

Article

# Scrutinising the Conformational Ensemble of the Intrinsically Mixed-Folded Protein Galectin-3

Midhun Mohan Anila <sup>†</sup>, Paweł Rogowski <sup>†</sup> and Bartosz Różycki <sup>\*†</sup>

Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland; midhun@ifpan.edu.pl (M.M.A.); progowski@ifpan.edu.pl (P.R.)

\* Correspondence: rozycki@ifpan.edu.pl

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** Galectin-3 is a protein involved in many intra- and extra-cellular processes. It has been identified as a diagnostic or prognostic biomarker for certain types of heart disease, kidney disease and cancer. Galectin-3 comprises a carbohydrate recognition domain (CRD) and an N-terminal domain (NTD), which is unstructured and contains eight collagen-like Pro-Gly-rich tandem repeats. While the structure of the CRD has been solved using protein crystallography, current knowledge about conformations of full-length galectin-3 is limited. To fill in this knowledge gap, we performed molecular dynamics (MD) simulations of full-length galectin-3. We systematically re-scaled the solute-solvent interactions in the Martini 3 force field to obtain the best possible agreement between available data from SAXS experiments and the ensemble of conformations generated in the MD simulations. The simulation conformations were found to be very diverse, as reflected, e.g., by (i) large fluctuations in the radius of gyration, ranging from about 2 to 5 nm, and (ii) multiple transient contacts made by amino acid residues in the NTD. Consistent with evidence from NMR experiments, contacts between the CRD and NTD were observed to not involve the carbohydrate-binding site on the CRD surface. Contacts within the NTD were found to be made most frequently by aromatic residues. Formation of fuzzy complexes with unspecific stoichiometry was observed to be mediated mostly by the NTD. Taken together, we offer a detailed picture of the conformational ensemble of full-length galectin-3, which will be important for explaining the biological functions of this protein at the molecular level.

**Keywords:** galectin-3; conformational ensemble; fuzzy complexes; molecular dynamics simulations; Martini 3



**Citation:** Anila, M.M.; Rogowski, P.; Różycki, B. Scrutinising the Conformational Ensemble of the Intrinsically Mixed-Folded Protein Galectin-3. *Molecules* **2024**, *29*, 2768. <https://doi.org/10.3390/molecules29122768>

Academic Editor: Ricardo L. Mancera

Received: 26 April 2024

Revised: 24 May 2024

Accepted: 6 June 2024

Published: 11 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Galectins are soluble proteins that contain at least one carbohydrate recognition domain (CRD) but no other types of folded protein domains [1,2]. They bind specifically to  $\beta$ -galactosides. Their molecular weights range from about 14 to 39 kDa [2,3]. In mammals, the galectin family consists of 15 members, which are involved in many biological processes, including immune responses, cell migration, autophagy and signaling [4]. Within this protein family, galectin-3 is unique in the sense that it bears a chimeric molecular architecture. Namely, it consists of a single CRD and an unstructured, non-lectin N-terminal domain (NTD) containing eight collagen-like Pro-Gly-rich tandem repeats. This unique architecture enables galectin-3 not only to interact with a plethora of ligands in a carbohydrate-dependent or -independent manner but also to form oligomers [5–7].

Galectin-3 is shuttled between the nucleus and cytoplasm [8]. It is also secreted to the cell surface and into extra-cellular fluids [9]. The different locations of galectin-3 contribute to its diverse functions, which include cell growth, cell adhesion, angiogenesis, apoptosis and mRNA processing [10]. Overexpression or changes in the localization of galectin-3 are commonly seen in various types of cancers. Increasing evidence indicates that galectin-3 is required for the regulation of various cell activities during cancer development, progression

and metastasis [10]. Consequently, galectin-3 has been identified as a biomarker for disease diagnosis and a target for therapy [11].

Galectin-3 has been studied extensively using molecular biophysics and structural biology methods. In particular, the structure of the CRD has been solved using protein crystallography, revealing in atomic detail the carbohydrate-binding specificity of galectin-3 [12,13]. Small-angle X-ray scattering (SAXS) experiments have demonstrated that the NTD is unstructured and that the radius of gyration of the full-length galectin-3 molecule in solution is about 2.9 nm [7]. Nuclear magnetic resonance (NMR) methods have been used to study the inter- and intra-molecular interactions of galectin-3 [6,7]. However, despite this significant research progress, the current knowledge about the conformational dynamics of full-length galectin-3 is still limited. To fill in this knowledge gap, we performed molecular dynamics (MD) simulations of full-length galectin-3. We used the Martini 3 force field [14] and, following the approach of Thomassen et al. [15], re-scaled the protein–water interactions in such a way that the ensemble of conformations generated in the MD simulations was in good quantitative agreement with the data from SAXS experiments [7]. Our simulation results revealed the following picture: Full-length galectin-3 exhibits large-scale fluctuations between diverse compact and extended conformations, with the radius of gyration ranging from about 2 to 5 nm. The NTD appears to be as flexible as a Gaussian chain, and virtually any pair of amino acid residues in the NTD makes transient contacts. Many different residues in the NTD make transient contacts with the same regions on the CRD surface, which is a characteristic of fuzzy interactions, as identified in the NMR experiments [7]. Aromatic residues in the NTD make rather frequent contacts, both with the CRD and within the NTD. The convex face of the CRD comes into contact with the NTD significantly more often than the concave face of the CRD, which appears to be qualitatively consistent with enhanced chemical shifts for residues 200–220 located on the convex side of the CRD [7]. Moreover, the carbohydrate-binding site—positioned on the concave surface of the CRD—is practically never occluded by the NTD, which potentially has biological implications. Taken together, our simulation results are in good quantitative agreement with the SAXS data and qualitatively consistent with conclusions drawn from the NMR experiments, and provide a detailed account of the conformational ensemble of full-length galectin-3 under SAXS experimental conditions.

The CRD of galectin-3 has been studied to some extent using MD methods. Specifically, MD simulations have elucidated the mechanisms of oligosaccharide recognition in the CRD [16] and revealed the mode of CRD binding to ganglioside GM1 in a model lipid membrane [17]. In addition, a combination of NMR and MD has been used to explore interactions between the NTD and the CRD in full-length galectin-3 [18]. In contrast to the aforementioned studies employing all-atom MD simulations, we used the coarse-grained Martini 3 model to explore the conformational dynamics as well as the intra- and inter-molecular interactions of full-length galectin-3.

MD simulations of intrinsically disordered proteins (IDPs) are often challenging and time consuming. Firstly, IDPs typically exhibit large-scale conformational fluctuations that occur on long time scales. Therefore, the simulation time has to be sufficiently long to capture the IDP dynamics [19]. Secondly, the simulation box has to be large enough to accommodate the most extended conformations of the IDP under study; and the computational cost of explicit solvent simulations certainly increases as the simulation box is made larger. One way to significantly reduce the computational cost of MD simulations is to use coarse-grained models, where groups of atoms are represented by beads. Although coarse-grained models have a lower spatial resolution and weaker predictive power than all-atom models, they have proven very useful in diverse studies on biomacromolecules, including IDPs [15,20].

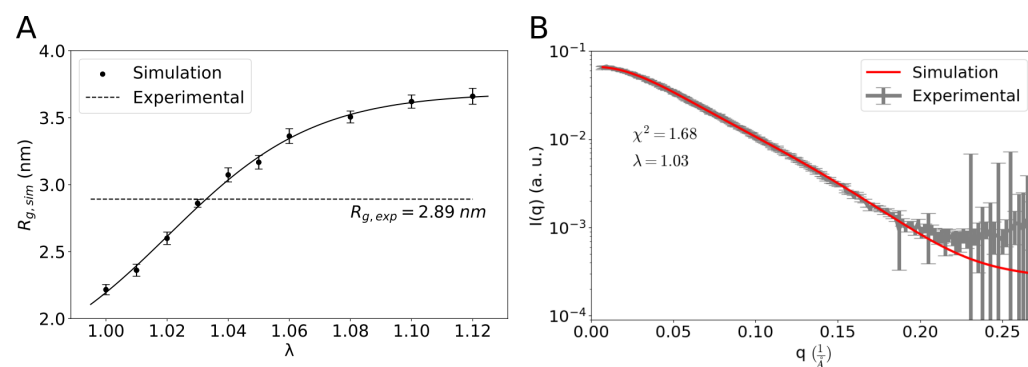
One of the most commonly used coarse-grained models for simulating biomacromolecules is currently Martini 3 [14]. We performed the MD simulations of full-length galectin-3 using Martini 3, which allowed us to study the conformational dynamics and inter-molecular interactions of this IDP on time scales up to 100  $\mu$ s. The MD simulation

results show, in particular, that the carbohydrate-binding site on the CRD surface barely ever makes intra-molecular contacts with the NTD and virtually no inter-molecular contacts at the protein concentration of 400  $\mu\text{M}$ . These results imply that neither inter- nor intra-molecular interactions occlude the carbohydrate-binding site on the CRD surface. The MD simulations show also, consistent with experimental evidence [7,21], that interactions between galectin-3 molecules are mediated mostly by the NTDs. These results together may be important for explaining the biological functions of galectin-3 at the molecular level.

## 2. Results

We performed MD simulations of full-length galectin-3 using the Martini 3 force field [14] and noticed that the simulation structures were, on average, significantly more compact than those observed in SAXS experiments [7]. Indeed, from the MD simulations, we obtained the average radius of gyration  $R_{g,\text{sim}} = 2.22 \pm 0.08$  nm, which was clearly smaller than  $R_{g,\text{exp}} = 2.89$  nm determined from the SAXS experimental data [7]. Therefore, we followed the approach of Thomassen et al. [15] and re-scaled the protein–water interactions by a factor  $\lambda$  in the Martini 3 force field. Specifically, we performed a series of MD simulations of one galectin-3 molecule in a cubic box of side length  $L = 20$  nm with  $\lambda = 1, 1.01, \dots, 1.1, 1.12$ . As expected, the average radius of gyration,  $R_{g,\text{sim}}$ , was found to increase monotonically with  $\lambda$  (Figure 1A). At  $\lambda = 1.03$ , we obtained  $R_{g,\text{sim}} = 2.86 \pm 0.06$  nm, consistent with the  $R_{g,\text{exp}} = 2.89$  nm obtained from the SAXS experiment [7].

