Biophysical Chemistry 288 (2022) 106843

Contents lists available at ScienceDirect

Biophysical Chemistry

journal homepage: www.elsevier.com/locate/biophyschem

Conformational ensemble of the full-length SARS-CoV-2 nucleocapsid (N) protein based on molecular simulations and SAXS data

Bartosz Różycki [a,*], Evzen Boura [b,*]

[a] Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland
[b] Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, v.v.i, Flemingovo nam. 2, 166 10, Prague 6, Czech Republic

ARTICLE INFO

Keywords:
SAXS
Nucleocapsid
SARS-CoV-2
EROS

ABSTRACT

The nucleocapsid protein of the SARS-CoV-2 virus comprises two RNA-binding domains and three regions that are intrinsically disordered. While the structures of the RNA-binding domains have been solved using protein crystallography and NMR, current knowledge of the conformations of the full-length nucleocapsid protein is rather limited. To fill in this knowledge gap, we combined coarse-grained molecular simulations with data from small-angle X-ray scattering (SAXS) experiments using the ensemble refinement of SAXS (EROS) method. Our results show that the dimer of the full-length nucleocapsid protein exhibits large conformational fluctuations with its radius of gyration ranging from about 4 to 8 nm. The RNA-binding domains do not make direct contacts. The disordered region that links these two domains comprises a hydrophobic α-helix which makes frequent and nonspecific contacts with the RNA-binding domains. Each of the intrinsically disordered regions adopts conformations that are locally compact, yet on average, much more extended than Gaussian chains of equivalent lengths. We offer a detailed picture of the conformational ensemble of the nucleocapsid protein dimer under near-physiological conditions, which will be important for understanding the nucleocapsid assembly process.

1. Introduction

Recent advances in cryo-EM microscopy have allowed us to structurally characterize very large proteins and their complexes [1,2], and structure prediction by AlphaFold has reached previously unseen accuracy [3]. Together with classical macromolecular crystallography and protein NMR, these advances could create an impression that most – if not all – protein structures can be rather easily solved or accurately predicted. Yet one class of proteins is outstanding in this respect, the intrinsically disordered proteins (IDPs) – especially proteins where well-folded domains are connected by long, flexible, intrinsically disordered polypeptide segments. Such proteins are usually too big for NMR analysis and too flexible for crystallographic or cryo-EM analysis [4].

SARS-CoV-2 encodes for more than 20 proteins including 16 non-structural proteins (nsp1–16), spike glycoprotein (S), nucleocapsid protein (N), membrane protein (M), envelope protein (E) and several accessory proteins [5]. Most of these are well folded and were structurally characterized using cryo-EM or macromolecular crystallography. However, the N protein is composed of two globular domains and three intrinsically disordered regions (IDR1–3, Fig. 1). The structured domains are named NTD (N-terminal domain) and CTD (C-terminal domain). The NTD has a right hand-like fold composed of a β-sheet core with a large protruding central loop that resembles a finger (Fig. 1) [6]. Importantly, it binds both single-stranded and double-stranded RNA [7]. The CTD also binds RNA with a preference for viral genomic intergenic transcriptional regulatory sequences [8]. It is composed of eight helices and two β-strands and its shape reminds of the letter C (Fig. 1) [8,9]. Two C-shaped monomers form a dimer which is responsible for the dimerization of the N protein [8–11]. As outlined above, SARS-CoV-2 nucleocapsid NTD and CTD are well structurally characterized. However, the current knowledge about conformations of the full-length N protein is more limited. Importantly, small angle X-ray scattering (SAXS) experiments have shown that the full-length N protein in solution is a dimer with a radius of gyration of 5.9 nm [12].

SAXS is a powerful method to characterize macromolecules in solution [13,14]. Despite the loss of structural information due to orientational averaging of the scattering signal, SAXS provides useful information about the shapes and dimensions of macromolecules in solution. In particular, standard analysis of SAXS data permits the determination of low-resolution structures of proteins in the form of molecular envelopes, where protein shapes are represented by a set of dummy atoms [15]. However, this approach is only valid for proteins

* Corresponding authors.
E-mail addresses: rozycki@ifpan.edu.pl (B. Różycki), boura@uochb.cas.cz (E. Boura).

https://doi.org/10.1016/j.bpc.2022.106843
Received 8 April 2022; Received in revised form 10 May 2022; Accepted 2 June 2022
Available online 7 June 2022

