

RESEARCH ARTICLE

Decoding the mechanism governing the structural stability of wheat germ agglutinin and its isolated domains: A combined calorimetric, NMR, and MD simulation study

Jorge Luis Medrano-Cerano¹ | Luis Fernando Cofas-Vargas²  |
Eduardo Leyva¹ | Jesús Antonio Rauda-Ceja¹ | Mateo Calderón-Vargas¹ |
Patricia Cano-Sánchez¹ | Gustavo Titau-Delgado¹ |
Carolina Monserrath Melchor-Meneses¹ | Andrés Hernández-Arana³ |
Federico del Río-Portilla¹  | Enrique García-Hernández¹ 

¹Universidad Nacional Autónoma de México, Instituto de Química, Ciudad Universitaria, Ciudad de México, Mexico

²Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland

³Área de Biofísicoquímica, Departamento de Química, Universidad Autónoma Metropolitana Iztapalapa, Ciudad de México, Mexico

Correspondence

Enrique García-Hernández and Federico del Río-Portilla, Universidad Nacional Autónoma de México, Instituto de Química, Ciudad Universitaria, Ciudad de México, 04510, Mexico.
Email: egarciah@unam.mx and jfrp@unam.mx

Funding information

Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México, Grant/Award Numbers: IN206221, IN214722, IN209624

Review Editor: Nir Ben-Tal

Abstract

Wheat germ agglutinin (WGA) demonstrates potential as an oral delivery agent owing to its selective binding to carbohydrates and its capacity to traverse biological membranes. In this study, we employed differential scanning calorimetry and molecular dynamics simulations to comprehensively characterize the thermal unfolding process of both the complete lectin and its four isolated domains. Furthermore, we present the nuclear magnetic resonance structures of three domains that were previously lacking experimental structures in their isolated forms. Our results provide a collective understanding of the energetic and structural factors governing the intricate unfolding mechanism of the complete agglutinin, shedding light on the specific role played by each domain in this process. The analysis revealed negligible interdomain cooperativity, highlighting instead significant coupling between dimer dissociation and the unfolding of the more labile domains. By comparing the dominant interactions, we rationalized the stability differences among the domains. Understanding the structural stability of WGA opens avenues for enhanced drug delivery strategies, underscoring its potential as a promising carrier throughout the gastrointestinal environment.

KEYWORDS

homodimer, hydrogen bonding, lectin, multidomain protein, structural stability, thermal unfolding

Jorge Luis Medrano-Cerano, Luis Fernando Cofas-Vargas, and Eduardo Leyva contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

1 | INTRODUCTION

Lectins participate in a wide range of highly specific recognition events by selectively attaching to carbohydrates and glycoconjugates. These interactions are crucial for myriads of biological phenomena, including self-recognition, cell-to-cell communication, pathogen invasion, and neutralization of invasive organisms (Gabiús, 2018; Solís et al., 2015). One protein of particular significance within this group is wheat germ agglutinin (WGA), which has played a pivotal role in uncovering abnormal glycosylation patterns in cancerous cells (Aub et al., 1965), a breakthrough that opened the door to the molecular identification of malignant cells. Subsequently, WGA has been widely employed for cellular labeling and isolation, and for investigating the significance of glycosylation in cellular adhesion, migration, and signaling (Balčiūnaitė-Murzienė and Dzikaras, 2021). Fluorescent probes covalently linked to this lectin facilitate the visualization of individual cells and the resolution of expression patterns of glycan structures. Moreover, WGA proves useful for the identification of specific bacterial and viral infections, exhibiting potential in the field of medical diagnostics (Raghu and Kumar, 2020; Tu et al., 2022). The conjugation of drug-like compounds with WGA allows the specific targeting of malignant cells or tissues that exhibit a high abundance of WGA-binding carbohydrates (Apfelthaler et al., 2017; Chiu and Lim, 2021). This has the potential to improve the efficacy of treatments while reducing the occurrence of unpleasant responses (Hädrich et al., 2022; Liu et al., 2016). The utilization of WGA-coated surfaces enables the immobilization of cells or proteins, hence facilitating the advancement of biochip production and other analytical methodologies. The remarkable ability of WGA to cross several epithelial barriers, including the blood–brain barrier, has garnered significant attention because of its potential in enhancing oral medication administration (Balčiūnaitė-Murzienė and Dzikaras, 2021; Wang et al., 2023).

WGA recognizes N-acetylglucosamine, N-acetylneuraminic acid, and their derived oligosaccharides (Wright, 1992). It functions by selectively attaching to carbohydrates present on the surfaces of bacteria, fungi, and the exoskeleton of insects, contributing to the neutralization of the offensive activities exerted by these organisms. It forms a non-obligate head-to-tail weak homodimer at neutral pH (dissociation constant (K_d) $\sim 20 \mu\text{M}$), which reversibly dissociates under acidic conditions (Chavelas et al., 2004; Portillo-Télez et al., 2011). The WGA subunit has a modular structure consisting of four isostructural hevein-like domains (here referenced as WGA.A to WGA.D), with four disulfide bridges per domain that are essential for the protein's stability (Jiménez-Barbero et al., 2006; Slavokhotova

et al., 2017). Eight carbohydrate-binding sites are located along the dimer interface. The modular structure of WGA is most probably the result of the duplication and fusion of hevein-like genes, one of the smallest lectins known (García-Hernández et al., 1997). Because of the independent carbohydrate-binding capabilities exhibited by each of its four domains (Espinosa et al., 2000; Leyva et al., 2019), the monomerized subunit of WGA can bind two or more glycoconjugates simultaneously. This multivalent binding ability is essential for the functioning of many lectins, especially those that contribute to the innate defense against pathogens and other aggressor organisms via cell agglutination. Presumably, this is why WGA has not undergone intense evolutionary pressure to form a robust homooligomer, with its dimeric form coexisting in solution with native-like folded monomers. Self-association, on the other hand, represents a significant functional improvement for this agglutinin because it not only provides additional stability to the active conformation, but also creates expanded binding sites composed of residues from both subunits, which increases both binding affinity and the potential for multivalent interaction with glycoconjugates (Schwefel et al., 2010). On the contrary, because of the presence of only a single recognition site per subunit, the majority of agglutinins have evolved to form strong homooligomers in order to fulfill their functional roles ($K_d = 10^{-8}$ – 10^{-41} M^{-1}) (Portillo-Télez et al., 2011).

To improve the efficacy of WGA as a potential agent for peroral delivery, it is crucial to have a thorough understanding of its structural stability and ability to withstand environmental changes, such as the fluctuating acidity it meets throughout its passage in the digestive tract. The multistep thermal unfolding of WGA combines dimer dissociation and independent domain unfolding events in a complex manner (Leyva et al., 2019; Portillo-Télez et al., 2011). To shed new light on this intricate behavior, in this study we have performed a characterization of the thermal unfolding of the complete protein and of each of its isolated domains using differential scanning calorimetry (DSC) and molecular dynamics (MD) simulations. In addition, we report for the first time the NMR structures of isolated WGA.A, WGA.C, and WGA.D (referred to as $_{\text{isol}}$ WGA.X, where X corresponds to each domain). The aforementioned structures, in conjunction with the previously reported $_{\text{isol}}$ WGA.B structure (Espinosa et al., 2000), complete the picture that each WGA domain functions as a cooperative folding unit (Porter and Rose, 2012). The results given in this study provide insights into the underlying factors that contribute to the variations in stability observed between the four WGA domains. These findings hold significant implications for utilizing WGA as a drug carrier across the gastrointestinal tract, particularly concerning the crucial aspect of structural stability.