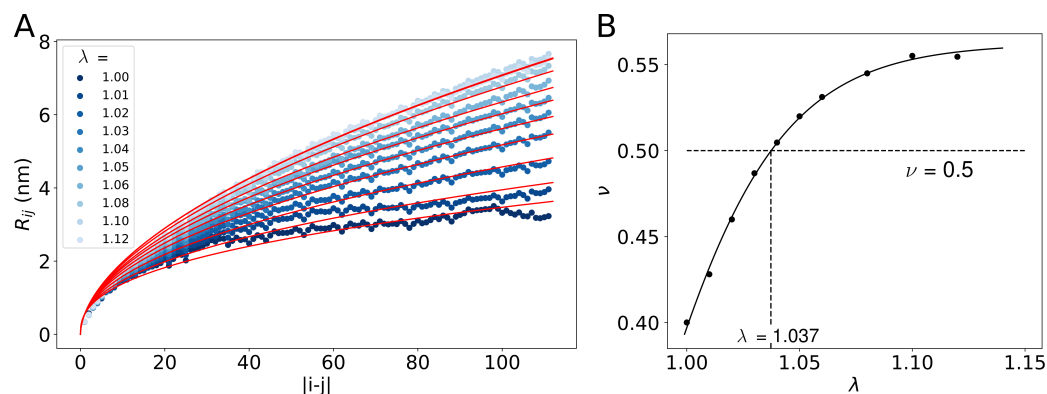
We quantitatively compared the SAXS experimental data with the ensemble of galectin-3 structures generated in the MD simulations with  $\lambda = 1.03$  (Figure 1B). Firstly, we computed scattering intensity profiles for all of the simulation structures individually. Then, we averaged the scattering intensity profiles and used the least-squares method to fit the average profile to the SAXS data published by Lin et al. [7]. We found that the ensemble-averaged scattering intensity profile was in good agreement with the SAXS data ( $\chi^2 = 1.68$ , Figure 1B), validating our simulation outcome.



**Figure 1.** (A) The average radius of gyration,  $R_{g,\text{sim}}$ , as a function of the protein–water interaction re-scaling parameter  $\lambda$ . The error bars indicate standard deviation calculated over a single trajectory using block averaging. The solid line is shown to guide the eye. The dashed line indicates  $R_{g,\text{exp}} = 2.89$  nm, obtained from SAXS experiments [7]. (B) Comparison of the SAXS experimental data taken from Ref. [7] (gray points with error bars) with the scattering intensity profile obtained from the ensemble of the simulation structures with  $\lambda = 1.03$  (red line).

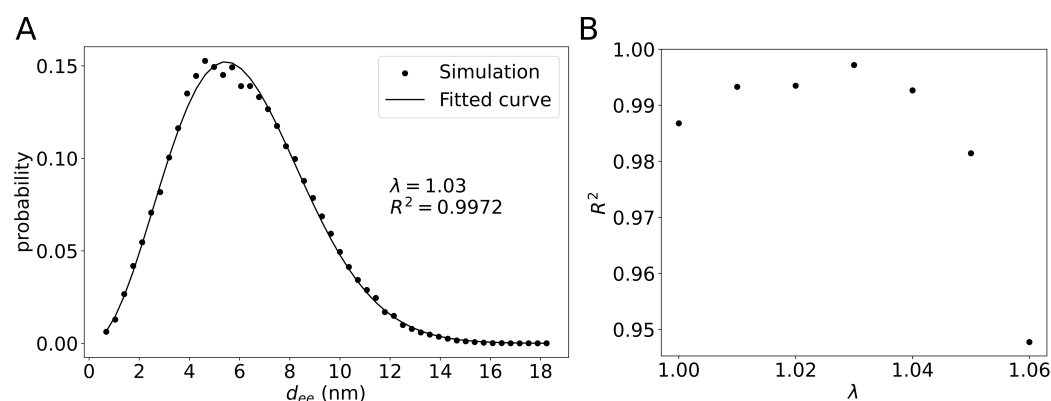
Interestingly, the re-scaling of the protein–water interactions by just a few percent resulted in significant changes in the average radius of gyration (Figure 1A), which suggested that the spatial extent of the NTD could depend sensitively on the value of parameter  $\lambda$ . We set out to investigate this effect. To this end, we followed the approach of Dignon et al. [22] and computed the average inter-residue distance  $R_{ij}$  as a function of sequence separation  $|i - j|$ , with  $i$  and  $j$  being indices of residues in the NTD (Figure 2A). For each of the  $\lambda$ -values, we fitted the resulting distances  $R_{ij}$  to the function  $R_{ij}(\nu) = b|i - j|^\nu$ , where the Kuhn length  $b$  was set to 0.55 nm, as suggested for IDPs [22–24]. We found the Flory scaling exponent  $\nu$  increased monotonically from about 0.4 to 0.55 as  $\lambda$  was increased from 1 to 1.1

(Figure 2B). Thus the solvent equality could be conveniently controlled by tuning the value of parameter  $\lambda$ . The  $\theta$ -point defined by  $\nu = \frac{1}{2}$  was identified at  $\lambda = 1.037$ . Interestingly, the  $\lambda$ -value at which the ensemble of simulation structures was found to fit the SAXS data best,  $\lambda = 1.03$ , was very close to the  $\theta$ -point with  $\lambda = 1.037$ , at which the NTD should effectively behave like a Gaussian chain.



**Figure 2.** (A) The average inter-residue distance  $R_{ij}$  as a function of sequence separation  $|i - j|$ , where  $i$  and  $j$  denote indices of residues in the NTD. The points in different hues of blue represent  $R_{ij}$  values obtained from the MD simulation trajectories with different values of  $\lambda$ , as indicated in the inset. The solid lines in red represent fits of the function  $R_{ij}(\nu) = b|i - j|^\nu$  to the simulation data, where  $b = 0.55$  nm, and  $\nu$  is the Flory scaling exponent. (B) The scaling exponent  $\nu$  as a function of parameter  $\lambda$ . The points in black represent the results of fitting in panel (A). The solid line represents a smooth fit to the points shown in black. The intersection of  $\nu(\lambda)$  with  $\nu = \frac{1}{2}$  yields the  $\theta$ -point at  $\lambda = 1.037$ .

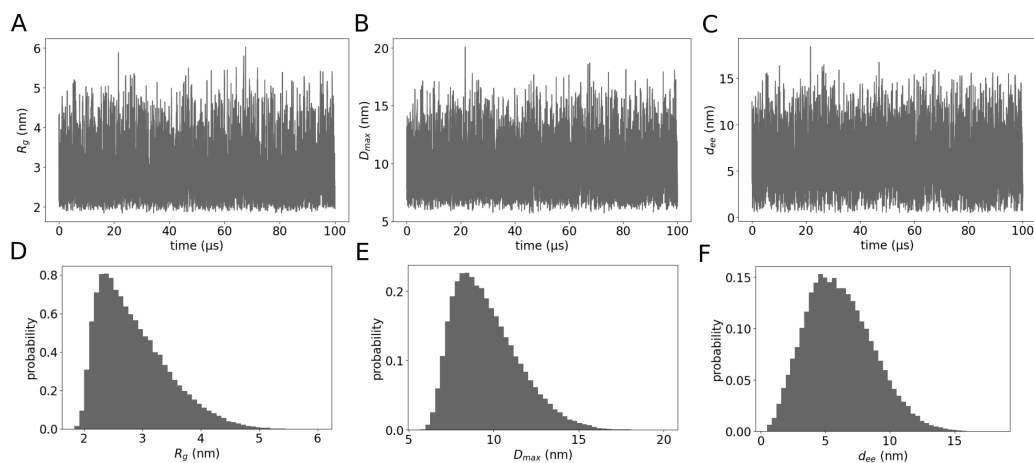
We also determined distributions of the end-to-end distance  $d_{ee}$  for the different values of parameter  $\lambda$ . The end-to-end distance  $d_{ee}$  was defined simply as the distance between the N-terminus and the C-terminus of galectin-3. We noticed that, for  $\lambda = 1.03$ , the end-to-end distance distribution was approximated very well by the Gaussian chain end-to-end distance distribution  $p(d_{ee}) = 4d_{ee}^2 \exp[-(d_{ee}/\zeta)^2]/(\sqrt{\pi}\zeta^3)$  [25] with the fitting parameter  $\zeta = 5.44$  nm (Figure 3A). More generally, the end-to-end distance distributions obtained from the MD simulations could be fitted well to the function  $p(d_{ee})$  with the Pearson correlation coefficient  $R^2 > 0.98$  as long as  $\lambda < 1.06$  (Figure 3B). For  $\lambda > 1.06$ , we obtained much worse fits with  $R^2 < 0.9$ , indicating that conformations of the NTD were significantly different to Gaussian chain conformations.



**Figure 3.** (A) Distribution of the end-to-end distance  $d_{ee}$ . The points represent results obtained from the MD simulation with  $\lambda = 1.03$ . The solid line shows the Gaussian chain end-to-end distance distribution  $p(d_{ee}) = 4d_{ee}^2 \exp[-(d_{ee}/\zeta)^2]/(\sqrt{\pi}\zeta^3)$  with  $\zeta = 5.44$  nm. (B) The Pearson correlation coefficient  $R^2$  obtained by fitting the function  $p(d_{ee})$  to the end-to-end distance distributions from the MD simulations with different  $\lambda$ -values.