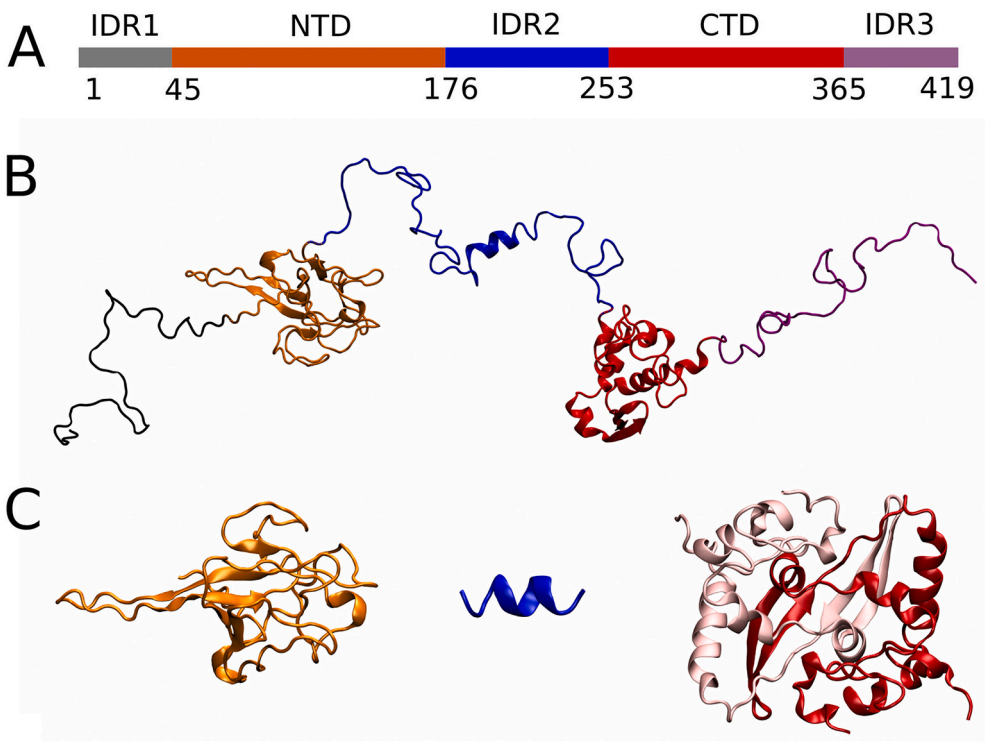

Fig. 1. Domain architecture of the N protein. A) Schematic representation of the N protein domain architecture. The boundaries of the globular domains and the IDRs are indicated. B) A random conformation of a monomeric N protein, which is composed of two folded domains (NTD shown in orange and CTD shown in red) and three intrinsically disordered regions (IDR1 in gray, IDR2 in blue, and IDR3 in purple). C) Structural constrains used in our simulations; from left: the NTD (PDB entry 6YI3), a short α-helix within the IDR2 (PDB entry 7PKU), and a dimer of the CTDs (PDB entry 7DE1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with a stable ternary structure that can be represented as a rigid body, which the N protein is certainly not. SAXS analysis of IDPs requires different approaches that are based on generating an ensemble of protein conformations [16–22]. Useful and appropriate tools for generating conformational ensembles of biomolecules are molecular dynamics (MD) simulations. However, despite steady developments in this field [23–25], all-atom MD simulations of multi-domain and partially disordered proteins are still challenging – not only because of the large sizes of such proteins but also because of the long time scales on which large conformational fluctuations occur. Coarse-grained molecular simulations, on the other hand, provide means to efficiently sample conformational ensembles of large, multi-domain and dynamic proteins and their complexes [4,20,26]. However, the predictive power of coarse-grained simulations certainly lags behind that of all-atom MD simulations. One way to overcome these challenges is to use experimental data to guide or bias coarse-grained simulations.

Recently, original data are more and more reused within the scientific community and many studies are published as open-access articles. Usually the data are deposited in open-access databases such as PDB (Protein Data Bank) or EMDB (Electron Microscopy Data Bank). SASBDB (Small Angle Scattering Biological Data Bank) and PED (Protein Ensemble Database) [27,28] are also available but not yet broadly established and their use is not mandatory (unlike PDB) by most journals. Nevertheless, professors Tengchuan Jin and Hongliang He kindly provided us with the SAXS data they recently acquired on the SARS-CoV-2 N protein [12]. Here, we combined coarse-grained simulations of the N protein dimer with the data from SAXS experiments using the ensemble refinement of SAXS (EROS) method [18]. Our results provide a detailed picture of the conformational ensemble of the SARS-CoV-2 N protein dimer.

2. Materials and methods

2.1. Coarse-grained simulations

We used the coarse-grained model for multi-domain IDPs introduced by Kim and Hummer [29]. Replica exchange Monte Carlo (REMC) simulations of the coarse-grained model were performed with replicas at 24 different temperatures T_i below and above the room temperature $T = 300$ K. Specifically, we used $T/T_i = 0.45, 0.49, 0.52, 0.55, \ldots, 1.09, 1.12, 1.15$. Following the simulation methods of Kim and Hummer [29], the basic MC steps were rotations and translations of each of the rigid domains. For the flexible linkers and termini, in addition to local MC moves on each of the residue beads, crank-shaft moves were employed to enhance sampling. The REMC simulation consisted of 1.1×10^8 MC cycles, where the initial 10^7 MC cycles were used for equilibration and the subsequent 10^8 MC cycles for data collection. The protein conformations were saved every 5000 MC cycles at room temperature, $T_i = T$, which resulted in an ensemble of $N = 20{,}000$ conformations for further analysis. Selected conformations were visualized using VMD [30].

2.2. Ensemble refinement of SAXS

We computed scattering intensity profiles of the recorded conformations using an algorithm co-developed with the EROS method [18] taking the default value of the electron density of the protein hydration shell (0.03 e/Å^3). Then the ensemble-averaged scattering intensity profile was computed as

$$I(q) = \sum_{k=1}^{N} w_k I_k(q)$$

where the index $k = 1, \ldots, N$ labels the conformations in the ensemble, w_k is the statistical weight assigned to conformation k, $I_k(q)$ denotes the scattering intensity profile of the k-th conformation, and q denotes the momentum transfer. (We use the notation $q = 4\pi\sin(\theta/2)/\lambda$, where θ is the scattering angle and λ is the X-ray wavelength). The discrepancy between the experimental SAXS data, $I_{\mathrm{exp}}(q)$, and the ensemble-averaged scattering intensity profile, $I_{\mathrm{sim}}(q)$, was quantified by

$$x^2 = \sum_{i=1}^{N_q} \frac{\left(I_{exp}(q_i) - aI_{sim}(q_i) - b\right)^2}{\sigma^2(q_i)}$$

where $i = 1, \ldots, N_q$ labels points in the SAXS dataset, q_i are the values of