2 | RESULTS

2.1 | Thermal stability of wild-type WGA and its isolated recombinant domains

The thermal unfolding of wild-type WGA (purified from wheat germs) and its four recombinant domains (produced in *E. coli*) was characterized using DSC (Figure 1). Measurements were conducted at pH 4.7, a condition where a spectroscopic characterization of the homodimer and the recombinant domains has been reported elsewhere (Leyva et al., 2019). CD spectra confirmed the

structural integrity of the proteins at room temperature (data not shown). The calorimetric traces obtained from the isolated domains exhibited a single peak (Figure 1a). Based on the area recovered (>97%) in a temperature rescanning, a high degree of reversibility was observed in all the measurements conducted. The unfolding of each domain was consistent with a two-state process, based on the calorimetric criterion ($\Delta H_{\text{vH}}/\Delta H_{\text{cal}} \sim 1$) and a fit to a single-transition model. The order of thermoresistance was $\text{isolWGA.D} < \text{isolWGA.A} < \text{isolWGA.C} < \text{isolWGA.B}$. Notably, there was a large T_m difference ($\sim 25^\circ\text{C}$) between the most labile and the most resistant domains (Table 1).

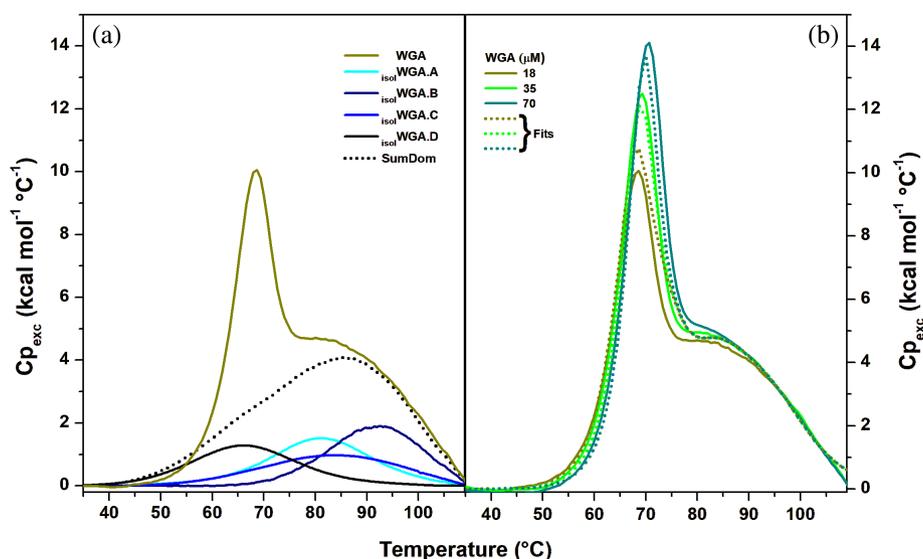


FIGURE 1 DSC traces of WGA and its isolated domains. (a) Endotherms for isolated domains and wild-type WGA. The solid lines correspond to experimental traces. The dotted line is the sum of the four isolated domain endotherms. Domain endotherms were satisfactorily fitted to a two-state unfolding model (fittings not shown). Domain concentrations ranged between 0.1 and 0.4 mM. (b) Wild-type WGA endotherms at different monomer equivalent concentrations. An unfolding model that included dimer dissociation and four independent two-state transitions (dotted lines) was simultaneously best fitted to the three experimental endotherms. All measurements were performed at pH 4.7, in a 30 mM acetate buffer. DSC, differential scanning calorimetry; WGA, wheat germ agglutinin.

TABLE 1 Calorimetrically determined stability parameters of WGA and the isolated domains.

Isolated domains				Complete WGA ^c		
Domain ^a	T_m (°C)	ΔH_{cal} (kcal/mol)	ΔH_{vH}^b (kcal/mol)	Transition ^d	T_{m_i} (°C)	ΔH_i (kcal/mol)
isolWGA.D	66.5 ± 0.0	34.3 ± 0.2	34.4	1	59.2 ± 3.2	53.2 ± 2.1
isolWGA.A	81.2 ± 0.1	38.8 ± 0.1	38.8	2	73.6 ± 0.3	70.8 ± 1.7
isolWGA.C	83.8 ± 0.1	32.1 ± 0.1	30.6	3	83.8 ± 0.5	57.1 ± 1.0
isolWGA.B	91.8 ± 0.0	45.2 ± 0.1	45.2	4	93.4 ± 0.4	50.1 ± 1.4

Abbreviation: WGA, wheat germ agglutinin.

^aDomains are arranged in ascending order according to their melting temperatures.

^bvan't Hoff enthalpy change, calculated from DSC endotherms as $4RT_{1/2}^2 C_{\text{exc } 1/2} / \Delta H_{\text{cal}}$, where $T_{1/2}$ and $C_{\text{exc } 1/2}$ are the temperature and the excess heat capacity at half transition, and ΔH_{cal} is the total endotherm area.

^cBest fitted parameters resulting from the simultaneous analysis of the wild-type WGA endotherms at protein concentrations of 18, 35, and 70 μM, according to an unfolding model that included dimer dissociation and four independent two-state transitions, as described in Appendix S1.

^dDeconvoluted transitions of complete WGA arranged in ascending order according to their melting temperatures.

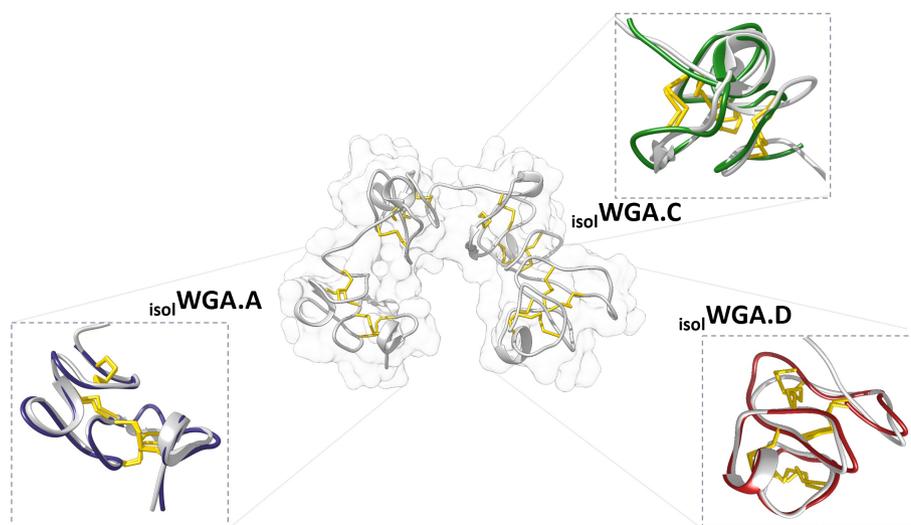


FIGURE 2 NMR structures of isolated WGA domains. Alignment of the NMR-derived best-energy conformers with the corresponding domain structures in the complete WGA crystal (PDB ID: 2uvo). WGA, wheat germ agglutinin.

The calorimetric profiles of the complete protein exhibited two peaks (Figure 1b). The lowest temperature peak was markedly dependent on protein concentration. In contrast, the second peak remained nearly constant. Hence, the first peak was coupled to the dissociation of the dimer, whereas the other one was mostly linked to the unfolding of the monomerized subunits (Portillo-Télez et al., 2011). It is worth mentioning that an endotherm resulting from the cumulative calorimetric signals of the isolated domains exhibited considerable overlap with the high temperature peak observed in the WGA traces (Figure 1a). This agreement further supports the picture that each domain represents a cooperative folding unit, and that their unfolding within the same subunit can be described as an event with minimal interdomain cooperativity (Portillo-Télez et al., 2011). However, it should be recalled that the concentration dependence of the dissociation event can also affect subsequent unfolding transitions of individual domains, thus the unfolding transitions of one or more domains could significantly contribute to the area of the first peak in the DSC curves.