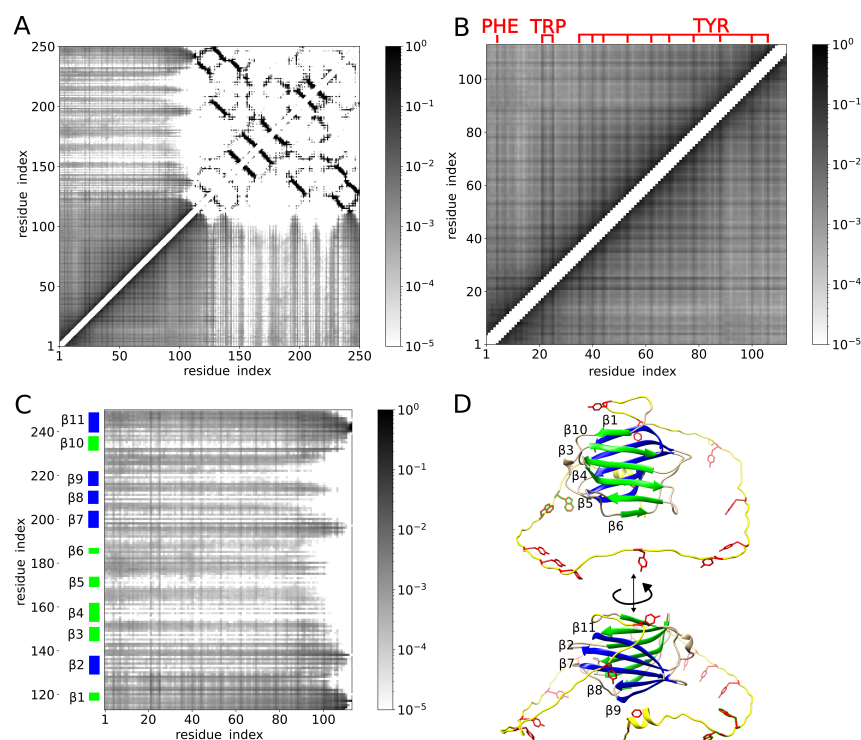
We quantified the ensemble of the full-length galectin-3 structures generated in the MD simulations with  $\lambda = 1.03$ . Firstly, we computed the gyration radius  $R_g$ , the maximum diameter  $D_{\max}$  and the end-to-end distance  $d_{ee}$  for each of the simulation structures. The maximum diameter was defined as the largest distance between any pair of amino acid residues in the protein. We found galectin-3 undergoes large conformational fluctuations, with  $R_g$  ranging from 2 to 6 nm (Figure 4A,D),  $D_{\max}$  fluctuating in the range from about 6 to 18 nm (Figure 4B,E) and  $d_{ee}$  varying between 1 and 16 nm (Figure 4C,F). The values of  $D_{\max}$  and  $d_{ee}$  were thus smaller by at least a couple of nanometers than the simulation box size  $L = 20$  nm, ensuring that galectin-3 did not interact with its periodic image during the MD simulations. In addition, we found the fluctuations in  $R_g$ ,  $D_{\max}$  and  $d_{ee}$  to take place on time scales much shorter than the simulation time of 100  $\mu\text{s}$  (Figure 4A–C), indicating that the simulation system was at equilibrium. We also noticed the distributions of  $R_g$  and  $D_{\max}$  to be asymmetric, exhibiting rather long tails (Figure 4A,B). The latter observation suggests that the NTD of galectin-3 undergoes spontaneous transitions between compact and extended conformations.

Next, we determined contacts between amino acid residues in each of the simulation structures and analyzed how many times each of the contacts was present in the structural ensemble. The resulting map of contacts is shown in Figure 5A, where the hues of gray correspond to the decimal logarithm of contact probability. Contacts found frequently in the structural ensemble are marked in dark gray. Rare contacts are marked in light gray. Contacts between amino acid residues that are close in sequence (i.e., contacts between residue pairs  $(i, i + 1)$ ,  $(i, i + 2)$  and  $(i, i + 3)$ , where index  $i$  labels the amino acid residues) are excluded from the map. The upper-right part of the map shows contacts within the CRD. These are mostly contacts between  $\beta$ -strands stabilizing the CRD structure. The lower-left part of the map corresponds to contacts within the NTD. The lower-right and upper-left parts of the map show contacts between the NTD and the CRD.



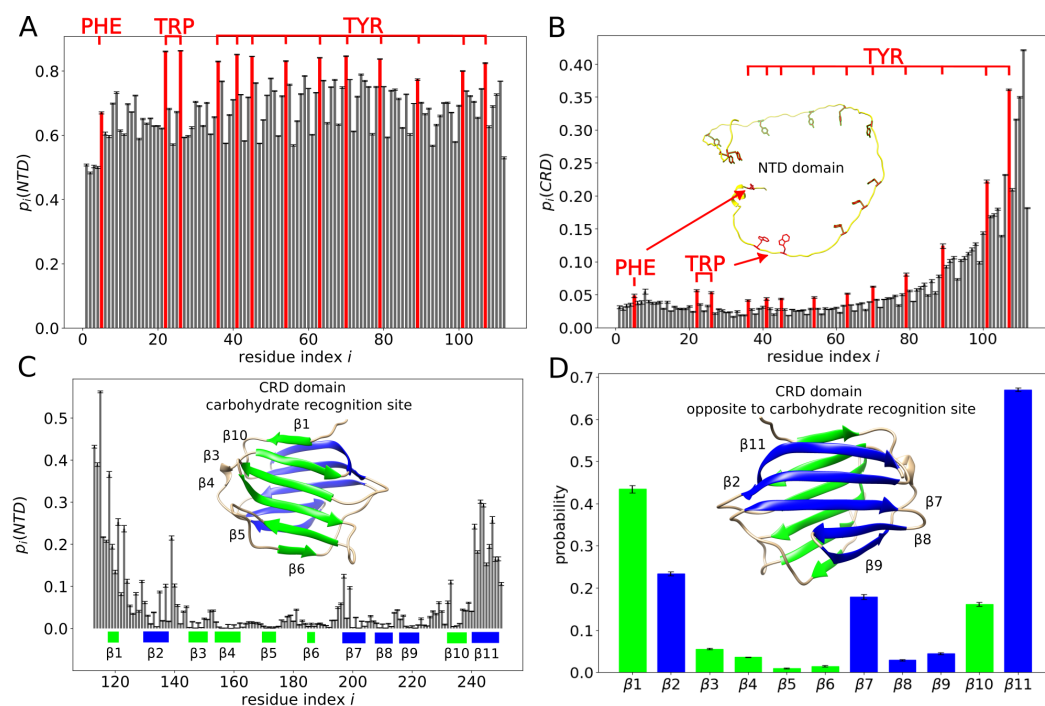
**Figure 4.** (A–C) The gyration radius  $R_g$ , the maximum diameter  $D_{\max}$  and the end-to-end distance  $d_{ee}$  in the course of the MD simulation with  $\lambda = 1.03$ . (D–F) Histograms of  $R_g$ ,  $D_{\max}$  and  $d_{ee}$  generated from the time series shown in panels (A–C).

Interestingly, inspection of the contact map revealed contacts between virtually all pairs of amino acid residues in the NTD (Figure 5B), indicating a large diversity of conformations of the NTD. However, we found that the most frequent contacts within the NTD were those involving aromatic residues (Figure 5B). In addition, we found the aromatic residues in the NTD making relatively frequent contact also with many amino acid residues of the CRD (Figure 5C). However, we also noticed the NTD contacting only very rarely certain regions of the CRD (Figure 5C), including the  $\beta$ -strands  $\beta_4$  and  $\beta_5$  in particular.



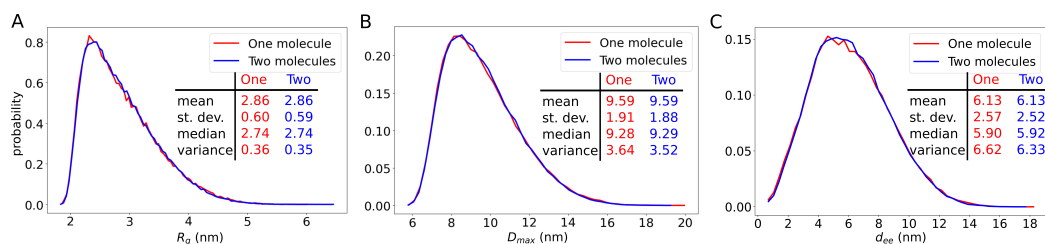
**Figure 5.** (A) Map of contacts between pairs of amino acid residues. The hues of gray correspond to the decimal logarithm of the contact probability. The frequent and rare contacts are marked in dark and light gray, respectively. Contacts between amino acid residues that are close in sequence (namely, contacts between residue pairs  $(i, i + 1)$ ,  $(i, i + 2)$  and  $(i, i + 3)$ ) are excluded from the map. (B) Map of contacts between amino acid residues within the NTD. The location of aromatic residues (PHE, TRP and TYR) in the amino acid sequence is highlighted in red. (C) Map of residue contacts between the NTD and the CRD. The  $\beta$ -strands within the CRD are indicated on the vertical axis. The  $\beta$ -strands on the concave and convex side of the CRD are marked in green and blue, respectively. (D) Cartoon of the galectin-3 molecule. The NTD is marked in yellow, except for the aromatic residues, which are shown in red in the stick representation. The concave face of the CRD is shown in green. It consists of  $\beta$ -strands  $\beta_1$ ,  $\beta_{10}$ ,  $\beta_3$ ,  $\beta_4$ ,  $\beta_5$ , and  $\beta_6$ . Notably, the carbohydrate-binding site is formed by  $\beta_4$ ,  $\beta_5$  and  $\beta_6$ . The convex face of the CRD is shown in blue. It consists of  $\beta$ -strands  $\beta_{11}$ ,  $\beta_2$ ,  $\beta_7$ ,  $\beta_8$  and  $\beta_9$ .

To further explore these observations, we computed the probability  $p_i(D)$  of contact between a given residue with index  $i$  and any residue within domain  $D = \text{NTD}$  or  $\text{CRD}$ . Contacts between residue pairs  $(i, i + 1)$ ,  $(i, i + 2)$  and  $(i, i + 3)$  were excluded from the analysis. We found  $p_i(\text{NTD}) > 0.5$  for any residue  $i$  in the NTD (Figure 6A). Moreover, for almost all aromatic residues in the NTD, we found  $p_i(\text{NTD}) \approx 0.8$ , which was clearly larger than for other residues in the NTD (Figure 6A). The aromatic residues in the NTD were found to exhibit also larger probabilities  $p_i(\text{CRD})$  than residues in their vicinity (Figure 6B). The residues in the CRD that we found to exhibit the largest probabilities  $p_i(\text{NTD})$  of contact with the NTD were those within and around the  $\beta$ -strands  $\beta_1$  and  $\beta_{11}$  (Figure 6C,D). Most of the residues forming  $\beta$ -strands  $\beta_3$ ,  $\beta_4$ ,  $\beta_5$ ,  $\beta_6$ ,  $\beta_8$  and  $\beta_9$  were found to exhibit rather small probabilities  $p_i(\text{NTD})$  (Figure 6C,D). In particular, we found the  $\beta$ -strands  $\beta_3$  and  $\beta_4$  making practically no contacts with the NTD (Figure 6D). This result is interesting because the carbohydrate-binding site on the CRD surface is located within the  $\beta$ -strands  $\beta_3$ ,  $\beta_4$  and  $\beta_5$ .