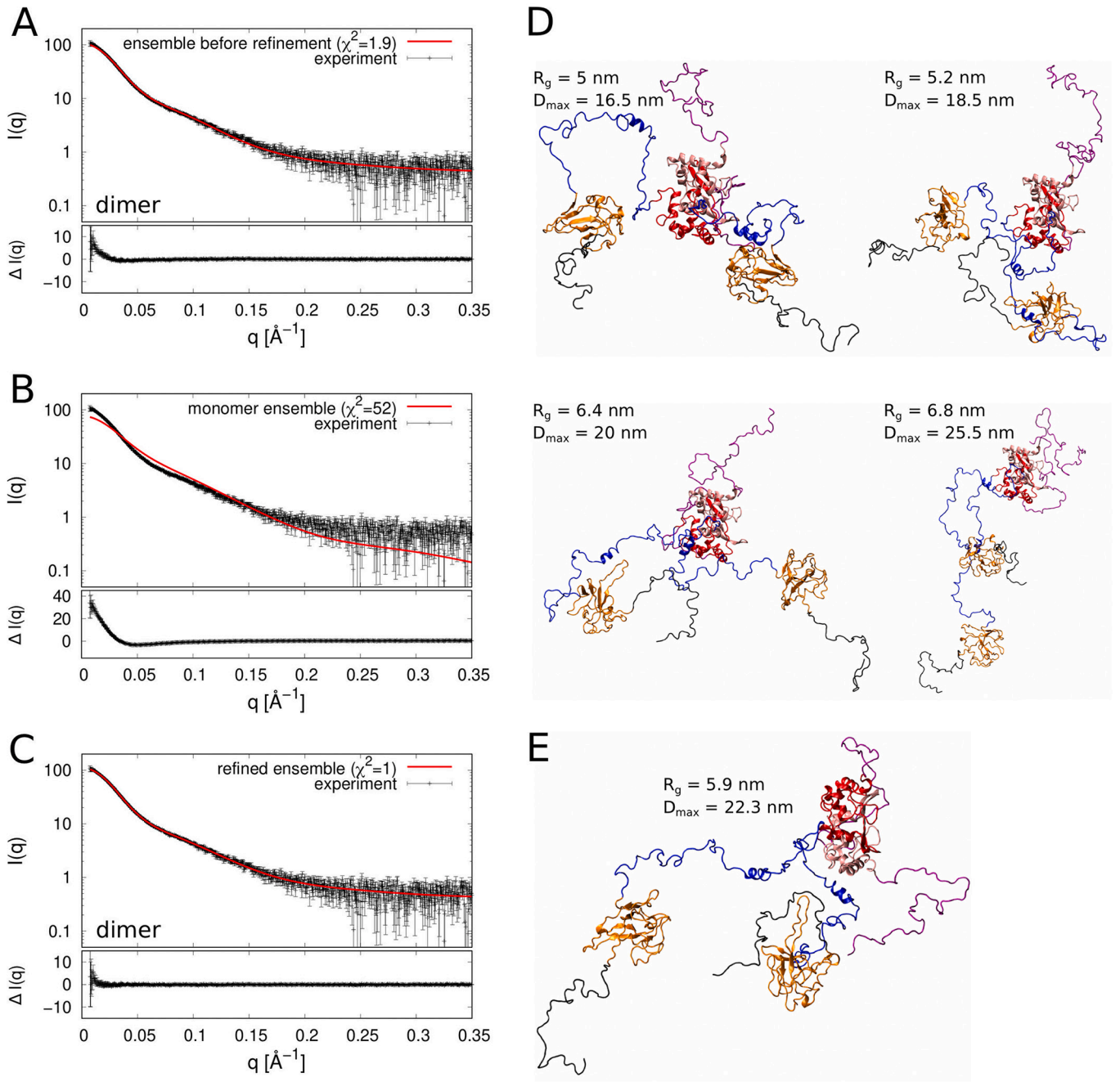

Fig. 2. SAXS of the N protein. A) Comparison of the experimental SAXS data taken from [12] (black points with error bars) with the scattering intensity profile computed from the ensemble of the simulation structures of the N protein dimer (red line). B) Comparison of the experimental SAXS data with the scattering intensity profile computed from the ensemble of the simulation structures of the N protein monomer. C) Comparison of the experimental SAXS data with the scattering intensity profile of the refined ensemble of the dimer structures. D) Cartoon representation of four structures which jointly fit the experimental SAXS data with $\chi^2 = 1$. Their R_g and D_{max} values are indicated. E) Cartoon representation of a dimer structure that best fits the experimental SAXS data ($\chi^2 = 1.12$). The colour code is as in Fig. 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

momentum transfer in the SAXS dataset, and $\sigma(q_i)$ is the statistical error of intensity $I_{exp}(q_i)$. The scale factor a and the offset b result from the conditions $\partial\chi^2/\partial a = 0$ and $\partial\chi^2/\partial b = 0$. The offset parameter b accounts for uncertainties in the buffer subtraction procedures.

At first, all conformations in the simulation ensemble were taken to be equally relevant, which implied $w_k = w_k^{(0)} = 1/N$ for $k = 1, \ldots, N$ and resulted in $\chi^2 = 1.9$ (Fig. 2A). The agreement between the simulation ensemble and the experimental SAXS data was thus good but still could be improved. We therefore refined the conformational ensemble using the EROS method [18], which prevents over-fitting and keeps the refined ensemble (with weights w_k) as close as possible to the reference ensemble (with weights $w_k^{(0)} = 1/N$). Following the original EROS method, we introduced a function $F = \chi^2 - \theta S$, where θ is a parameter that expresses the confidence in the reference ensemble whereas

$$S = -\sum_{k=1}^{N} w_k \, In\left(\frac{w_k}{w_k^{(0)}}\right)$$

is the relative entropy, which equals to the negative Kullback-Leiber divergence. We minimized F with respect to the set of weights w_k using a steepest descent method within the log-weights formulation