The DSC traces of wild-type WGA dimer at three concentrations were analyzed globally (i.e., with shared domain unfolding enthalpy and T_m values) according to the model described in Appendix S1 and Data S1 (Figure 1b). In agreement with previous calorimetric dissociation measurements (Portillo-Télez et al., 2011), the fitted dissociation enthalpy resulted in a modest value of 0.22 ± 0.02 kcal/mol. Fitted C_p exc tracings explained reasonably well both the overall displacement to higher temperatures and the enthalpy increase in the low-temperature peak in the experimental curves as WGA concentration increased. The endotherms were consistent with the occurrence of four two-state transitions. Encouraging as these results are, it is worth to note that the fitted values differed from those of the isolated domains

(Table 1). These differences minimized between $_{\text{isol}}\text{WGA.B}$ and the highest-temperature transition for the wild-type protein, indicating that the dimer imparts a significant energetic effect on the more labile domains.

2.2 | NMR determined structures of isolated domains

The solution structures of $_{\text{isol}}\text{WGA.A}$, $_{\text{isol}}\text{WGA.C}$, and $_{\text{isol}}\text{WGA.D}$ were deposited in the PDB under the PDB IDs 8vu6, 8vu7, and 8vu8, respectively (Figure 2 and Figure SM1 in Data S1). The three domains exhibited a clearly defined hevein-like structural motif, featuring eight cysteines appropriately positioned to form four disulfide bridges. Each domain displayed a close conformational overlap with the crystal structure of the entire agglutinin (Figure 2), with pairwise backbone root-mean-square deviation (RMSD) for residues 4–41, that is, excluding the flexible regions from both the N- and C-terminal regions, around 1 Å. Root-mean-square fluctuation (RMSF) calculations among the NMR conformers of each domain revealed that $_{\text{isol}}\text{WGA.D}$ exhibited higher conformational flexibility (0.6 ± 0.2 Å) compared to $_{\text{isol}}\text{WGA.A}$ (0.3 ± 0.1 Å) or $_{\text{isol}}\text{WGA.C}$ (0.4 ± 0.1 Å) (Figure SM1 in Data S1).

2.3 | Structural dynamics of WGA domains characterized by MD simulations

To investigate the factors influencing the varying thermal stabilities of the WGA domains, MD simulations were conducted at 20 and 120°C. These temperatures were selected based on the preceding experimental findings, which indicated that the folded state prevails at

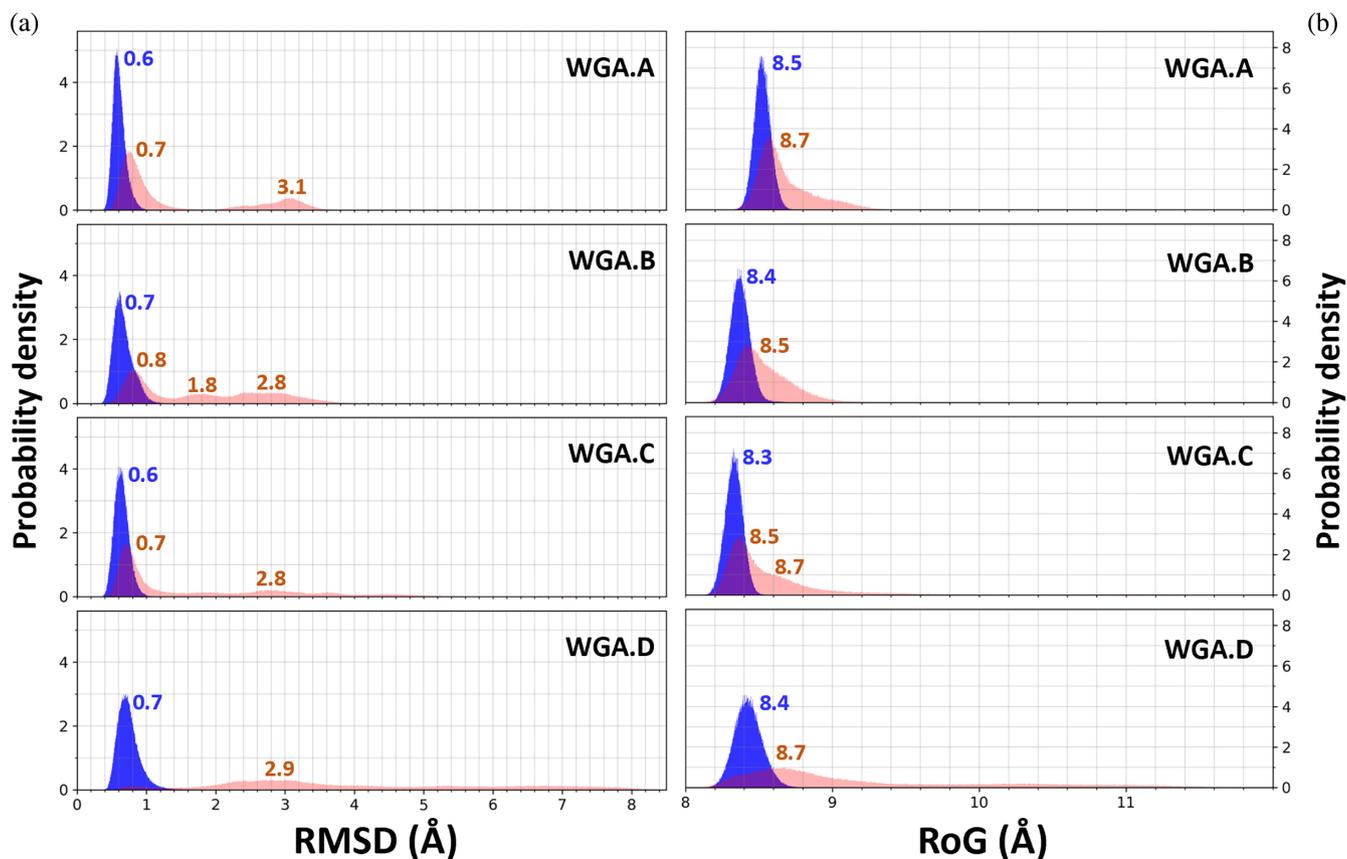


FIGURE 3 Molecular size distributions of WGA domains within the isolated monomer during MD simulations. (a) RMSD probability densities. Values were computed for individual domains, taking the initial structure as a reference. (b) RoG probability densities. Results for trajectories produced at 20 and 120°C are in blue and red, respectively. The numerical values in the plots correspond to the maximum of the closest peak. MD, molecular dynamics; RMSD, root-mean-square deviation; RoG, radius of gyration; WGA, wheat germ agglutinin.

20°C while the unfolded conformation predominates at 120°C across all four domains. The simulations were individually carried out for each isolated domain and, additionally, for the monomerized subunit to evaluate potential interdomain cooperative effects. For each system at both temperatures, five replicates were simulated, each spanning 500 ns, resulting in a cumulative simulation time of 25 μ s. The results corresponding to each assessed conformational metric are initially presented for the domains within the subunit. Subsequently, they are briefly compared with those obtained from simulations conducted on the isolated domains.