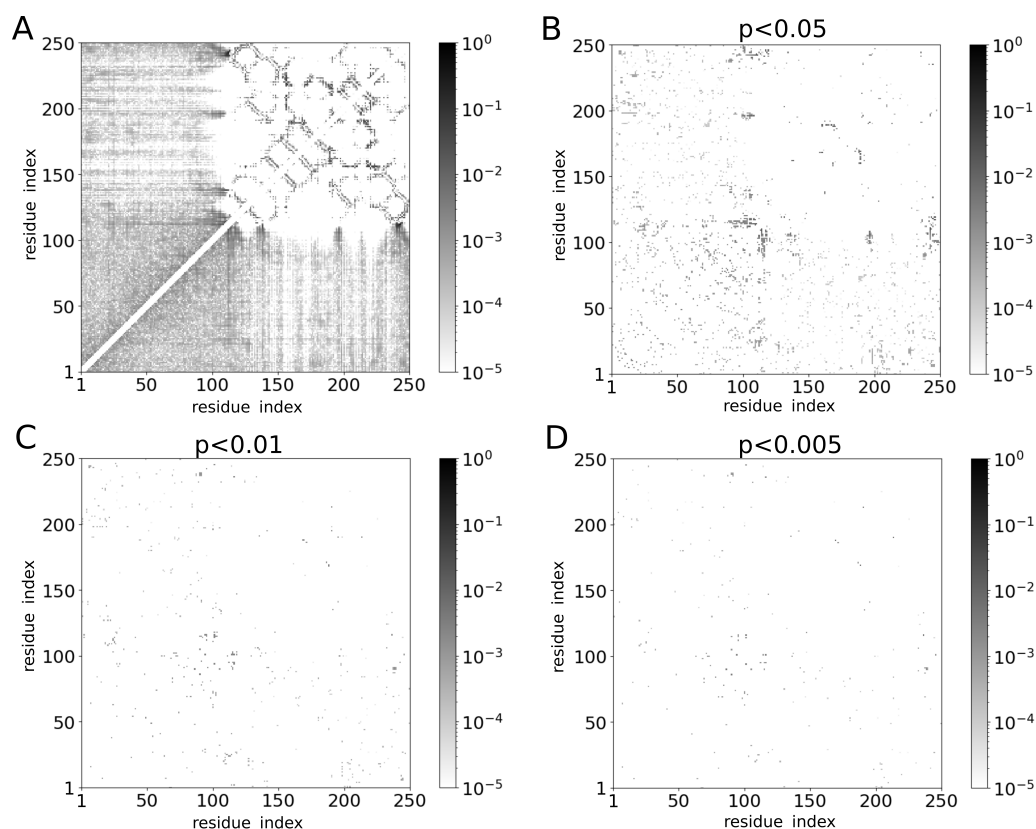


**Figure 6.** (A–C) Probability  $p_i(D)$  of contact between a given residue with index  $i$  and any residue within domain  $D = \text{NTD}$  or  $\text{CRD}$ . The color code is as in Figure 5. (A)  $p_i(\text{NTD})$  for  $i$  within the NTD. (B)  $p_i(\text{CRD})$  for  $i$  within the NTD. (C)  $p_i(\text{NTD})$  for  $i$  within the CRD. (D) Probability of contact between the NTD and a specific  $\beta$ -strand in the CRD.

We next performed MD simulations of two molecules of galectin-3 in a cubic box of side length  $L = 20$  nm using  $\lambda = 1.03$ . The protein concentration in these simulations was 0.4 mM, as in the NMR experiments reported by Lin et al. [7]. We determined the gyration radius  $R_g$  (Figure 7A), the maximum diameter  $D_{\max}$  (Figure 7B) and the end-to-end distance  $d_{ee}$  (Figure 7C) for each of the galectin-3 structures generated in these simulations. Unexpectedly, we found that distributions of these quantities were practically identical to the corresponding distributions obtained from the MD simulations of a single molecule of galectin-3 (Figure 7), suggesting that the conformations of galectin-3 were unaffected by inter-molecular interactions at this protein concentration. To investigate this observation in more detail, we analyzed intra-molecular contacts. Specifically, we computed the probabilities  $p_{ij}^{(\text{mon})}$  and  $p_{ij}^{(\text{dim})}$  of the intra-molecular contacts between residues  $i$  and  $j$  in the MD trajectory with one and two molecules of galectin-3, respectively, and plotted a map of  $|p_{ij}^{(\text{mon})} - p_{ij}^{(\text{dim})}|$  (Figure 8). Although differences between  $p_{ij}^{(\text{mon})}$  and  $p_{ij}^{(\text{dim})}$  were noticeable for many residue pairs (Figure 8A), the ANOVA test showed that the great majority of these differences were statistically insignificant (Figure 8B–D). We thus concluded that the intra-molecular contacts and conformations of galectin-3 were almost unaffected by inter-molecular interactions at the protein concentration of 0.4 mM (Figure S1).



**Figure 7.** Distributions of (A) the radius  $R_g$  of gyration, (B) the maximum diameter  $D_{\max}$  and (C) the end-to-end distance  $d_{ee}$ . The lines in red and blue were obtained from the MD simulations of one and two molecules of galectin-3, respectively.



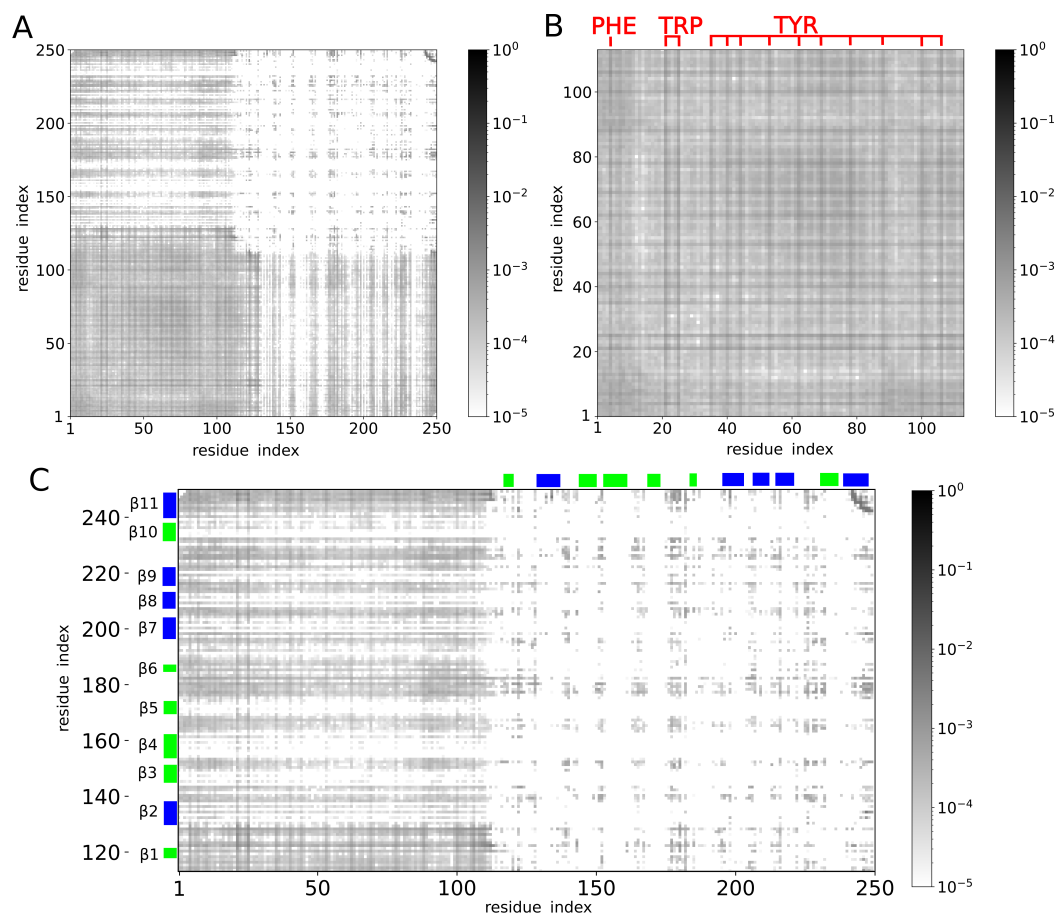
**Figure 8.** (A) Differences between probabilities of intra-molecular contacts observed in the MD trajectories with one and two molecules of galectin-3. The hues of gray correspond to  $\log_{10} |p_{ij}^{(\text{mon})} - p_{ij}^{(\text{dim})}|$ , where  $p_{ij}^{(\text{mon})}$  and  $p_{ij}^{(\text{dim})}$  denote the probabilities of intra-molecular contacts between residues  $i$  and  $j$  in the MD simulation with one and two molecules of galectin-3, respectively. (B–D) Analogous to panel A but for contacts with the ANOVA test  $p$ -values of (B) 0.05, (C) 0.01 and (D) 0.005.

We analyzed the maps shown in Figure 8C,D (corresponding to the ANOVA test  $p$ -values of 0.01 and 0.005, respectively) and selected all contacts with  $|p_{ij}^{(\text{mon})} - p_{ij}^{(\text{dim})}| > 0.001$ . Interestingly, we found that the great majority of these contacts were formed by aromatic residues and prolines (and, to a much smaller extent, also by glycines, serines and alanines) in the NTD and at the NTD–CRD interface (i.e., around residues 110 to 120). These amino acid residues (especially the aromatic ones) were also found to be involved in inter-molecular contacts (Figure 9).

We next analyzed inter-molecular contacts, i.e., contacts between the two molecules of galectin-3 (Figure 9A). We found that contacts between the NTDs were rather unspecific and fuzzy, although formed relatively often by aromatic residues (Figure 9B). The aromatic residues in the NTD were also noticed to make inter-molecular contacts with many residues on the CRD surface (Figure 9C). The amino acid residues in the CRD that were identified as making inter-molecular contacts were mostly those positioned within the loops and turns between the  $\beta$ -strands as well as within the  $\beta$ -strands  $\beta_1$  and  $\beta_{11}$  (Figure 9C).