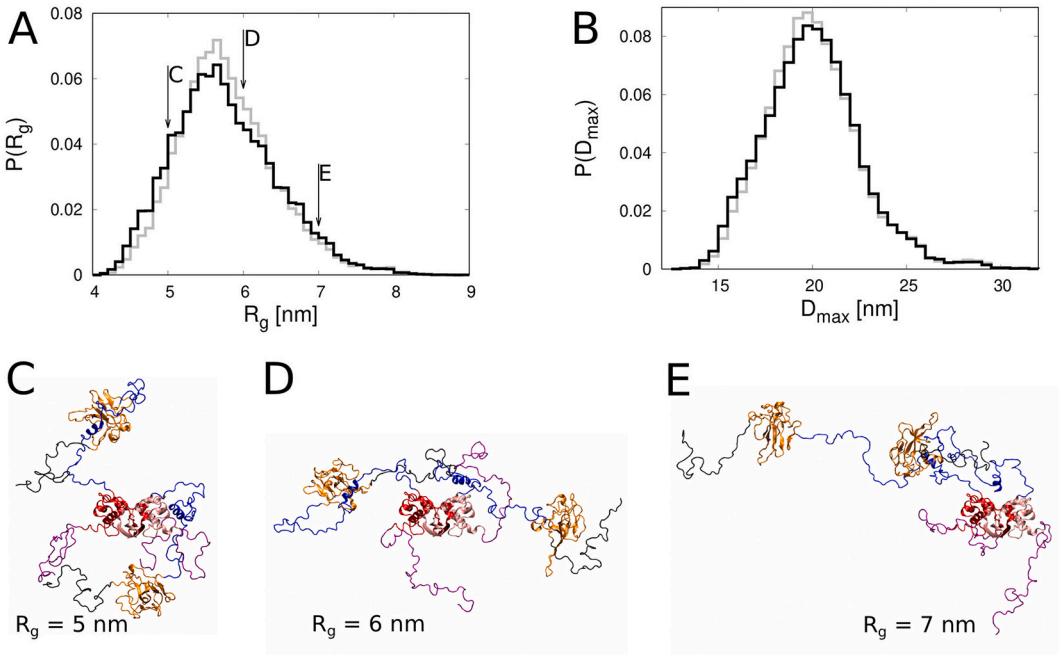

Fig. 3. Analysis of the radius of gyration (R_g) and maximum dimension (D_{max}) of the N protein dimer. Histograms of A) R_g and B) D_{max}. The lines in gray and black correspond to the ensemble of simulation structures before and after refinement, respectively. C-E) Snapshots of selected conformations with C) $R_g = 5$ nm D) $R_g = 6$ nm and E) $R_g = 7$ nm, as indicated by the arrows in panel A. The colour code is as in Fig. 1.

[31,32]. Fig. S1 shows χ^2 as a function of S at $F = F_{min}$ for different values of θ. For large values of θ, when the relative entropy term dominates over χ^2, minimizing F leads to small perturbations in the statistical weights and, thus, $w_k \approx w_k^{(0)}$ for the majority of conformations k. Decreasing the value of θ, the value of χ^2 decreases and approaches the least-χ^2 fit value, under the constraints of positive and normalized weights w_k. Minimization of F at $\theta = 0$ produces the best possible agreement with experiment but also largest changes in the statistical weights, possibly as a result of over-fitting. In the L-shaped curve shown in Fig. S1 we chose a point in the elbow region, as indicated by the gray horizontal line, where the refined ensemble agrees very well with the experimental data ($\chi^2 = 1$, Fig. 2B) without undue over-fitting. The optimal weights at this value of θ are shown in Fig. S2.

2.3. Contacts between residues

Each residue bead was assigned a van der Waals radius as introduced by Kim and Hummer [29]. We assumed that a pair of beads, i and j, with van der Waals radii σ_i and σ_j, respectively, were in contact if their distance r_{ij} was smaller than $1.5 \times \sigma_{ij}$, where $\sigma_{ij} = (\sigma_i + \sigma_j)/2$ [33]. According to this criterion, we calculated a map of residue contacts for each conformation obtained in the REMC simulations. The relative weight of a given contact was then taken as a sum of statistical weights w_k of the conformations in which this contact was present.

3. Results

To efficiently sample conformations of the nucleocapsid protein in solution, we used an implicit-solvent, coarse-grained model introduced by Kim and Hummer [29]. This model is equipped with a transferable energy function and devised to simulate conformational ensembles of large, dynamic, multi-domain proteins. It has been successfully applied to systems ranging from membrane-protein trafficking machineries [34,35] to lipid kinases in dynamic complexes with regulatory proteins [36,37] to hijacking of host protein by viruses [38,39] and to protein complexes controlling the biogenesis of autophagosomes [40]. In the framework of this model, amino acid residues are represented as spherical beads centered at the α-carbon atoms. The interactions between the residue beads are described by amino-acid dependent pair potentials and Debye-Hückel-type electrostatics. Folded protein domains are treated as rigid bodies whereas inter-domain linkers and flexible termini (IDR1–3, Fig. 1) are represented as polymers of amino acid beads with bending, stretching and torsion potentials. Here, the rigid domains were the NTD (PDB entry 6YI3) [7], the dimer of the CTD (PDB entry 7DE1) [8] and the hydrophobic α-helix within the polypeptide segment joining the NTD and CTD (PDB entry 7PKU) [41] (Fig. 1).

We performed extensive replica exchange Monte Carlo simulations of the coarse-grained model (see Materials and methods). From the simulations we obtained an ensemble of $N = 20{,}000$ structures of the N protein dimer at room temperature ($T = 300$ K). We computed the scattering intensity profile for each of the simulation structures using an algorithm co-developed with the EROS method [18]. The ensemble-averaged scattering intensity profile was found to be in good agreement ($\chi^2 = 1.9$) with the SAXS data published by Zeng et al. [12] (Fig. 2A). We also performed analogous simulations of the monomeric form of the N protein (Fig. 1B), and computed the scattering intensity profile corresponding to the ensemble of the simulation structures of the N protein monomer (Fig. 2B). This scattering intensity profile is distinctly inconsistent with the experimental SAXS data ($\chi^2 = 52$), which confirms that monomers of the N protein are very unlikely to exist in solution at the concentrations used in the SAXS experiments. (To our knowledge, the N protein has never been reported to be in the monomeric form). This result illustrates how SAXS and computer simulations can be used together to determine the multimeric state of a macromolecule.

Next, we refined the conformational ensemble of the N protein dimer using the EROS approach, which employs the maximum-entropy method to minimally modify the statistical weights of the simulation structures in an attempt to match the conformational ensemble to the experimental SAXS data (see Materials and methods). The refined ensemble was very close to the original ensemble and in perfect agreement with the SAXS data ($\chi^2 = 1$, Fig. 2C). We also employed the minimum ensemble method [34] and picked out four structures of the N