The autocorrelation function of the main-chain dihedral angles and RMSD were analyzed for each replicate to investigate the structural relaxation of the systems on the spanned timescale. Through the autocorrelation function, it is possible to determine whether the system has evolved to the point where the results are no longer significantly influenced by its initial conditions, whereas the RMSD analysis enables the visualization of large-scale changes in protein conformation. This last aspect is especially significant to assess in the simulations performed

at high temperature, where major conformational changes relative to the initial structure were anticipated. Analysis of the dihedral angles of the main chain revealed that relaxation of each domain occurred within the first 100 ns in each trajectory (Figure SM2 in Data S1). Therefore, subsequent analyses were conducted for each system utilizing a combined trajectory of 2 μ s, comprising the final 400 ns extracted from the five individual replicates.

Probability densities for RMSD and radius of gyration (RoG) were calculated from the respective time course plots (Figure SM3 in Data S1) to characterize the central and dispersion trends of structural flexibility of the domains (Figure 3). At 20°C, the average RMSD and RoG values of the four domains oscillated around 0.6–0.7 and 8.3–8.5 Å, respectively, with a dispersion toward slightly larger values for WGA.D. At 120°C, the four domains exhibited different degrees of expansion, including conformers as compact as the native-like state. WGA.B and WGA.D showed the lowest and highest degree of expansion, respectively. In general, the trajectories of the four isolated domains mirrored those within the subunit,

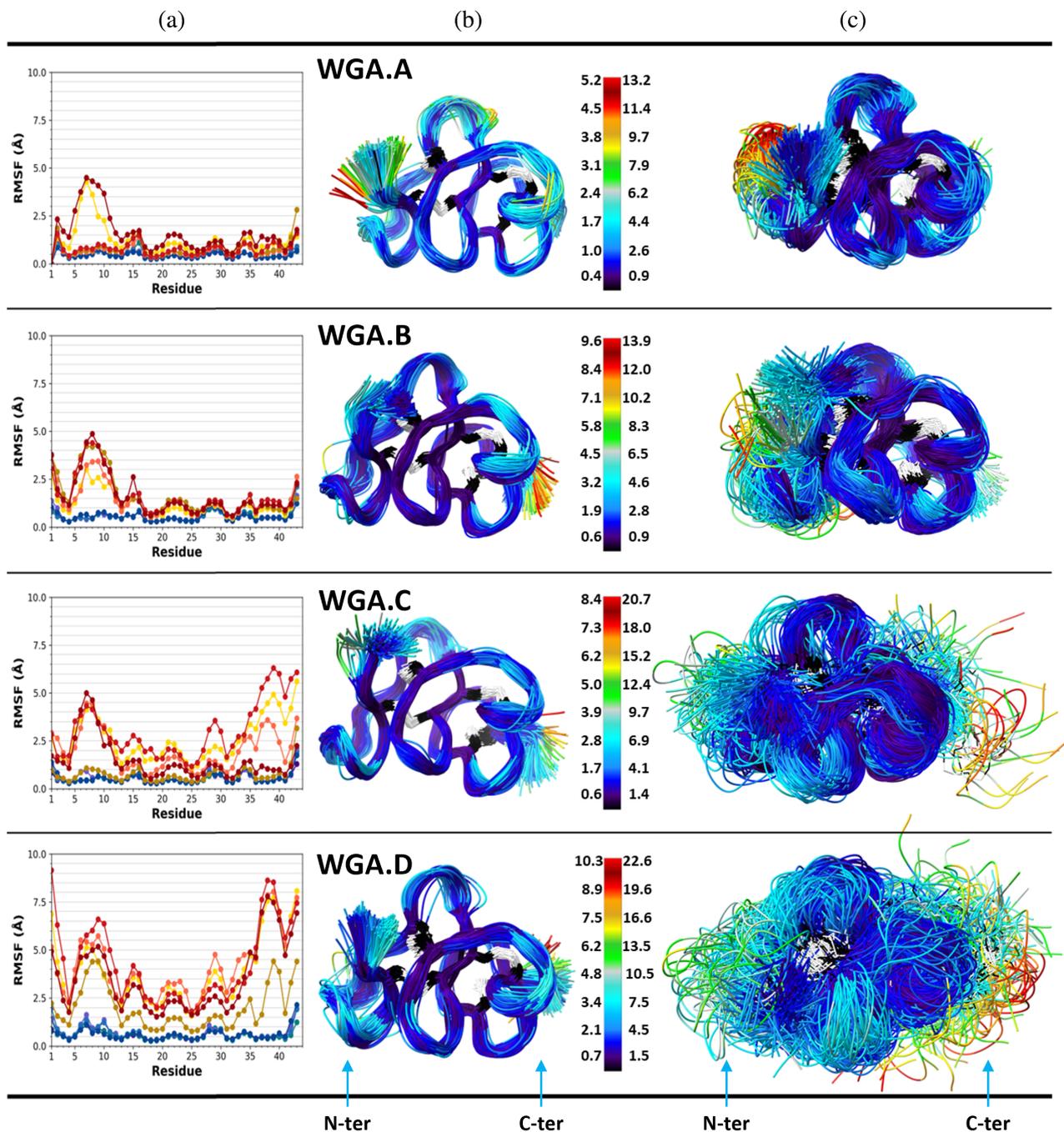


FIGURE 4 Conformational flexibility of WGA domains within the isolated monomer. (a) RMSF plots. Values for the five individual trajectories at 20 and 120°C are depicted in blue and red hues, respectively. (b and c) Snapshots extracted from the MD trajectories at 20 and 120°C, respectively. These snapshots, taken every 5 ns, were optimally aligned based on a subgroup of residues exhibiting the lowest mobility (RMSD) in each concatenated trajectory (Martínez, 2015). Using RMSD values of heavy backbone atoms, regions within each snapshot were color-coded according to the adjacent color scale (RMSD in Å). Disulfide bridges are represented by black-white sticks. MD, molecular dynamics; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; WGA, wheat germ agglutinin.

although exhibiting somewhat larger RMSD and RoG values (Figure SM4 in Data S1). These differences were more noticeable at 120°C.

Figure 4 depicts the conformational flexibility of each domain within the subunit in terms of RMSF and snapshots sampled every 5 ns (Martínez, 2015). At both

temperatures, the central region of each domain remained stiffer than the ends of the chain. According to these data, the higher expansion of WGA.D at 20°C (Figure 4) was due to a higher degree of unpacking of both terminal segments. The N-ter segment of WGA.A and WGA.B exhibited greater mobility at 120°C, while