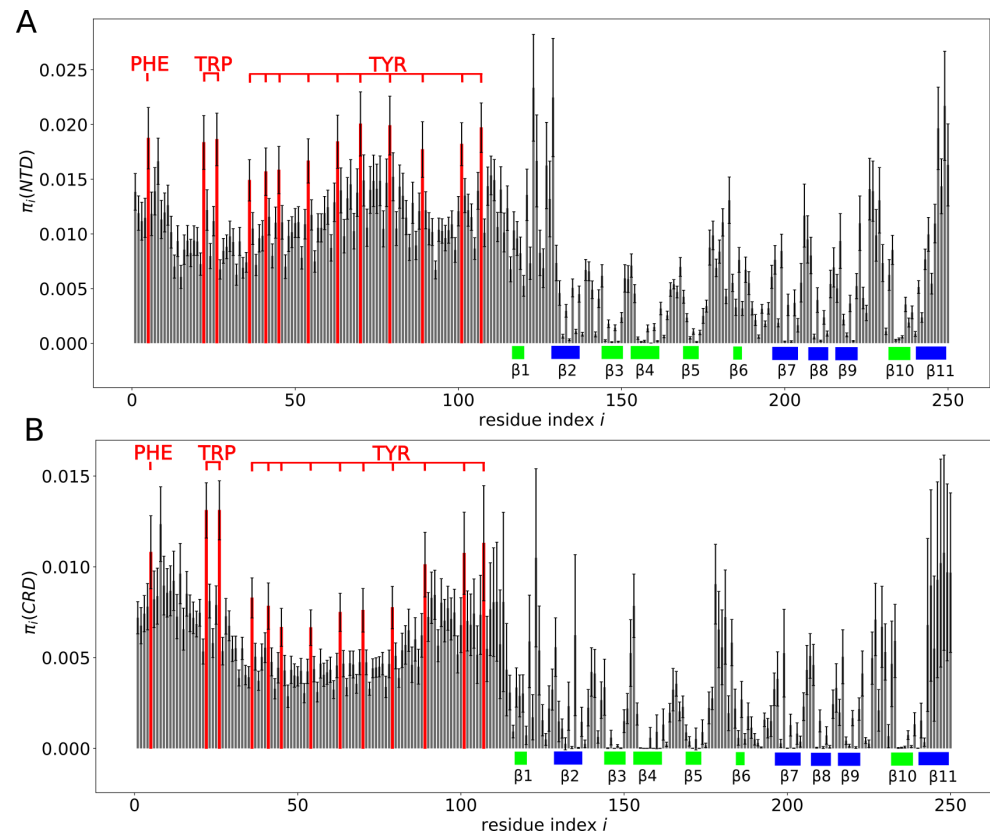
We also determined the probability  $\pi_i(D)$  of contact between a given residue with index  $i$  in one molecule of galectin-3 and a given domain  $D = \text{NTD}$  or  $\text{CRD}$  of the other molecule of galectin-3 (Figure 10). We found  $\pi_i(\text{NTD}) \approx 0.01$  and  $\pi_i(\text{CRD}) \approx 0.005$  for almost any amino acid residue  $i$  within the NTD, as is characteristic for fuzzy inter-molecular interactions. In contrast, for amino acid residues within the CRD, both  $\pi_i(\text{NTD})$  and  $\pi_i(\text{CRD})$  were found to depend sensitively on the residue index  $i$ . In particular, we observed a vanishing  $\pi_i(\text{NTD})$  and  $\pi_i(\text{CRD})$  for most of residues within the  $\beta$ -strands  $\beta_3$ ,  $\beta_4$ ,  $\beta_5$ ,  $\beta_6$  and  $\beta_{10}$ , i.e., on the concave face of the CRD.



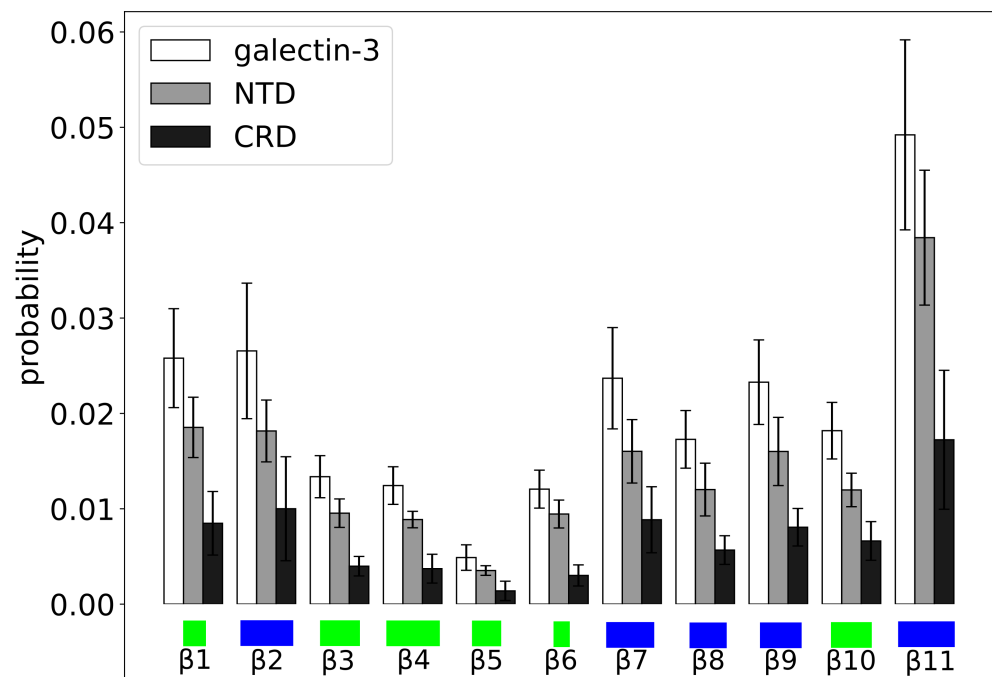


**Figure 9.** (A) Map of contacts between two molecules of galectin-3 at the protein concentration of 0.4 mM. The hues of gray correspond to the decimal logarithm of the contact probability. (B) Map of inter-molecular contacts between the NTDs. The location of aromatic residues in the amino acid sequence is highlighted in red. (C) Map of inter-molecular contacts formed by amino acid residues in the CRD. The  $\beta$ -strands on the concave and convex side of the CRD are marked in green and blue, respectively.

We also computed the probability of contact between a given  $\beta$ -strand ( $\beta_1, \dots, \beta_{11}$ ) in the CRD of one molecule of galectin-3 and a given domain (NTD or CRD) of the other molecule of galectin-3 (Figure 11). We found  $\beta_{11}$  making inter-molecular contacts most frequently, with both the NTD and the CRD. We identified  $\beta_3, \beta_4, \beta_5$  and  $\beta_6$  (forming the concave face of the CRD) as the  $\beta$ -strands to be making inter-molecular contacts most rarely with both the NTD and the CRD. Interestingly, we found each of the  $\beta$ -strands making inter-molecular contacts more often with the NTD than with the CRD, indicating that the NTD–CRD contacts occurred in the MD trajectory more often than the CRD–CRD contacts. To further quantify this observation, we computed the probability  $\pi(D_1, D_2)$  of contact between domain  $D_1 = \text{NTD}$  or CRD of one molecule of galectin-3 and domain  $D_2 = \text{NTD}$  or CRD of the other molecule of galectin-3. We obtained  $\pi(\text{NTD}, \text{NTD}) = 0.14 \pm 0.03$ ,  $\pi(\text{NTD}, \text{CRD}) = 0.11 \pm 0.03$  and  $\pi(\text{CRD}, \text{CRD}) = 0.07 \pm 0.02$ , implying that interactions between galectin-3 molecules in the MD trajectory were mediated mostly by the NTDs, consistent with evidence from NMR experiments [7].



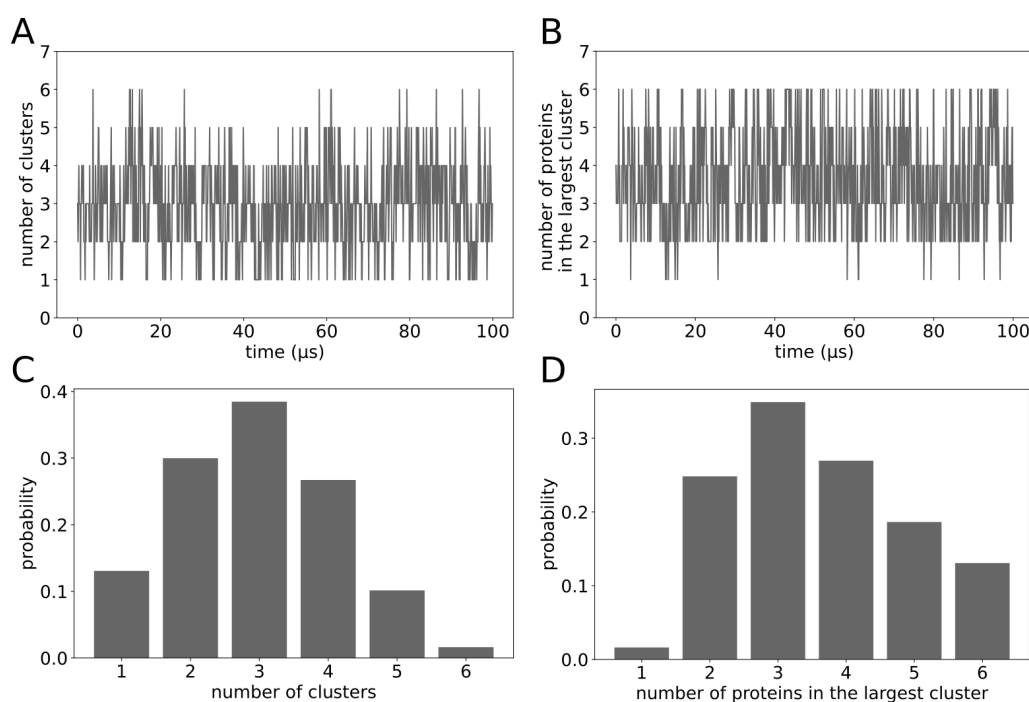
**Figure 10.** Probabilities of contact between a given residue in one molecule of galectin-3 and a given domain (A) NTD or (B) CRD of the other molecule of galectin-3. The location of aromatic residues in the NTD is highlighted in red. The  $\beta$ -strands on the concave and convex face of the CRD are marked in green and blue, respectively.



**Figure 11.** Probability of contact between a given  $\beta$ -strand in the CRD of one molecule of galectin-3 and the NTD (bars in gray) or the CRD (bars in black) of the other molecule of galectin-3. The  $\beta$ -strands on the concave and convex side of the CRD are marked in green and blue, respectively.

In a case of insufficient sampling in the MD simulations, inter-molecular contact probabilities (Figures 9–11) may be biased by the initial placement of the galectin-3 molecules in the simulation box. To rule out this possibility, we analyzed the relative positions and orientations of the two proteins in the MD trajectory (Figure S2). We found that the distance between the centers of the two CRDs fluctuated rapidly between about 3 and 17 nm on time scales much shorter than the simulation time of 100  $\mu$ s (Figure S2A). Also the relative orientation of the two CRDs was found to change quickly between 0 and 180° on time scales much shorter than 100  $\mu$ s (Figure S2B). We thus concluded that 100  $\mu$ s of the coarse-grained dynamics was sufficient to sample all possible relative positions and orientations of the two molecules of galectin-3 in the simulation box.

To study galectin-3 oligomerization [26], we performed MD simulations of six molecules of galectin-3 in a cubic box of side length  $L = 20$  nm with  $\lambda = 1.03$  and determined clusters of the protein molecules. Two molecules of galectin-3 were assigned to the same cluster if they were identified as making at least one inter-molecular contact. We determined the number of clusters (Figure 12A) and the number of molecules forming the largest cluster (Figure 12B) as a function of time. Interestingly, we found both of these numbers to fluctuate between 1 and 6, indicating that galectin-3 could form fuzzy complexes [7]. We found the most likely number of clusters to be three (Figure 12C) and the largest cluster to comprise most typically three molecules of galectin-3 (Figure 12D). However, the probability of all six molecules of galectin-3 forming one cluster was found to be over  $\frac{1}{8}$  (Figure 12C,D).