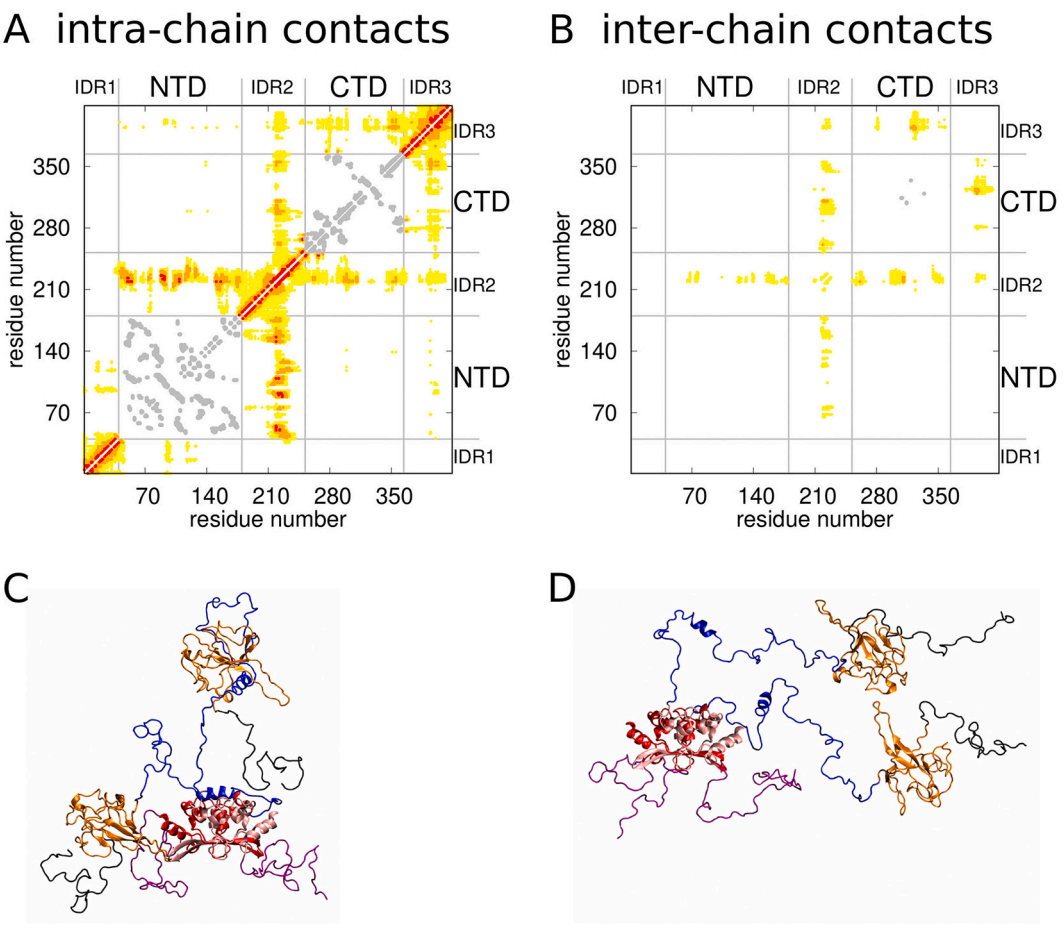

Fig. 4. Analysis of intra- and inter-chain contacts. A, B) Maps of intra- and inter-chain contacts. The points in red, orange and yellow represent frequent, transient and rare contacts, respectively. The points in gray indicate contacts within the folded domains (NTD and CTD) which do not change in the course of the simulations. C) Cartoon representation of a simulation structure with the largest number of contacts. The hydrophobic α-helices (blue) are bound to the NTD (orange) and CTD (red). The IDRs adopt compact conformations. D) Snapshot of a simulation structure with the smallest number of contacts. The hydrophobic α-helices are unbound and the IDRs are rather extended. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

protein dimer that jointly fit the SAXS data perfectly well ($\chi^2 = 1$, Fig. 2D). This set of structures gives us a glimpse into conformational fluctuations of the N protein dimer. We also identified a single structure that best fits the SAXS data ($\chi^2 = 1.12$). We note that this structure (Fig. 2E) involves very few inter-domain contacts, thus it cannot be stable in solution: it represents only a single "snapshot" of the flexible N dimer.

To quantify the degree of conformational fluctuations of the N protein dimer, we computed the radius of gyration (R_g) and the maximum dimension (D_{max}) of each of the simulation structures. The maximum dimension is defined as the maximum distance between any two points of the protein. The resulting histograms of R_g and D_{max} together with three representative structures are shown in Fig. 3. We note that in the simulation ensemble (i.e. before refinement), structures with an R_g between 5.3 and 6.3 nm and a D_{max} between about 18 and 20 nm are slightly over-represented, whereas more compact structures with an R_g smaller than 5 nm and a D_{max} smaller than 18 nm are slightly under-represented. Taken together, the histograms in Fig. 3 show that the N protein dimer exhibits large conformational fluctuations with an R_g ranging from about 4 to over 8 nm and a D_{max} between 12 and 32 nm.

To characterize the structural ensemble of the dimer, we identified contacts between pairs of amino acid residues and the relative frequency of their occurrence. To this end, we determined contacts between pairs of amino-acid beads in each of the simulation structures using a distance criterion [33]. The relative frequency f of a given contact in the refined ensemble was then taken as a sum of statistical weights of the conformations in which this contact was present (see Materials and methods). The maps of the relative frequency of the intra- and inter-chain contacts are shown in Fig. 4A and B, respectively. The intra-chain contacts are the contacts formed within each of the polypeptide chains of the dimer. The inter-chain contacts, on the other hand, are those between the two polypeptide chains of the dimer. The points in red indicate frequent contacts with $f > 0.1$. The points in orange correspond to transient contacts with $0.01 < f < 0.1$. The points in yellow indicate rare contacts with $0.001 < f < 0.01$. Finally, the points in gray show contacts within rigid domains that do not evolve in the course of the REMC simulations.

The contact map in Fig. 4A shows that the frequent contacts can be grouped in two categories: *i*) contacts formed within each of the three IDRs and *ii*) contacts formed by either the NTD or the CTD with the hydrophobic α-helix in the IDR2. The contact map in Fig. 4B shows that there are rather few contacts between the two chains and that these contacts are mostly rare or transient.

To characterize the level of extension or compactness of the IDRs, we determined their end-to-end distances in each of the simulation structures. Fig. 5 shows the resulting histograms of the end-to-end distances. The end-to-end distances of IDR1, IDR2 and IDR3 are found to vary from 1 to 9 nm, from 1 to 13 nm, and from 0.5 to 11.5 nm, respectively. The average and root-mean-square values of the end-to-end distances are summarized in Table 1.