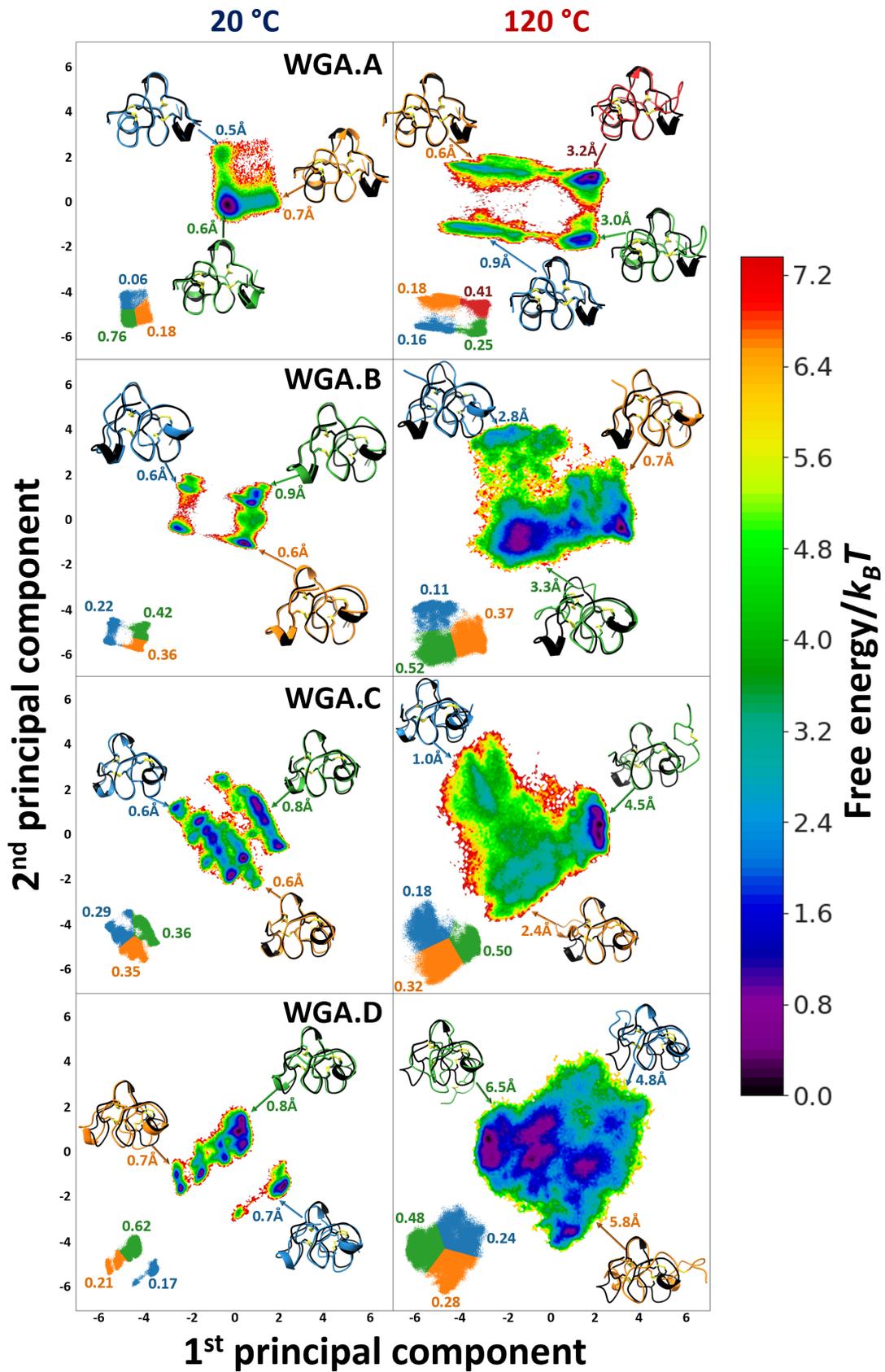


FIGURE 5 Legend on next page.

the rest of the structure remained relatively compact. In contrast, WGA.C and, to a greater extent, WGA.D exhibited a more pronounced unfolding. In WGA.C and WGA.D, the C-ter segment containing the 36:41 disulfide bridge temporarily lost contact with the protein core, a phenomenon not observed in WGA.A and WGA.B. In isolation, the four domains exhibited greater flexibility compared to their counterparts within the subunit (Figure SM5 in Data S1), an observation consistent with the RMSD and RoG results. In this case, only *isol*WGA.B kept the 36:41 disulfide bridge packed along all trajectories.

To characterize the conformational heterogeneity of the domains, free energy landscapes (FEL) were generated from a principal component analysis on the backbone dihedral angles (Figure 5 and Figure SM6 in Data S1) (Altis et al., 2008; Scherer et al., 2015). Conformational clusters and their corresponding abundances were determined using the K-means clustering approach, which was guided by silhouette scores. At 20°C, three distinct clusters were solved for each of the domains assembling the subunit, although they varied in their dispersion and apparent number of basins of attraction. The centroids of these clusters corresponded to conformers that exhibited RMSD < 1 Å. The three clusters were enriched in a roughly equal manner in WGA.B and WGA.C. Conversely, the most similar group of conformers to the reference structure was predominant in WGA.A, and the most dissimilar group was in WGA.D, emphasizing distinct structural preferences in each variant. At 120°C, the four domains exhibited a significantly increased FEL dispersion, composed of a higher number of shallow and broad basins of attraction. In all cases, the dominant conformers were those of the cluster with the highest RMSD. WGA.A and WGA.B exhibited a higher frequency of visits to native-like conformations, with their unfolded forms experiencing a reduced degree of expansion. For these domains, the dominant conformers were preferentially unfolded in their C-ter and N-ter regions, respectively, with minor changes in the central region. Like WGA.A, WGA.C kept native-like conformations in its N-ter and central regions, while its C-ter exhibited much larger unfolding. In contrast, WGA.D unfolded not only at its two termini, but its most abundant cluster also showed the largest structural perturbations in

the central region. Similar behaviors were observed for the isolated domains, with a notable increase in the dispersion and number of basins of attraction in the FELs at 120°C compared with the subunit domains (Figure SM6 in Data S1). Thus, the increased structural fluctuations of isolated domains, especially at high temperatures, appear to be a key distinction compared to their assembly as subunits.

The above-mentioned results unveiled substantial conformational differences between the four domains, regardless of their state as independent entities or as constituents of the subunit. To explore the contributing factors to these differences, a comprehensive analysis of hydrogen bonding was undertaken. Figure 6 depicts hydrogen bonds identified in the domains constituting the whole subunit, based on the heat maps presented in Figure SM7 in Data S1. Based on their persistence, they were categorized as stable (with cumulative frequency >0.85) or moderately stable (with cumulative frequency ranging from 0.6 to 0.85). At 20°C, a shared network of hydrogen bonds, encompassing approximately 60% of the entire domain sequence (see Figure 6a), was identified consistently across all four domains (Figure 6b). This interaction core comprised 14 hydrogen bonds, nine of them exhibiting high stability and the remaining with moderate stability. Residues 18, 19, 24, and 26 (central region), as well as residues 28, 32, and 37 (C-ter region), were involved in the formation of eight cooperative bridges. S²⁰ was observed to have a pivotal position in all domains, as it participated in the establishment of four hydrogen bonds with residues 23, 24, and 37. The four domains shared a single hydrogen bond (8:11) at the N-ter region.

In addition to the completely conserved hydrogen bonds, there were unshared bridges or with varying cumulative frequencies between the domains. Figure 6c compares WGA.B and WGA.D, whose hydrogen bonding patterns were the most dissimilar. Q³⁷ side chain in WGA.B formed three additional hydrogen bonds, with the S⁶ and Q⁷ side chains and the S⁶ backbone. The same pattern was observed in WGA.C. Within these two domains, Q³⁷ engaged in the formation of more bridges than any other residue. In contrast, the E⁶ and K⁶ side chains in WGA.A and WGA.D, respectively, were fully exposed to the solvent, forming no interaction with the

FIGURE 5 Conformational space of WGA domains within the isolated monomer. Free energy landscapes were crafted by projecting a PCA of backbone dihedral angles projected onto the first two principal components. Results for trajectories generated at 20 and 120°C are shown in the left and right columns, respectively. The main clusters in the FELs, along with the corresponding abundance fractions, were identified through a K-means clustering (shown in the left-bottom corners). Conformers corresponding to cluster centroids are shown as cartoons. The adjacent values denote the RMSD of each centroid compared to the reference initial structure, depicted in black. FELs, free energy landscapes; PCA, principal component analysis; RMSD, root-mean-square deviation; WGA, wheat germ agglutinin.