**Figure 12.** (A) Number of clusters and (B) number of proteins forming the largest cluster as a function of time. (C,D) Corresponding histograms of the number of clusters and the number of proteins forming the largest clusters. Results obtained from the MD simulations of six molecules of galectin-3 in a cubic box of side length  $L = 20$  nm with  $\lambda = 1.03$ .

### 3. Methods

#### 3.1. Molecular Dynamics Simulations

A structural model of full-length galectin-3 was built based on (i) the crystal structure of the CRD deposited in the Protein Data Bank (PDB) with the entry code of 2NMO [13] and (ii) atomic coordinates of the NTD generated using the Molefacture Plugin in VMD [27]. The structural model of full-length galectin-3 was converted to the Martini 3 coarse-grained representation using Martinize2. Secondary structure elements within the CRD were

assigned using DSSP [28] in Martinize2. The NTD was assigned as a random coil. An elastic network was imposed on the CRD to preserve the domain structure in the course of the MD simulations. Specifically, Martinize2 was called with flags `-elastic -ef 700.0 -e1 0.5 -eu 0.9 -ea 0 -ep 0 -scfix`. In this way, harmonic restraints between backbone beads between the lower and upper distance cut-offs of 0.5 and 0.9 nm, respectively, were imposed. The spring constant of the harmonic restraints was set to  $700 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ . The harmonic restraints within the NTD, as well as those between the NTD and CRD, were subsequently removed from the topology file using a Python-3 script developed by Thomassen et al. [15]. In addition, by applying the flag `-scfix` in Martinize2, extra torsional potentials between backbone beads and side chain beads were added [29]. These potentials were included in the force field to prevent unphysical flipping of side chains in  $\beta$ -strands of the CRD.

The galectin-3 coarse-grained model was placed in a cubic box of side length  $L = 20 \text{ nm}$  using the Gromacs `editconf` tool. Then, the protein was solvated using Insane [30] with sodium and chloride ions added to neutralize the simulation system and to reach the salt concentration of 100 mM. Consequently, there were 489 sodium ions and 491 chloride ions in the simulation box. The energy of the system was minimized in 200 steps of the steepest descent algorithm implemented in Gromacs. The force field parameters describing interactions between the protein and water beads were re-scaled by a factor  $\lambda$  using a Python script provided by Thomassen et al. [15]. The values of parameter  $\lambda$  were set to 1, 1.01, ... 1.1, 1.12. Then, for each of the  $\lambda$ -values, an equilibration simulation of 10 ns was performed with the integration time step of 2 fs to allow for slow relaxation of the system.

The MD simulations were performed using Gromacs version 2021.5. Production runs were performed with the integration time step of 20 fs. Nonbonded interactions were treated with the Verlet cut-off scheme. The cut-off for van der Waals interactions was set to 1.1 nm. Coulomb interactions were treated using the reaction field method with a cut-off of 1.1 nm and dielectric constant of 15. The temperature and pressure were kept constant at  $T = 300 \text{ K}$  and  $p = 1 \text{ bar}$ , respectively, using the velocity-rescaling thermostat and the Parrinello–Rahman barostat. System configurations were saved at regular intervals of 1 ns. For most of the  $\lambda$ -values, the simulation trajectories were 20  $\mu\text{s}$  long. However, for  $\lambda = 1.03$ , the total simulation time was 100  $\mu\text{s}$  to obtain sufficient statistics for detailed analyses of amino acid residue contacts; see Section 3.2.

MD simulations of two and six molecules of galectin-3 in a cubic box of side length  $L = 20 \text{ nm}$  and with  $\lambda = 1.03$  were set up and carried out analogously. The flag `-merge` was used in Martinize2. The galectin-3 molecules were placed in the simulation box in random orientations using Packmol [31] with a tolerance of 0.4 nm and a maximum of 20 iterations. The protein concentration was 0.42 and 1.25 mM, respectively, in the simulations of two and six molecules of galectin-3. At each of the two concentrations of galectin-3, the MD time was 100  $\mu\text{s}$ .

### 3.2. Analysis of Molecular Dynamics Trajectories

Periodic boundary conditions were handled using the Gromacs tool `trjconv` with the flags `-pbc whole -center`. However, to analyze inter-molecular contacts in the MD trajectories with two and six molecules of galectin-3, we used the Python script `mdvwhole` [32] with the flags `-sel 'name BB SC1 SC2 SC3 SC4 SC5' -res -0.7` instead.

The radius of gyration ( $R_g$ ), the maximum diameter ( $D_{\text{max}}$ ) and the end-to-end distance ( $d_{\text{ee}}$ ) of galectin-3 were computed on the basis of the Cartesian coordinates of protein beads (BB, SC1, SC2, SC3, SC4 and SC5) using Python with the `MDTraj` library [33]. For a given conformation of galectin-3,  $D_{\text{max}}$  was computed as the largest of the distances between all pairs of protein beads, whereas  $d_{\text{ee}}$  was taken as the distance between the backbone beads of the amino acid residues at the N- and the C-terminus. Histograms of  $R_g$ ,  $D_{\text{max}}$  and  $d_{\text{ee}}$  were determined using Python with the `NumPy` library [34]. The average value of  $R_g$  was determined separately for each of the trajectories with specific  $\lambda$ -values. The statistical error of the average  $R_g$  was estimated using block error analysis [35].

The average inter-residue distance  $R_{ij}$  as a function of sequence separation  $|i - j|$  was determined from the Cartesian coordinates of the protein beads in the following way: For a given conformation of galectin-3, the distance  $d_{ij}$  between two amino acid residues with indices  $i$  and  $j$  was taken as the smallest of all the distances between the beads forming residue  $i$  and the beads forming residue  $j$ . The distances  $d_{ij}$  were computed using the function `compute_contacts` from the Python library `MDTraj`. The average inter-residue distance  $R_{ij}$  was then obtained by averaging  $d_{ij}$  over all conformations sampled in the MD simulation with a given  $\lambda$ -value and over all residue pairs  $i$  and  $j$  with a given sequence separation  $|i - j|$ .

For SAXS analysis, every 5th frame was extracted from the MD trajectory of one molecule of galectin-3 with  $\lambda = 1.03$ . The resulting  $N = 20,000$  conformations of galectin-3 (recorded evenly every 5 ns in the MD trajectory) were converted from the coarse-grained Martini representation to the all-atom representation using the Backward algorithm with default parameters [36]. The Python script `backward.py` was used to conduct the actual backmapping. The bash wrapper `initral-v5.sh` was used to perform the energy minimization and position-restraint MD simulations using Gromacs with the CHARMM36m force field [37].

The resulting  $N = 20,000$  structures of galectin-3 in the all-atom representation were used as input for SAXS analysis. Namely, the scattering intensity profile was computed for each of the structures individually using Pepsi-SAXS [38]. The ensemble-averaged scattering intensity profile was computed as

$$I_{\text{sim}}(q) = \frac{1}{N} \sum_{k=1}^N I_k(q)$$

where  $k = 1, \dots, N$  labels the galectin-3 structures in the ensemble,  $I_k(q)$  is the scattering intensity profile of the  $k$ th structure and  $q$  is the X-ray momentum transfer (i.e.,  $q = 4\pi \sin(\theta/2)/\lambda$ , where  $\theta$  and  $\lambda$  denote the scattering angle and the X-ray wavelength, respectively). The discrepancy between the SAXS experimental data,  $I_{\text{exp}}(q)$ , and the ensemble-averaged scattering intensity profile,  $I_{\text{sim}}(q)$ , was quantified by

$$\chi^2(a, b) = \frac{1}{n} \sum_{i=1}^n (I_{\text{exp}}(q_i) - a I_{\text{sim}}(q_i) - b)^2 / \sigma_{\text{exp}}^2(q_i)$$

where  $n = 221$  is the number of points in the experimental dataset, and  $\sigma(q)$  denotes the statistical error of the SAXS intensity  $I_{\text{exp}}(q)$ . The scale parameter  $a$  and the offset  $b$  were fitted to minimize  $\chi^2(a, b)$  using the Python library `SciPy` [39]. The calculations were performed with different contrasts of the hydration shell electron density  $\delta\rho$  ranging from 4% to 10% of the bulk water electron density  $\rho = 334 \text{ e/nm}^3$ . The best fit with  $\chi^2 = 1.68$  was obtained for  $\delta\rho = 0.075\rho \approx 25 \text{ e/nm}^3$ .

Contact maps were determined using the `ContactTrajectory` function from the Python library `Contact Map Explorer 0.7.0` [40]. A pair of amino acid residues was considered to be in contact if the distance between any bead in one of the residues and any bead in the other residue was smaller than 0.7 nm. Contacts between residue pairs  $(i, i + 1)$ ,  $(i, i + 2)$  and  $(i, i + 3)$  were excluded from the analysis. Consequently, the `ContactTrajectory` function was called with parameters `cutoff = 0.7` and `n_neighbors_ignored = 3`. The same criteria were used to determine intra- and inter-molecular contacts.

The probability of contact between two residues was defined as the number of frames in which the two residues were in contact divided by the total number of frames in the MD trajectory, which was  $10^5$  for the simulations with  $\lambda = 1.03$ . All analyses involving the probability of contact occurrence were performed using the `ContactFrequency` function from the Python library `Contact Map Explorer 0.7.0` with parameters `cutoff = 0.7` and `n_neighbors_ignored = 3`. The statistical error of the probability of contact occurrence was estimated using block error analysis [35].

To compute the probability of contact between a given residue and a specific domain, the NTD was assumed to comprise amino acid residues with indices from 1 to 112. Consequently, the boundaries of the CRD were taken to be amino acid residues with indices 113 and 250. To determine the probability of contact formed by a specific  $\beta$ -strand in the CRD, the  $\beta$ -strands were taken to comprise amino acid residues with the following indices: from 118 to 121 ( $\beta_1$ ), from 130 to 138 ( $\beta_2$ ), from 145 to 151 ( $\beta_3$ ), from 154 to 162 ( $\beta_4$ ), from 170 to 174 ( $\beta_5$ ), from 185 to 187 ( $\beta_6$ ), from 197 to 204 ( $\beta_7$ ), from 208 to 213 ( $\beta_8$ ), from 216 to 222 ( $\beta_9$ ), from 233 to 238 ( $\beta_{10}$ ) and from 240 to 249 ( $\beta_{11}$ ).