It is interesting to compare the root-mean-square end-to-end distances of the IDRs with those of freely-jointed chains, which are also

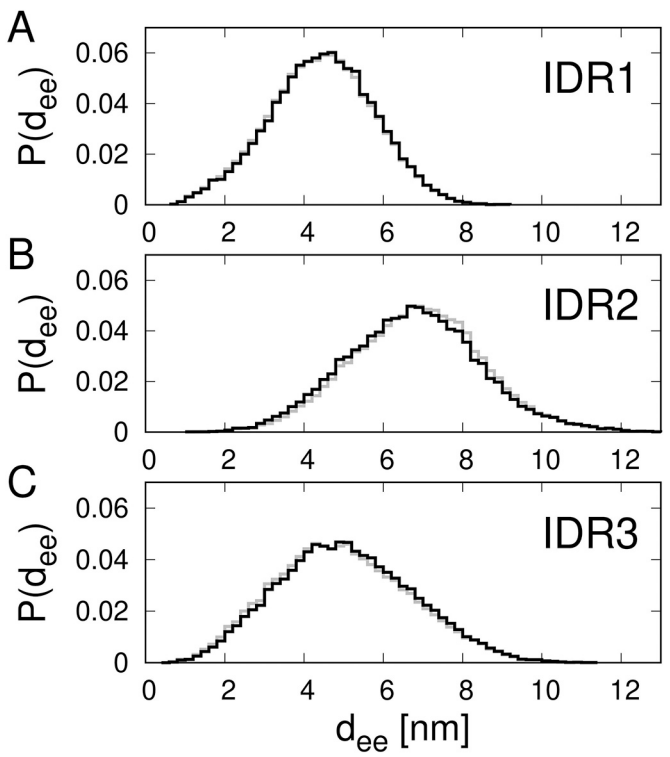

Fig. 5. Histograms of the end-to-end distance of IDR1,2,3 (A,B,C). The lines in gray and black correspond to the ensemble of simulation structures before and after refinement, respectively.

Table 1
Statistical properties of the end-to-end distances of the IDRs.

IDR	Average end-to-end distance [nm]	Root-mean-square end-to-end distance [nm]	Number of amino acid residues (n)	Root-mean-square end-to-end distance of a Gaussian chain of n monomers [nm]	Ratio of the root-mean-square end-to-end distances in columns 3 and 5
1	4.32 ± 0.08	4.52 ± 0.07	40	2.40	1.88
2	6.77 ± 0.29	6.95 ± 0.25	71	3.20	2.17
3	4.87 ± 0.16	5.15 ± 0.15	55	2.82	1.83

often called ideal chains or Gaussian chains. The root-mean-square end-to-end distance of a Gaussian chain of n beads is equal to the square root of n times the length of the bond between consecutive beads in the chain. As in the coarse-grained model for multi-domain IDPs introduced by Kim and Hummer, here we take the chain beads to be centered at the α-carbon atoms and the length of the pseudo-bond between consecutive α-carbon atoms to be 0.38 nm. The results presented in Table 1 show clearly that each of the three IDRs exhibits a root-mean-square end-to-end distance larger than the Gaussian chains of the corresponding length. This observation implies that although the conformations of the IDRs are locally compact (see the contact map in Fig. 4A), they are clearly more extended than those of the Gaussian chains overall. Interestingly, the N- and C-terminal tails (IDR1 and IDR3) behave somewhat more like Gaussian chains compared to the middle linker (IDR2).

4. Discussion

The thermodynamic state of an IDP is an ensemble of rapidly interconverting conformations. Important examples of IDPs are large, multi-domain proteins in which several autonomously folded domains are held together by intrinsically disordered regions [42]. Their structural analysis is difficult because of their large sizes and dynamic nature [4]. A notable example is the nucleocapsid protein of SARS-CoV-2 (Fig. 1). The question we posed in this study was about the degree of conformational fluctuations of this flexible protein complex.

We performed extensive REMC simulations of the N protein dimer and refined the resulting ensemble of conformations using the available SAXS data (Fig. 2). Our results show that the N protein dimer exhibits large conformational fluctuations in solution. Its radius of gyration varies between about 4 and 8 nm with an average of 5.9 nm (Fig. 3A). The maximum dimension exhibits large fluctuations in a range of 12 to 32 nm (Fig. 3B). The NTD does not make direct contacts with the CTD (Fig. 4); however, the hydrophobic α-helix within the IDR2 makes frequent and transient contacts with both the NTD and the CTD. Frequent contacts within each of the three IDRs indicate that these regions can adopt conformations that are locally compact. Yet, our analysis of the distances between termini of the IDRs (Fig. 5 and Table 1) shows that each of the three IDRs exhibits conformations that are more extended than Gaussian chains of corresponding lengths. Taken together, our results provide a detailed picture of the conformational ensemble of the SARS-CoV-2 N protein dimer under near-physiological conditions.

The structure and flexibility of the SARS-CoV-2 N protein is likely to be important for the assembly of the nucleocapsid. Interestingly, both the NTD and CTD can bind RNA, yet it is evident from the contact maps in Fig. 4 that there are practically no direct contacts between the NTD and the CTD. Thus our structural analysis suggests that the NTD and CTD do not directly cooperate in RNA binding. However, the hydrophobic α-helix within the IDR2 makes multiple intra- and inter-chain contacts with both the NTD and CTD. This α-helix makes most frequent contacts with the NTD of the same chain (Fig. 4A) and least frequent contacts with the NTD of the other chain in the N dimer (Fig. 4B). Frequent contacts are observed also within each of the disordered segments IDR1–3. The presence of these contacts indicates that the IDRs adopt conformations that are locally compact. It remains to be shown how IDRs influence other interesting properties of the N protein such as its ability to phase separate with RNA [43–47]. However, we note that the conformations adopted by the N protein could be modified by RNA or protein binding [41,48].

Author contributions

EB and BR planned the research, BR performed the simulations and analyzed the results, EB and BR wrote the manuscript.

Declaration of Competing Interest

The authors declare no conflicts of interests.

Acknowledgements

We thank professors Tengchuan Jin and Hongliang He from the University of Science and Technology of China for sharing their SAXS data with us. We are also grateful to Michael Downey (University of Alberta) for language corrections. The research has been supported by the Polish National Science Centre within the international CEUS-UNISONO program, grant number 2020/02/Y/NZ1/00020 (to BR), and European Regional Development Fund; OP RDE; Project: “Chemical biology for drugging undruggable targets (ChemBioDrug)” (No. CZ.02.1.01/0.0/0.0/16_019/0000729). The Academy of Sciences of the Czech Republic (RVO: 61388963) is also acknowledged. The simulations were carried out using the supercomputer resources at the Centre of Informatics – Tricity Academic Supercomputer & networK (CI TASK) in Gdańsk, Poland.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bpc.2022.106843.