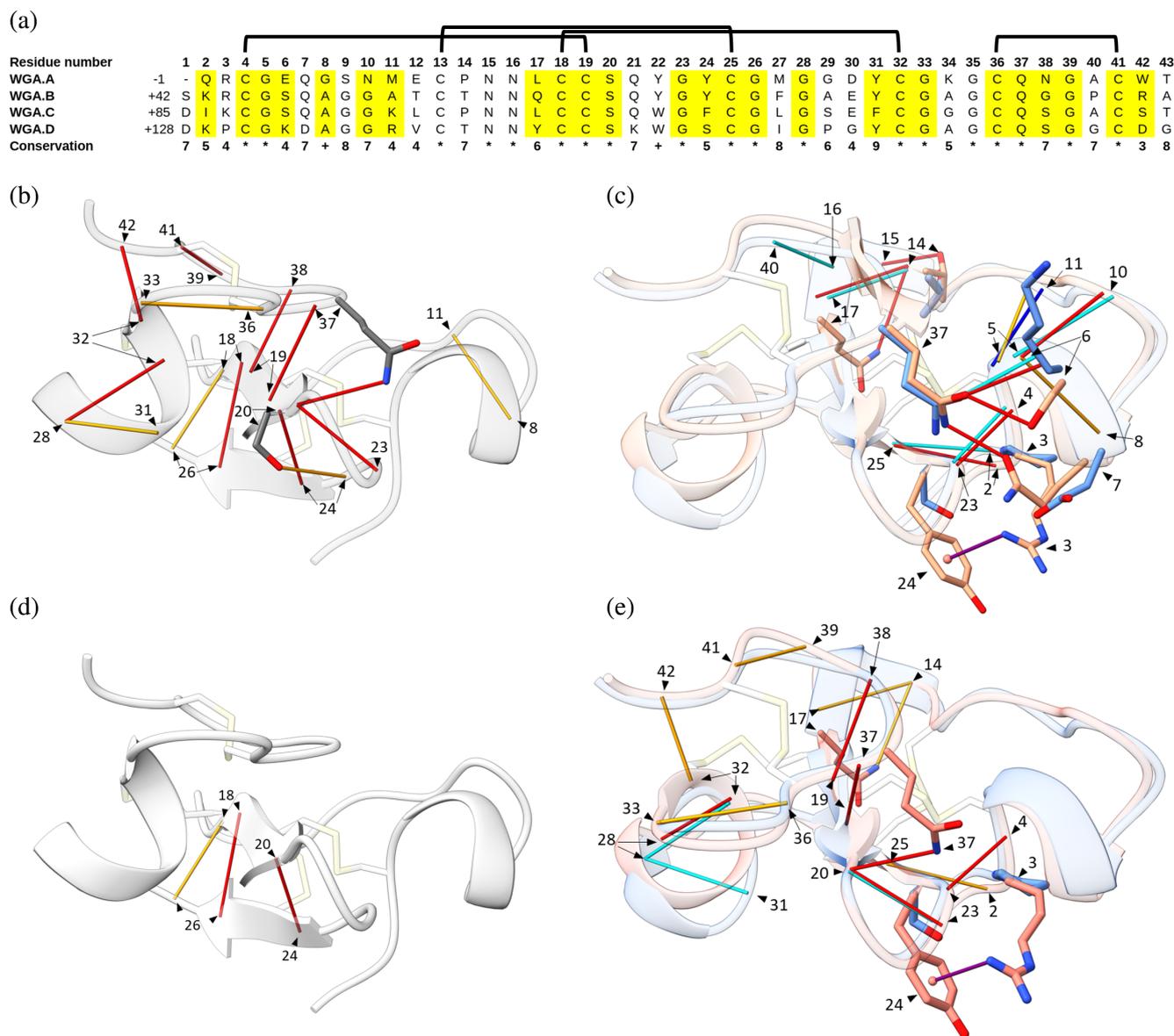


FIGURE 6 Hydrogen bonding patterns in WGA domains within the isolated monomer. (a) Sequence alignment of the four domains. The residue number heading was generated according to this alignment. The actual position for a specific residue within the protein can be determined by summing the numerical value situated to the left of each domain sequence. Residues forming hydrogen bonds in the four domains at 20°C were identified and highlighted in yellow. The Conservation row reflects a scale from 0 (null conservation of physicochemical properties of the amino acid group) to 10 (complete conservation, denoted as “+”) as defined in Livingstone and Barton (1993). Cysteine pairs are connected by horizontal brackets. (b) Schematic representation of hydrogen bonds consistently identified across the four domains during MD simulations at 20°C. Red lines indicate bonds with high persistence, while orange lines denote those with moderate persistence. (c) Contrast of non-shared hydrogen bonds between WG.A.B (in red hues) and WG.A.D (in blue hues) at 20°C. The purple line highlights a highly persistent cation- π interaction. (d and e) Similar to (b) and (c) for trajectories at 120°C. MD, molecular dynamics; WGA, wheat germ agglutinin.

Q³⁷ side chain. WG.A.D lacked an additional hydrogen bond because the shorter side chain of D⁷ precluded its interaction with Q³⁷. The presence of an enduring cation- π contact between the side chain of a basic residue at position 3 and an aromatic residue at position 24 was seen in WG.A.A, WG.A.B, and WG.A.C. This interaction was not formed in WG.A.D, which has the P³-S²⁴ pair instead. The

reduced number of hydrogen bonds and the lack of the cation- π bond in WG.A.D correlated with a weakening of hydrogen bonds formed by the residue pairs 2:25, 4:23, 5:10, 5:37, and 5:11. Finally, the region of residues 14–17 showed hydrogen bonds that were more stable in WG.A.B.

At 120°C, significant alterations were observed in the hydrogen bonding networks, although complete

disruption was not observed in any of the domains. As depicted in Figure 6d, three hydrogen bonds of the common core remained in all domains as prevalent as at low temperature. Other core hydrogen bonds (i.e., 2:25, 4:23, and 14:17) were kept with high stability, except in WGA.D (Figure 6e). It is worth mentioning that the cation- π interaction between residues 3 and 24 exhibited a consistently high prevalence in the three domains. The most significant disparities were noted in the C-ter region. The bonds located at positions 32:42, 33:36, and 39:41 exhibited a moderate level of stability in WGA.A and WGA.B, while they were not present in the other two domains. The hydrogen bond at position 28:32 exhibited a high level of strength in WGA.A and WGA.B, but it was weaker in WGA.C and WGA.D. In contrast, the 28:31 bond was disrupted in WGA.B and WGA.C, whereas it persisted moderately in WGA.A and WGA.D. Overall, these results seemed to account for the varying stabilities of the domains, with the largest contrasts observed between WGA.B and WGA.D.

At each temperature, the intradomain hydrogen bonding patterns within the protein were very similar to those observed in the isolated domains (Figure S8 in Data S1). Regarding interdomain interactions, an inspection of the crystal structure of the whole protein showed four hydrogen bonds between WGA.A and WGA.B and one between WGA.C and WGA.D (Schwefel et al., 2010). Only the one between G³⁸(WGA.A) and the side chain of N⁵⁷(WGA.B) had a moderate persistence in the 20°C simulations, whereas it was completely lost at 120°C. No further interdomain hydrogen bonding was observed at any of the temperatures.