To identify oligomers formed in the MD simulation with six molecules of galectin-3 in a cubic box of side length  $L = 20$  nm with  $\lambda = 1.03$ , agglomerative clustering was applied to each of the configurations obtained from the MD simulation [41]. At the beginning of the clustering procedure, each of the galectin-3 molecules was taken as a separate cluster. Then, clusters were merged successively. Two clusters were merged if any molecule from one cluster was identified to make at least one contact with any molecule from the other cluster. The clustering procedure was stopped when no more clusters could be merged. This clustering procedure was implemented in Python and applied successively to each of the molecular coarse-grained configurations recorded in the MD simulation.

#### 4. Discussion

SAXS is a useful and widely used method to characterize macromolecules in solution [42,43]. Standard analysis of SAXS data of a globular protein yields the values of  $R_g$  and  $D_{max}$ , as well as the so-called molecular envelope that represents the overall shape of the protein [44]. However, SAXS analysis of proteins containing intrinsically disordered regions requires different approaches based on generating an ensemble of protein conformations [45–48]. In particular, a combination of SAXS experiments and coarse-grained molecular simulations has proven very useful in delineating conformations of proteins comprising both folded domains and intrinsically disordered regions [49–56]. Here, we followed this approach and used coarse-grained MD simulations together with available SAXS data to explore the conformational ensemble of full-length galectin-3.

Galectin-3 has been reported to be involved in numerous intra- and extra-cellular processes [10]. In particular, galectin-3 has been demonstrated to drive glycosphingolipid-dependent biogenesis of clathrin-independent carriers [57]. Although galectin-3 oligomerization has been implicated in this process, the molecular mechanisms underlying the endocytic pit formation remain elusive. However, the discovery that galectin-3 forms biomolecular condensates sheds new light on how this protein can perform its biological functions at the cell membrane [21]. Recently, we devised a mesoscopic model with dissipative particle dynamics (DPD) to simulate biomolecular condensates of galectin-3 and their interactions with lipid membranes [41]. The DPD simulations reveal the mesoscopic mechanisms by which the biomolecular condensates can sense and generate membrane curvature but do not provide insights into molecular details of these mechanisms. In the current study, in contrast, we used MD simulations with the Martini 3 force field to explore the conformations of full-length galectin-3 in solution. We slightly altered the magnitude of protein–water interactions in the the force field to obtain an ensemble of galectin-3 conformations in good quantitative agreement with available data from SAXS experiments [7] (Figure 1). We quantified the conformational ensemble in terms of protein size, flexibility and intra-molecular contacts (Figures 4–6). Interestingly, we found that the conformational ensemble at infinite dilution—as obtained from the MD simulations of a single molecule of galectin-3—was barely distinguishable from the conformational ensemble at the protein concentration of 400  $\mu$ M, i.e., obtained from the MD simulations of two molecules of galectin-3 in the cubic box of side length  $L = 20$  nm (Figures 7 and 8) This result indicates that the protein concentration range between 40 and 400  $\mu$ M used in the NMR experiments [7] is appropriate for telling apart the intra- and inter-molecular interactions.

Dignon et al. have identified a strong correlation between the critical temperature and the  $\theta$ -point temperature of IDPs [22]. Our simulation results show that the galectin-3

SAXS experimental conditions are very close to the  $\theta$ -point conditions (Figures 2B and 3A) and that galectin-3 molecules associated only transiently form to diverse oligomers at elevated concentrations (Figure 12), indicating together that galectin-3 does not aggregate or condensate at these conditions. As a matter of fact, galectin-3 has been shown to undergo liquid–liquid phase separation only at relatively high salt concentrations [21]. Future MD studies may help to elucidate this phenomenon and delineate the structure of galectin-3 condensates.

## 5. Conclusions

Our detailed analysis of the MD trajectory with one molecule of galectin-3 shows that (i) full-length galectin-3 exhibits large-scale fluctuations between diverse compact and extended conformations (Figures 4 and 5); (ii) the NTD is practically as flexible as a Gaussian chain (Figure 3A), and virtually any pair of amino acid residues in the NTD makes transient contacts (Figure 5B); (iii) aromatic residues in the NTD make the most frequent intra-molecular contacts—both with the CRD and within the NTD (Figure 5); (iv) many different residues in the NTD make transient contacts with single regions on the CRD surface (Figure 5C), which is a characteristic of fuzzy interactions and has been identified in NMR experiments [7]; and (v) the convex face of the CRD comes into contact with the NTD significantly more often than the concave face of the CRD (Figures 5 and 6), which appears to be consistent with NMR chemical shifts for residues 200–220 located on the convex side of the CRD [7]. On the one hand, such a detailed picture of the conformational ensemble of full-length galectin-3 cannot be at present obtained from experiments alone. On the other hand, our simulation results are not only in good quantitative agreement with the data from SAXS experiments ( $\chi^2 = 1.68$ , Figure 1B), but also consistent with evidence from the NMR experiments [7].

Our analysis of the MD trajectories with two and six molecules of galectin-3 shows additionally that the  $\beta$ -strands  $\beta_4$ ,  $\beta_5$  and  $\beta_6$  (forming the carbohydrate-binding site on the CRD surface) barely ever make intra-molecular contacts with the NTD (Figure 6) and virtually no inter-molecular contacts at the protein concentration of 400  $\mu$ M (Figures 10 and 11). These results mean that neither inter- or intra-molecular interactions occlude the carbohydrate-binding site on the CRD surface, which potentially has biological implications. In addition, consistent with experimental evidence [7], the MD simulations show that interactions between galectin-3 molecules are mediated mostly by the NTDs (Figures 9–11).

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/molecules29122768/s1>: Figure S1: Data supporting Figure 8, Figure S2: Data supporting Figure 9, Figure S3: Snapshots from subsequent quarters of the MD trajectory of one molecule of galectin-3 with  $\lambda = 1.03$ , Figure S4: Snapshots from subsequent quarters of the MD trajectory of two molecules of galectin-3, Figure S5: Snapshots from subsequent quarters of the MD trajectory of six molecules of galectin-3.

**Author Contributions:** M.M.A.: data curation, formal analysis, investigation, methodology, validation, writing—original draft preparation; P.R.: conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing—original draft preparation; B.R.: conceptualization, formal analysis, funding acquisition, methodology, project administration, resources, supervision, validation, writing—original draft preparation, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research has been funded in whole by the National Science Centre of Poland via grant number 2020/39/B/NZ1/00377.

**Data Availability Statement:** The original contributions presented in the study are included in the article and supplementary material. Further inquiries can be directed to the corresponding author.

**Acknowledgments:** We thank Jie-rong Huang from the Institute of Biochemistry and Molecular Biology in Taiwan for sharing the SAXS data. We also thank the Centre of Informatics—Tricity Academic Supercomputer and Network (CI TASK) in Gdańsk, Poland, for access to computer

resources. We acknowledge the Polish high-performance computing infrastructure PLGrid for awarding this project access to the LUMI supercomputer, owned by the EuroHPC Joint Undertaking, hosted by CSC (Finland) and the LUMI consortium through PLL/2023/04/016485. For the purpose of open access, the authors have applied a CC-BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Hirabayashi, J.; Kasai, K.-I. The family of metazoan metal-independent  $\beta$ -galactoside-binding lectins: Structure, function and molecular evolution. *Glycobiology* **1993**, *3*, 297–304. [[CrossRef](#)]
2. Barondes, S.H.; Castronovo, V.; Cooper, D.; Cummings, R.D.; Drickamer, K.; Feizi, T.; Gitt, M.A.; Hirabayashi, J.; Hughes, C.; Kasai, K.-I. Galectins: A family of animal beta-galactoside-binding lectins. *Cell* **1994**, *76*, 597–598. [[CrossRef](#)] [[PubMed](#)]
3. Bänfer, S.; Jacob, R. Galectins in intra- and extracellular vesicles. *Biomolecules* **2020**, *10*, 1232. [[CrossRef](#)] [[PubMed](#)]
4. Johannes, L.; Jacob, R.; Leffler, H. Galectins at a glance. *J. Cell Sci.* **2018**, *131*, jcs208884. [[CrossRef](#)] [[PubMed](#)]
5. Lepur, A.; Salomonsson, E.; Nilsson, U.J.; Leffler, H. Ligand induced galectin-3 protein self-association. *J. Biol. Chem.* **2012**, *287*, 21751–21756. [[CrossRef](#)]
6. Ippel, H.; Miller, M.C.; Vértesy, S.; Zheng, Y.; Cañada, F.J.; Suylen, D.; Umemoto, K.; Romanò, C.; Hackeng, T.; Tai, G.; et al. Intra- and intermolecular interactions of human galectin-3: Assessment by full-assignment-based NMR. *Glycobiology* **2016**, *26*, 888–903. [[CrossRef](#)] [[PubMed](#)]
7. Lin, Y.H.; Qiu, D.C.; Chang, W.H.; Yeh, Y.Q.; Jeng, U.S.; Liu, F.T.; Huang, J.R. The intrinsically disordered N-terminal domain of galectin-3 dynamically mediates multisite self-association of the protein through fuzzy interactions. *J. Biol. Chem.* **2017**, *292*, 17845–17856. [[CrossRef](#)] [[PubMed](#)]
8. Funasaka, T.; Raz, A.; Nangia-Makker, P. Nuclear transport of galectin-3 and its therapeutic implications. *Semin. Cancer Biol.* **2014**, *27*, 30–38. [[CrossRef](#)] [[PubMed](#)]
9. Dumic, J.; Dabelic, S.; Flögel, M. Galectin-3: An open-ended story. *Biochim. Biophys. Acta-(Bba)-Gen. Subj.* **2006**, *1760*, 616–635. [[CrossRef](#)]
10. Newlaczyl, A.U.; Yu, L.G. Galectin-3—A jack-of-all-trades in cancer. *Cancer Lett.* **2011**, *313*, 123–128. [[CrossRef](#)]
11. Dong, R.; Zhang, M.; Hu, Q.; Zheng, S.; Soh, A.; Zheng, Y.; Yuan, H. Galectin-3 as a novel biomarker for disease diagnosis and a target for therapy. *Int. J. Mol. Med.* **2018**, *41*, 599–614. [[CrossRef](#)] [[PubMed](#)]
12. Seetharaman, J.; Kanigsberg, A.; Slaaby, R.; Leffler, H.; Barondes, S.H.; Rini, J.M. X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1-Å resolution. *J. Biol. Chem.* **1998**, *273*, 13047–13052. [[CrossRef](#)] [[PubMed](#)]
13. Collins, P.M.; Hidari, K.I.; Blanchard, H. Slow diffusion of lactose out of galectin-3 crystals monitored by X-ray crystallography: Possible implications for ligand-exchange protocols. *Acta Crystallogr. Sect. Biol. Crystallogr.* **2007**, *63*, 415–419. [[CrossRef](#)] [[PubMed](#)]
14. Souza, P.C.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B.M.; Wassenaar, T.A.; et al. Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382–388. [[CrossRef](#)] [[PubMed](#)]
15. Thomasen, F.E.; Pesce, F.; Roesgaard, M.A.; Tesei, G.; Lindorff-Larsen, K. Improving Martini 3 for disordered and multidomain proteins. *J. Chem. Theory Comput.* **2022**, *18*, 2033–2041. [[CrossRef](#)] [[PubMed](#)]
16. Koneru, J.K.; Sinha, S.; Mondal, J. Molecular dynamics simulations elucidate oligosaccharide recognition pathways by galectin-3 at atomic resolution. *J. Biol. Chem.* **2021**, *297*, 101271. [[CrossRef](#)] [[PubMed](#)]
17. Vander Zanden, C.M.; Majewski, J.; Weissbarth, Y.; Browne, D.F.; Watkins, E.B.; Gabius, H.J. Structure of Galectin-3 bound to a model membrane containing ganglioside GM1. *Biophys. J.* **2023**, *122*, 1926–1937. [[CrossRef](#)] [[PubMed](#)]
18. Bhattacharya, S.; Zhang, M.; Hu, W.; Qi, T.; Heisterkamp, N. Targeting disordered-structured domain interactions in Galectin-3 based on NMR and enhanced MD. *Biophys. J.* **2022**, *121*, 4342–4357. [[CrossRef](#)] [[PubMed](#)]
19. Robustelli, P.; Piana, S.; Shaw, D.E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4758–E4766. [[CrossRef](#)]
20. Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A.E.; Kolinski, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016**, *116*, 7898–7936. [[CrossRef](#)]
21. Chiu, Y.P.; Sun, Y.C.; Qiu, D.C.; Lin, Y.H.; Chen, Y.Q.; Kuo, J.C.; Huang, J.R. Liquid-liquid phase separation and extracellular multivalent interactions in the tale of galectin-3. *Nat. Commun.* **2020**, *11*, 1229. [[CrossRef](#)] [[PubMed](#)]
22. Dignon, G.L.; Zheng, W.; Best, R.B.; Kim, Y.C.; Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9929–9934. [[CrossRef](#)] [[PubMed](#)]
23. Hofmann, H.; Soranno, A.; Borgia, A.; Gast, K.; Nettels, D.; Schuler, B. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 16155–16160. [[CrossRef](#)] [[PubMed](#)]
24. Zheng, W.; Zerze, G.; Borgia, A.; Mittal, J.; Schuler, B.; Best, R.B. Inferring properties of disordered chains from FRET transfer efficiencies. *Biophys. J.* **2018**, *114*, 367a. [[CrossRef](#)]



