, A. G. Kikhney, N. R. Hajizadeh, J. M. Franklin, C. M. Jeffries, D. I. Svergun, ATSAS 2.8 : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions, *Journal of Applied Crystallography* 50 (2017) 1212–1225.
- [24] S. Grudinin, M. Garkavenko, A. Kazennov, Pepsi-SAXS : an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles, *Acta Crystallographica Section D Structural Biology* 73 (2017) 449–464.

- [25] C. J. Knight, J. S. Hub, WAXSiS: a web server for the calculation of SAXS/WAXS curves based on explicit-solvent molecular dynamics, *Nucleic Acids Research* 43 (2015) W225–W230.
- [26] F. Persson, P. Söderhjelm, B. Halle, The geometry of protein hydration, *Journal of Chemical Physics* 148 (2018) 215101.
- [27] V. Ozenne, F. Bauer, L. Salmon, J. Huang, M. Jensen, S. Segard, P. Bernadó, C. Charavay, M. Blackledge, Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables, *Bioinformatics* 28 (2012) 1463–1470.
- [28] G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack, Improved prediction of protein side-chain conformations with SCWRL4, *Proteins* 77 (2009) 778–795.
- [29] M. Abraham, T. Murtola, R. Schulz, S. Pall, J. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1-2 (2015) 19–25.
- [30] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, A smooth particle mesh Ewald method, *Journal of chemical physics* 103 (1995) 8577–8593.
- [31] B. Hess, P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation, *Journal of Chemical Theory and Computation* 4 (2008) 116–122.
- [32] S. Miyamoto, P. A. Kollman, SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models, *Journal of computational chemistry* 13 (1992) 952–962.
- [33] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, *Journal of Chemical Physics* 81 (1984) 511–519.
- [34] W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, *Physical Review A* 31 (1985) 1695–1697.

- [35] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *Journal of Applied Physics* 52 (1981) 7182–7190.
- [36] M. Carballo-Pacheco, B. Strodel, Comparison of force fields for Alzheimer’s A $\beta$ 42: A case study for intrinsically disordered proteins, *Protein Science* 26 (2016) 174–185.
- [37] V. Duong, Z. Chen, M. Thapa, R. Luo, Computational Studies of Intrinsically Disordered Proteins, *J. Phys. Chem. B* 122 (2018) 10455–10469.
- [38] M. Chan-Yao-Chong, D. Durand, T. Ha-Duong, Molecular Dynamics Simulations Combined with Nuclear Magnetic Resonance and/or Small-Angle X-ray Scattering Data for Characterizing Intrinsically Disordered Protein Conformational Ensembles, *J. Chem. Inf. Model.* 59 (2019) 1743–1758.
- [39] R. Best, W. Zheng, J. Mittal, Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association, *J. Chem. Theory Comput.* 10 (2014) 5113–5124.
- [40] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, Sensitivity of secondary structure propensities to sequence differences between  $\alpha$ - and  $\gamma$ -synuclein: Implications for fibrillation, *Protein Sci* 15 (2006) 2795–2804.
- [41] X. Xu, D. Case, Automated prediction of  $^{15}\text{N}$ ,  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$  and  $^{13}\text{C}\epsilon$  chemical shifts in proteins using a density functional database, *J. Biomol. NMR* 21 (2001) 321–333.
- [42] M. Heinig, D. Frishman, STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins, *Nucleic Acids Res* 32 (2004) W500–W502.
- [43] M. Bonomi, G. T. Heller, C. Camilloni, M. Vendruscolo, Principles of protein structural ensemble determination, *Current Opinion in Structural Biology* 42 (2017) 106–116.
- [44] P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, D. I. Svergun, Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering, *Journal of the American Chemical Society* 129 (2007) 5656–5664.

- [45] G. Tria, H. D. T. Mertens, M. Kachala, D. I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering, *IUCrJ* 2 (2015) 207–217.
- [46] L. Whitmore, B. A. Wallace, Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases, *Biopolymers* 89 (2008) 392–400.
- [47] N. Sreerama, S. Y. Venyaminov, R. W. Woody, Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Inclusion of Denatured Proteins with Native Proteins in the Analysis, *Analytical Biochemistry* 287 (2000) 243–251.
- [48] N. Sreerama, S. Y. Venyaminov, R. W. Woody, Estimation of the number of  $\alpha$ -helical and  $\beta$ -strand segments in proteins using circular dichroism spectroscopy, *Protein Science* 8 (1999) 370–380.
- [49] I. H. van Stokkum, H. J. Spoelder, M. Bloemendal, R. van Grondelle, F. C. Groen, Estimation of protein secondary structure and error analysis from circular dichroism spectra, *Analytical Biochemistry* 191 (1990) 110–118.
- [50] W. C. Johnson, Analyzing protein circular dichroism spectra for accurate secondary structures, *Proteins: Structure, Function, and Bioinformatics* 35 (1999) 307–312.
- [51] N. Sreerama, R. W. Woody, Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set, *Analytical Biochemistry* 287 (2000) 252–260.
- [52] C. Ioannou, P. M. Schaeffer, N. E. Dixon, P. Sultanas, Helicase binding to DnaI exposes a cryptic DNA-binding site during helicase loading in *Bacillus subtilis*, *Nucleic Acids Res* 34 (2006) 5247–5258.
- [53] V. Receveur-Brechot, D. Durand, How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective, *Curr. Prot. Pept. Sci.* (2012) 55–75.
- [54] D. Mao, E. Wachter, B. A. Wallace, Folding of the mitochondrial proton adenosine triphosphatase proteolipid channel in phospholipid vesicles, *Biochemistry* 21 (1982) 4960–4968.