3 | DISCUSSION

Wild-type WGA demonstrated a complex unfolding mechanism, wherein the dimer plays a stabilizing role that becomes more pronounced as the pH of the environment approaches neutral values (Leyva et al., 2019; Portillo-Téllez et al., 2011). In a previous study, the unfolding of fully monomerized WGA at pH 2 was found adequately described by a mechanism of three independent two-state transitions (Portillo-Téllez et al., 2011). Building on this, the collective sum of the endotherms from the isolated domains obtained herein at pH 4.7 closely mirrored the shape and magnitude of the peak observed at high temperatures for the complete agglutinin, where the monomers were predominant. Furthermore, MD simulations revealed limited inter-domain communication within the monomeric subunit. Both results further supported the notion that unfolding of domains proceeds independently. However, analysis of

WGA endotherms suggested that the stabilities of the more thermolabile domains are affected within the dimer. These energetic perturbations likely arose from cooperative effects between subunits that, because of the intricate nature of the underlying unfolding mechanism, could not precisely be quantified from the analysis of calorimetric traces of the entire protein. In contrast, the most stable domain, whose unfolding occurred at temperatures where the protein is predominantly monomerized, and with the domains with lower stability mostly unfolded, exhibited a behavior closely resembling its isolated counterpart. Despite this complexity, the characterization of the isolated domains allowed us to ascertain the sequence of unfolding transitions that define the multistage mechanism of the complete protein.

MD simulations were performed at 20 and 120°C, that is, at pre- and post-transition temperatures for the four domains, respectively. In line with the calorimetric findings, the domains displayed similar behaviors whether they were linked together in the complete subunit or in an isolated form. Nevertheless, the isolated domains exhibited increased conformational freedom in the unfolded state, likely due to the absence of adjacent domains, allowing them to occupy a larger space. Through a comprehensive analysis of the trajectories, we pinpointed a set of 14 stable hydrogen bonds that exhibit evolutionary conservation across all four domains. S²⁰ maintained a pivotal position within this cooperative network of interactions by forming four hydrogen bonds simultaneously. Furthermore, the hydroxyl group of this residue plays a crucial role in ligand recognition across all domains (Itakura et al., 2017), suggesting that it has undergone evolutionary co-optimization concerning both stability and function. Q³⁷ is an additional crucial residue that contributed to the high structural stability of WGA domains. However, across the four domains, only two of the interactions formed by this residue were identical. In contrast to WGA.D, Q³⁷ established three additional strong hydrogen bonds in WGA.B, thereby forming a staple-like structure connecting the N-term and C-term segments. WGA.D also lacked the highly stable cation- π interaction between residues 3 and 24. Initial computational findings from our laboratory indicate that the substitution of residues P³, K⁶, E⁷, and S²⁴ in domain D with residues R³, S⁶, Q⁷, and Y²⁴ from domain B significantly enhances the compaction of the most labile domain. This behavior correlates with the establishment of two hydrogen bonds and a cation- π interaction, which are characteristic of the more stable domain. Conversely, substituting Q³⁷ with alanine destabilizes all four domains, consistent with our expectations from the aforementioned results. Comprehensive details on these findings, along with their experimental validation, will be presented in a forthcoming communication.

Collectively, these findings rationalize the reduced stability observed in the D domain, evidenced in its higher susceptibility for unfolding at both ends of the polypeptide chain. These insights provide a framework for the engineering of domain mutants with improved conformational stability, with the aim of maximizing the use of WGA as a carrier for drug delivery purposes.

4 | CONCLUDING REMARKS

The exceptional ability of lectins to recognize glycoconjugates has been increasingly exploited for the advancement of research tools and clinical applications. As a potential drug carrier, WGA stands out for its significant structural stability, high water solubility, specific recognition of cell lineages, especially malignant ones, and ability to surmount barriers such as the blood–brain barrier and the enterocytes of the intestinal epithelium. Notwithstanding these advantages and the research endeavors conducted thus far, the utilization of WGA in human and animal therapeutics is yet in its nascent stages of development. In this study, we thoroughly examined the thermodynamics, structure, and dynamics of both WGA and its four isolated domains. Through a comparative analysis of the domains as part of the complete protein and expressed independently, we pinpointed some key factors influencing the complex unfolding process of this agglutinin. Specifically, we identified crucial interactions that can serve as a blueprint for strategically enhancing the stability of the protein. This improvement aims to bolster its resilience against the harsh conditions of extreme acidity and proteolytic degradation encountered during the transit through the digestive tract. Such insights hold significant potential for fine-tuning the effectiveness and safety of WGA in drug transportation and delivery.

5 | MATERIALS AND METHODS

5.1 | Materials

Unless otherwise stated, all chemical reagents were obtained from Sigma Chemical Co (St. Louis, Missouri). WGA was purchased as a mixture of the three natural agglutinin variants. WGA isoform 1, purified using cationic-exchange chromatography in an HPLC system, as described elsewhere (Chavelas et al., 2004; Portillo-Télez et al., 2011), was used for the calorimetric measurements. Isolated domains were produced recombinantly as described elsewhere (Leyva et al., 2019). The purity of all protein samples was determined to be >97% through SDS-PAGE electrophoresis

and mass spectrometry analysis. All measurements were carried out in 30 mM acetate, pH 4.7.

5.2 | Circular dichroism (CD) spectroscopy

CD spectra were obtained on a JASCO J-720 spectropolarimeter, equipped with a PTC-348WI Peltier for temperature control, as detailed elsewhere (Arreguín-Espinosa et al., 2001). The spectra were recorded at room temperature in the far-UV region.

5.3 | DSC

DSC traces were obtained with a VP-Capillary DSC equipment (Malvern Panalytical Inc., Westborough, Massachusetts). All experiments were performed at a scan rate of 60°C/h. Buffer-buffer baselines were run at the same experimental conditions before recording the signal from the protein solution and subtracted from sample traces. Calorimetric traces of the isolated domains were fitted to a two-state model programmed in the Origin7 software package (Malvern Instruments; OriginLab, Northampton, Massachusetts). Wild-type WGA traces were fitted using an unfolding model that included dimer dissociation and independent two-state transitions for each domain (see Appendix S1). A total of 17 microstates are involved in this model. The partition function (Q) for this ensemble of states is calculated from the dimer dissociation constant (K_d) along with the unfolding constants of the four domains, denoted as K_1 to K_4 . The constants are calculated as functions of temperature (T) from their enthalpy and entropy changes (or, equivalently, from the enthalpy change (ΔH_i) and the temperature value at half transition, $T_{m,i}$), which were assumed to be independent of temperature. Once it is known, Q is used to calculate populations (fractions) of each microstate (f_i). Using these fractions and individual enthalpies, the excess enthalpy function of the entire system ($\langle \Delta H \rangle$) is established. Finally, the derivative of $\langle \Delta H \rangle$ gives the excess heat capacity tracing ($C_{p,exc}$) versus temperature (i.e., the DSC curve; see Appendix S1 and Data S1):

$$C_{p,exc} = \frac{1}{RT^2} \left[\sum_{i=1}^{16} f_i \Delta H_i^2 - \frac{\langle \Delta H \rangle^2 Q}{1+Q} \right],$$

where R is the ideal gas constant. A script to simultaneously fit the model equations to experimental DSC curves obtained as a function of protein concentration (expressed in terms of monomer equivalent) was implemented in LabTalk language (Origin Lab).