25. Mazur, J. Distribution function of the end-to-end distances of linear polymers with excluded volume effects. *J. Res. Natl. Bur. Stand. Sect. Phys. Chem.* **1965**, *69*, 355–363. [[CrossRef](#)] [[PubMed](#)]
26. Farhadi, S.A.; Liu, R.; Becker, M.W.; Phelps, E.A.; Hudalla, G.A. Physical tuning of galectin-3 signaling. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2024117118. [[CrossRef](#)] [[PubMed](#)]
27. Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
28. Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolym. Orig. Res. Biomol.* **1983**, *22*, 2577–2637. [[CrossRef](#)] [[PubMed](#)]
29. Herzog, F.A.; Braun, L.; Schoen, I.; Vogel, V. Improved side chain dynamics in MARTINI simulations of protein—Lipid interfaces. *J. Chem. Theory Comput.* **2016**, *12*, 2446–2458. [[CrossRef](#)]
30. Wassenaar, T.A.; Ingólfsson, H.I.; Bockmann, R.A.; Tieleman, D.P.; Marrink, S.J. Computational lipidomics with insane: A versatile tool for generating custom membranes for molecular simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2144–2155. [[CrossRef](#)]
31. Martínez, L.; Andrade, R.; Birgin, E.G.; Martínez, J.M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164. [[CrossRef](#)] [[PubMed](#)]
32. Bruininks, B.M.H.; Wassenaar, T.A.; Vattulainen, I. Unbreaking Assemblies in Molecular Simulations with Periodic Boundaries. *J. Chem. Inf. Model.* **2023**, *63*, 3448–3452. [[CrossRef](#)]
33. McGibbon, R.T.; Beauchamp, K.A.; Harrigan, M.P.; Klein, C.; Swails, J.M.; Hernández, C.X.; Schwantes, C.R.; Wang, L.P.; Lane, T.J.; Pande, V.S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532. [[CrossRef](#)] [[PubMed](#)]
34. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)]
35. Flyvbjerg, H.; Petersen, H.G. Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91*, 461–466. [[CrossRef](#)]
36. Wassenaar, T.A.; Pluhackova, K.; Bockmann, R.A.; Marrink, S.J.; Tieleman, D.P. Going backward: A flexible geometric approach to reverse transformation from coarse grained to atomistic models. *J. Chem. Theory Comput.* **2014**, *10*, 676–690. [[CrossRef](#)] [[PubMed](#)]
37. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B.L.; Grubmüller, H.; MacKerell, A.D., Jr. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71–73. [[CrossRef](#)] [[PubMed](#)]
38. Grudinin, S.; Garkavenko, M.; Kazennov, A. Pepsi-SAXS: An adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles. *Acta Crystallogr. Sect. Struct. Biol.* **2017**, *73*, 449–464. [[CrossRef](#)] [[PubMed](#)]
39. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272.
40. Swenson, D.W.H.; Revision, S.R. Contact Map Explorer. Available online: [https://github.com/dwhswenson/contact\\_map](https://github.com/dwhswenson/contact_map) (accessed on 25 April 2024).
41. Anila, M.M.; Ghosh, R.; Różycki, B. Membrane curvature sensing by model biomolecular condensates. *Soft Matter* **2023**, *19*, 3723–3732. [[CrossRef](#)]
42. Brosey, C.A.; Tainer, J.A. Evolving SAXS versatility: Solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. *Curr. Opin. Struct. Biol.* **2019**, *58*, 197–213. [[CrossRef](#)] [[PubMed](#)]
43. Gräwert, T.W.; Svergun, D.I. Structural modeling using solution small-angle X-ray scattering (SAXS). *J. Mol. Biol.* **2020**, *432*, 3078–3092. [[CrossRef](#)] [[PubMed](#)]
44. Petoukhov, M.V.; Svergun, D.I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **2005**, *89*, 1237–1250. [[CrossRef](#)] [[PubMed](#)]
45. Pelikan, M.; Hura, G.L.; Hammel, M. Structure and flexibility within proteins as identified through small angle X-ray scattering. *Gen. Physiol. Biophys.* **2009**, *28*, 174–189. [[CrossRef](#)] [[PubMed](#)]
46. Różycki, B.; Kim, Y.C.; Hummer, G. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* **2011**, *19*, 109–116. [[CrossRef](#)] [[PubMed](#)]
47. Tria, G.; Mertens, H.D.; Kachala, M.; Svergun, D.I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *Int. Union Crystallogr. J.* **2015**, *2*, 207–217. [[CrossRef](#)] [[PubMed](#)]
48. Cordeiro, T.N.; Herranz-Trillo, F.; Urbanek, A.; Estaña, A.; Cortés, J.; Sibille, N.; Bernadó, P. Small-angle scattering studies of intrinsically disordered proteins and their complexes. *Curr. Opin. Struct. Biol.* **2017**, *42*, 15–23. [[CrossRef](#)] [[PubMed](#)]
49. Różycki, B.; Boura, E. Large, dynamic, multi-protein complexes: A challenge for structural biology. *J. Phys. Condens. Matter* **2014**, *26*, 463103. [[CrossRef](#)] [[PubMed](#)]
50. Różycki, B.; Cieplak, M.; Czjzek, M. Large conformational fluctuations of the multi-domain xylanase Z of *Clostridium thermocellum*. *J. Struct. Biol.* **2015**, *191*, 68–75. [[CrossRef](#)]
51. Chalupska, D.; Eisenreichova, A.; Różycki, B.; Rezabkova, L.; Humpolickova, J.; Klima, M.; Boura, E. Structural analysis of phosphatidylinositol 4-kinase III $\beta$  (PI4KB)–14-3-3 protein complex reveals internal flexibility and explains 14-3-3 mediated protection from degradation in vitro. *J. Struct. Biol.* **2017**, *200*, 36–44. [[CrossRef](#)]
52. Chalupska, D.; Różycki, B.; Klima, M.; Boura, E. Structural insights into Acyl-coenzyme A binding domain containing 3 (ACBD3) protein hijacking by picornaviruses. *Protein Sci.* **2019**, *28*, 2073–2079. [[CrossRef](#)] [[PubMed](#)]
53. Chalupska, D.; Różycki, B.; Humpolickova, J.; Faltova, L.; Klima, M.; Boura, E. Phosphatidylinositol 4-kinase III $\beta$  (PI4KB) forms highly flexible heterocomplexes that include ACBD3, 14-3-3, and Rab11 proteins. *Sci. Rep.* **2019**, *9*, 567. [[CrossRef](#)] [[PubMed](#)]

54. Sicorello, A.; Różycki, B.; Konarev, P.V.; Svergun, D.I.; Pastore, A. Capturing the conformational ensemble of the mixed folded polyglutamine protein ataxin-3. *Structure* **2021**, *29*, 70–81. [[CrossRef](#)] [[PubMed](#)]
55. Dorival, J.; Morais, S.; Labourel, A.; Różycki, B.; Cazade, P.A.; Dabin, J.; Setter-Lamed, E.; Mizrahi, I.; Thompson, D.; Thureau, A.; et al. Mapping the deformability of natural and designed cellulosomes in solution. *Biotechnol. Biofuels Bioprod.* **2022**, *15*, 68. [[CrossRef](#)] [[PubMed](#)]
56. Różycki, B.; Boura, E. Conformational ensemble of the full-length SARS-CoV-2 nucleocapsid (N) protein based on molecular simulations and SAXS data. *Biophys. Chem.* **2022**, *288*, 106843. [[CrossRef](#)]
57. Lakshminarayan, R.; Wunder, C.; Becken, U.; Howes, M.T.; Benzing, C.; Arumugam, S.; Sales, S.; Ariotti, N.; Chambon, V.; Lamaze, C.; et al. Galectin-3 drives glycosphingolipid-dependent biogenesis of clathrin-independent carriers. *Nat. Cell Biol.* **2014**, *16*, 592–603. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.