References

[1] R.M. Glaeser, How good can single-particle Cryo-EM become? What remains before it approaches its physical limits? Annu. Rev. Biophys. 48 (48) (2019) 45–61.
[2] X.C. Bai, G. McMullan, S.H.W. Scheres, How cryo-EM is revolutionizing structural biology, Trends Biochem. Sci. 40 (1) (2015) 49–57.
[3] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589.
[4] B. Rozycki, E. Boura, Large, dynamic, multi-protein complexes: a challenge for structural biology, J. Phys. Condens. Matter 26 (46) (2014) 463103.
[5] P. V'Kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for SARS-CoV-2, Nat. Rev. Microbiol. 19 (3) (2021) 155–170.
[6] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q. Chen, Y. Yan, C. Zhang, H. Shan, S. Chen, Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, Acta Pharm. Sin. B 10 (7) (2020) 1228–1238.
[7] D.C. Dinesh, D. Chalupska, J. Silhan, E. Koutna, R. Nencka, V. Veverka, E. Boura, Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein, PLoS Pathog. 16 (12) (2020) e1009100.
[8] M. Yang, S. He, X. Chen, Z. Huang, Z. Zhou, Z. Zhou, Q. Chen, S. Chen, S. Kang, Structural insight into the SARS-CoV-2 nucleocapsid protein C-terminal domain reveals a novel recognition mechanism for viral transcriptional regulatory sequences, Front. Chem. 8 (2020) 624765.
[9] R. Zhou, R. Zeng, A. von Brunn, J. Lei, Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein, Mol. Biomed. 1 (1) (2020) 2.
[10] H. Luo, J. Chen, K. Chen, X. Shen, H. Jiang, Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization, Biochemistry 45 (39) (2006) 11827–11835.
[11] I.M. Yu, M.L. Oldham, J. Zhang, J. Chen, Crystal structure of the severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein dimerization domain reveals evolutionary linkage between corona- and arteriviridae, J. Biol. Chem. 281 (25) (2006) 17134–17139.
[12] W. Zeng, G. Liu, H. Ma, D. Zhao, Y. Yang, M. Liu, A. Mohammed, C. Zhao, Y. Yang, J. Xie, C. Ding, X. Ma, J. Weng, Y. Gao, H. He, T. Jin, Biochemical characterization of SARS-CoV-2 nucleocapsid protein, Biochem. Biophys. Res. Commun. 527 (3) (2020) 618–623.
[13] T.W. Grawert, D.I. Svergun, Structural Modeling using solution small-angle X-ray scattering (SAXS), J. Mol. Biol. 432 (9) (2020) 3078–3092.
[14] C.A. Brosey, J.A. Tainer, Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology, Curr. Opin. Struct. Biol. 58 (2019) 197–213.
[15] M.V. Petoukhov, D.I. Svergun, Global rigid body modeling of macromolecular complexes against small-angle scattering data, Biophys. J. 89 (2) (2005) 1237–1250.
[16] 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.
[17] M. Pelikan, G.L. Hura, M. Hammel, Structure and flexibility within proteins as identified through small angle X-ray scattering, Gen. Physiol. Biophys. 28 (2) (2009) 174–189.
[18] B. Rozycki, Y.C. Kim, G. Hummer, SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions, Structure 19 (1) (2011) 109–116.
[19] W. Peti, R. Page, E. Boura, B. Rozycki, Structures of dynamic protein complexes: hybrid techniques to study MAP kinase complexes and the ESCRT system, Methods Mol. Biol. 1688 (2018) 375–389.
[20] T.N. Cordeiro, F. Herranz-Trillo, A. Urbanek, A. Estana, J. Cortes, N. Sibille, P. Bernado, Small-angle scattering studies of intrinsically disordered proteins and their complexes, Curr. Opin. Struct. Biol. 42 (2017) 15–23.
[21] A. Estana, N. Sibille, E. Delaforge, M. Vaisset, J. Cortes, P. Bernado, Realistic ensemble models of intrinsically disordered proteins using a structure-encoding coil database, Structure 27 (2) (2019), p. 381-+.
[22] Z. Benayad, S. von Bulow, L.S. Stelzl, G. Hummer, Simulation of FUS protein condensates with an adapted coarse-grained model, J. Chem. Theory Comput. 17 (1) (2021) 525–537.
[23] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmuller, A.D. MacKerell Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins, Nat. Methods 14 (1) (2017) 71–73.
[24] B. Turonova, M. Sikora, C. Schurmann, W.J.H. Hagen, S. Welsch, F.E.C. Blanc, S. von Bulow, M. Gecht, K. Bagola, C. Horner, G. van Zandbergen, J. Landry, N.T. D. de Azevedo, S. Mosalaganti, A. Schwarz, R. Covino, M.D. Muhlebach, G. Hummer, J. Krijnse Locker, M. Beck, In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges, Science 370 (6513) (2020) 203–208.
[25] P.R. Pandey, B. Rozycki, R. Lipowsky, T.R. Weikl, Structural variability and concerted motions of the T cell receptor - CD3 complex, Elife 10 (2021).
[26] A.K. Sieradzan, C.R. Czaplewski, E.A. Lubecka, A.G. Lipska, A.S. Karczynska, A. P. Gieldon, R. Slusarz, M. Makowski, P. Krupa, M. Kogut, A. Antoniak, P. A. Wesolowski, A. Augustynowicz, H. Leszczynski, J.A. Liwo, Extension of the Unres package for physics-based coarse-grained simulations of proteins and protein complexes to very large systems, Biophys. J. 120 (3) (2021) 83a–84a.
[27] T. Lazar, E. Martinez-Perez, F. Quaglia, A. Hatos, L.B. Chemes, J.A. Iserte, N. A. Mendez, N.A. Garrone, T.E. Saldano, J. Marchetti, A.J.V. Rueda, P. Bernado, M. Blackledge, T.N. Cordeiro, E. Fagerberg, J.D. Forman-Kay, M.S. Fornasari, T. J. Gibson, G.W. Gomes, C.C. Gradinaru, T. Head-Gordon, M.R. Jensen, E.A. Lemke, S. Longhi, C. Marino-Buslje, G. Minervini, T. Mittag, A.M. Monzon, R.V. Pappu, G. Parisi, S. Ricard-Blum, K.M. Ruff, E. Salladini, M. Skepo, D. Svergun, S.D. Vallet, M. Varadi, P. Tompa, S.C.E. Tosatto, D. Piovesan, PED in 2021: a major update of the protein ensemble database for intrinsically disordered proteins, Nucleic Acids Res. 49 (D1) (2021) D404–D411.
[28] A.G. Kikhney, C.R. Borges, D.S. Molodenskiy, C.M. Jeffries, D.I. Svergun, SASBDB: towards an automatically curated and validated repository for biological scattering data, Protein Sci. 29 (1) (2020) 66–75.
[29] Y.C. Kim, G. Hummer, Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding, J. Mol. Biol. 375 (5) (2008) 1416–1433.
[30] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1) (1996), p. 33–8, 27–8.
[31] J. Kofinger, B. Rozycki, G. Hummer, Inferring structural ensembles of flexible and dynamic macromolecules using Bayesian, maximum entropy, and minimal-ensemble refinement methods, Biomolecul. Simul. 2022 (2019) 341–352.
[32] J. Kofinger, L.S. Stelzl, K. Reuter, C. Allande, K. Reichel, G. Hummer, Efficient ensemble refinement by reweighting, J. Chem. Theory Comput. 15 (5) (2019) 3390–3401.
[33] B. Rozycki, M. Cieplak, M. Czjzek, Large conformational fluctuations of the multi-domain xylanase Z of clostridium thermocellum, J. Struct. Biol. 191 (1) (2015) 68–75.
[34] E. Boura, B. Rozycki, D.Z. Herrick, H.S. Chung, J. Vecer, W.A. Eaton, D.S. Cafiso, G. Hummer, J.H. Hurley, Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 108 (23) (2011) 9437–9442.
[35] E. Boura, B. Rozycki, H.S. Chung, D.Z. Herrick, B. Canagarajah, D.S. Cafiso, W. A. Eaton, G. Hummer, J.H. Hurley, Solution structure of the ESCRT-I and -II supercomplex: implications for membrane budding and scission, Structure 20 (5) (2012) 874–886.
[36] D. Chalupska, A. Eisenreichova, B. Rozycki, L. Rezabkova, J. Humpolickova, M. Klima, E. Boura, Structural analysis of phosphatidylinositol 4-kinase IIIbeta (PI4KB) - 14-3-3 protein complex reveals internal flexibility and explains 14-3-3 mediated protection from degradation in vitro, J. Struct. Biol. 200 (1) (2017) 36–44.
[37] D. Chalupska, B. Rozycki, J. Humpolickova, L. Faltova, M. Klima, E. Boura, Phosphatidylinositol 4-kinase IIIbeta (PI4KB) forms highly flexible heterocomplexes that include ACBD3, 14-3-3, and Rab11 proteins, Sci. Rep. 9 (1) (2019) 567.
[38] D. Chalupska, B. Rozycki, M. Klima, E. Boura, Structural insights into acyl-coenzyme A binding domain containing 3 (ACBD3) protein hijacking by picornaviruses, Protein Sci. 28 (12) (2019) 2073–2079.
[39] V. Horova, H. Lyoo, B. Rozycki, D. Chalupska, M. Smola, J. Humpolickova, J. Strating, F.J.M. van Kuppeveld, E. Boura, M. Klima, Convergent evolution in the mechanisms of ACBD3 recruitment to picornavirus replication sites, PLoS Pathog. 15 (8) (2019), e1007962.
[40] J. Kofinger, M.J. Ragusa, I.H. Lee, G. Hummer, J.H. Hurley, Solution structure of the Atg1 complex: implications for the architecture of the phagophore assembly site, Structure 23 (5) (2015) 809–818.
[41] L.M. Bessa, S. Guseva, A.R. Camacho-Zarco, N. Salvi, D. Maurin, L.M. Perez, M. Botova, A. Malki, M. Nanao, M.R. Jensen, R.W.H. Ruigrok, M. Blackledge, The intrinsically disordered SARS-CoV-2 nucleoprotein in dynamic complex with its viral partner nsp3a, Sci. Adv. 8 (3) (2022) p. eabm4034.
[42] A.K. Dunker, C.J. Oldfield, J.W. Meng, P. Romero, J.Y. Yang, J.W. Chen, V. Vacic, Z. Obradovic, V.N. Uversky, The unfoldomics decade: an update on intrinsically disordered proteins, BMC Genomics 9 (2008).
[43] J. Cubuk, J.J. Alston, J.J. Incicco, S. Singh, M.D. Stuchell-Brereton, M.D. Ward, M. I. Zimmerman, N. Vithani, D. Griffith, J.A. Wagoner, G.R. Bowman, K.B. Hall, A. Soranno, A.S. Holehouse, The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA, Nat. Commun. 12 (1) (2021).
[44] H. Chen, Y. Cui, X.L. Han, W. Hu, M. Sun, Y. Zhang, P.H. Wang, G.T. Song, W. Chen, J.Z. Lou, Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA, Cell Res. 30 (12) (2020) 1143–1145.
[45] C. Iserman, C.A. Roden, M.A. Boerneke, R.S.G. Sealfon, G.A. McLaughlin, I. Jungreis, E.J. Fritch, Y.J. Hou, J. Ekena, C.A. Weidmann, C.L. Theesfeld, M. Kellis, O.G. Troyanskaya, R.S. Baric, T.P. Sheahan, K.M. Weeks, A.S. Gladfelter, Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid, Mol. Cell 80 (6) (2020) 1078–1091 e6.
[46] S.M. Cascarina, E.D. Ross, Phase separation by the SARS-CoV-2 nucleocapsid protein: consensus and open questions, J. Biol. Chem. 298 (3) (2022), 101677.
[47] G.L. Dignon, W.W. Zheng, R.B. Best, Y.C. Kim, J. Mittal, Relation between single-molecule properties and phase behavior of intrinsically disordered proteins, Proc. Natl. Acad. Sci. U. S. A. 115 (40) (2018) 9929–9934.
[48] C.K. Chang, Y.L. Hsu, Y.H. Chang, F.A. Chao, M.C. Wu, Y.S. Huang, C.K. Hu, T. H. Huang, Multiple nucleic acid binding sites and intrinsic disorder of severe acute

respiratory syndrome coronavirus nucleocapsid protein: implications for ribonucleocapsid protein packaging, J. Virol. 83 (5) (2009) 2255–2264.