5.4 | NMR spectroscopy

The proteins were dissolved in phosphate buffer containing 3% of D₂O. The NMR data were acquired in a Bruker AVANCE III 700 MHz spectrometer at 298 K as described elsewhere (Titau-Delgado et al., 2023). The acquired experiments were 2D-TOCSY (80 ms mixing time) and 2D-NOESY (150 and 300 ms mixing time). Water signal suppression was performed with DPGSE (Double Pulsed Field Gradient Spin Echo); these data were processed with NMRPipe software (Delaglio et al., 1995; Nguyen et al., 2007). The assignment was done using the strategy previously described by Wüthrich using the 2D-TOCSY experiment (80 ms mixing time) and the NOESY experiment (300 ms mixing time) (Wüthrich, 2003). Distance constraints were obtained from the 2D-NOESY (150 ms mixing time). After the assignment, 200 structures were calculated with CYANA 2.1, picking the 20 structures with the lowest energy values for refinement (Güntert, 2004). These structures were refined by MD calculations using AMBER 16 (Salomon-Ferrer et al., 2013a). The structures were validated and deposited in the PDB Protein Data Bank (PDB IDs 8vu8, 8vu7, and 8vu8 for *isol*WGA.A, *isol*WGA.C, and *isol*WGA.D, respectively).

5.5 | MD simulations

MD simulations were performed using the pmemd.cuda module of AMBER20 and the FF19SB force field (Case et al., 2014; Salomon-Ferrer et al., 2013b; Tian et al., 2020). Simulations were carried out for both the WGA monomer and each of its isolated domains. The subunit coordinates were taken from the dimer crystallographic structure of WGA isoform 1 in complex with N-acetyl-D-glucosamine (PDBID: 2uvo [Schwefel et al., 2010]). The lowest energy structure determined in this study by NMR was selected for each isolated domain, except for domain B, which was isolated from the crystallographic structure of the whole lectin. Four alanine residues were attached at the N-terminal end of *isol*WGA.A to match experimental conditions (Leyva et al., 2019). Proteins were protonated at pH 4.7 using PDBfixer (Eastman et al., 2017). Each protein was centered in a dodecahedral box, initially spanning 12 Å further from the solute in each direction, and then solvated with the OPC water model (Izadi et al., 2014). The charges were neutralized by adding the appropriate number of Na⁺ or Cl⁻ ions. To adjust the solvent orientation and remove local clashes, the systems were geometrically optimized using the steepest descent algorithm for 5000 cycles. Initial velocities were assigned to get a 150 K distribution, slowly increasing to the production

temperature in 0.8 ns in the NVT ensemble. The system was further equilibrated for 1 ns at the corresponding temperature in the NPT ensemble as described elsewhere (Avila-Barrientos et al., 2022). Production trajectories were run in the NPT ensemble. Systems were simulated using periodic boundary conditions and Ewald sums (grid spacing of 1 Å) for treating long-range electrostatic interactions, with a 10 Å cutoff for direct interactions. Lennard-Jones interactions were calculated with a cutoff distance of 12 Å. Temperature control was achieved using Langevin dynamics with collision frequency of 4 ps⁻¹. Pressure control was achieved using the Berendsen barostat with a pressure relaxation time of 2 ps. The system was maintained at 1 bar. The SHAKE algorithm was enabled to fix any bond involving hydrogen atoms (Miyamoto and Kollman, 1992). Hydrogen atoms repartition scheme was applied using ParmEd (Shirts et al., 2017), which allowed the use of a 4 fs time step integration (Hopkins et al., 2015). Five 500-ns replicas for each system were run at 20 or 120°C, yielding a total simulation time of 25 μs. The sequences of the four domains were aligned to generate a unified residue numbering system, which was used in the rest of the manuscript (see Figure 6 legend for details).

5.6 | MD trajectory analysis

Structural analysis was performed using CPPTRAJ (Roe and Cheatham, 2013). It included autocorrelation, hydrogen bonds (HB), contacts, RoG, RMSF, RMSD, and dihedral angle principal component analysis (dPCA). For RoG, RMSD, and dihedral angle calculations, the region delimited by the first and the last C_{yx} residue was considered, excluding the highly mobile residues at the termini of each domain. HB and contact matrixes were calculated using SenseNet (Schneider and Antes, 2022) with the corresponding CPPTRAJ inputs via Cytoscape (Shannon et al., 2003). All molecular structure representations were created using UCSF ChimeraXv1.6 (Pettersen et al., 2021). Data analysis and visualization plots were created with the Python libraries Pandas (McKinney, 2010), Matplotlib (Hunter, 2007), and Seaborn (Waskom, 2021). The flexibility of the protein was evaluated using the MDLovofit program, which determines regions with the lowest RMSD value to be used as the aligning mask for all trajectory conformers (Martínez, 2015).

5.7 | dPCA

To correctly differentiate the general and internal motions, dihedral angle PCA (dPCA) was performed on

the concatenated trajectories of each system at the same temperature (Altis et al., 2008). dPCA was calculated using the functions $\sin(x)$ and $\cos(x)$ to transform backbone dihedral angles into a linear metric coordinate space. Afterward, a 2N covariance matrix was calculated. The covariance matrix was diagonalized and instantaneous linear correlations between variables were removed. The eigenvectors and eigenvalues were calculated and arranged in descending order. PyEMMA (Scherer et al., 2015) was used to build two-dimensional FEL with the two first principal components, PC1 and PC2 (Cofas-Vargas et al., 2022):

$$F(x) = -k_B \times T \times \ln P(x),$$

where $F(x)$ is the x -coordinate FEL, k_B the Boltzmann constant, T the absolute temperature, and $P(x)$ is the x -coordinate probability distribution, taken as the two-dimensional histogram of PC1 and PC2.

5.8 | Cluster analysis

PC1 and PC2 were combined using the NumPy library (Harris et al., 2020) into a single array. Based on the silhouette score, which measures the cohesion and separation of data points within clusters and provides a quantitative assessment of clustering quality, the optimal number of clusters for the dataset was determined. For each cluster, the K-means algorithm was employed, and the silhouette score was calculated by testing 2 to 10 clusters with the scikit-learn library (Pedregosa et al., 2011). The final clustering results were obtained by executing again the K-means algorithm with the optimal number of clusters. Upon clustering, the frames within each cluster were extracted and saved.

AUTHOR CONTRIBUTIONS

Jorge Luis Medrano-Cerano: Writing – review and editing; methodology; investigation; formal analysis. **Luis Fernando Cofas-Vargas:** Conceptualization; investigation; methodology; writing – original draft; writing – review and editing; formal analysis. **Eduardo Leyva:** Methodology; writing – review and editing. **Jesús Antonio Rauda-Ceja:** Methodology; writing – original draft; writing – review and editing; investigation; formal analysis. **Mateo Calderón-Vargas:** Writing – review and editing; methodology; investigation; formal analysis. **Patricia Cano-Sánchez:** Methodology; writing – review and editing. **Gustavo Titaux-Delgado:** Methodology; writing – review and editing. **Carolina Monserrath Melchor-Meneses:** Methodology; writing – review and editing. **Andrés Hernández-Arana:** Investigation;

writing – review and editing; writing – original draft; formal analysis. **Federico del Río-Portilla:** Methodology; writing – review and editing; funding acquisition; formal analysis. **Enrique García-Hernández:** Conceptualization; funding acquisition; writing – original draft; writing – review and editing; formal analysis.

ACKNOWLEDGMENTS

J.L.M.-C. and J.A.R.-C. are doctoral students from Programa de Doctorado en Ciencias Bioquímicas from Universidad Nacional Autónoma de México (UNAM); C.M.M.-M. is a doctoral student from Programa de Doctorado en Ciencias Químicas from UNAM, and received Ph.D. fellowships from CONAHCyT. We thank the LANEM-IQ-UNAM for the use of the CD Spectropolarimeter and Adriana Romo-Pérez for the technical support. We are grateful to Lucía del Carmen Márquez-Alonso, Lucero Ríos-Ruiz, and Eréndira García-Ríos for their assistance in carrying out HPLC analysis and MALDI-TOF spectra. Finally, we thank Dr. Beatriz Quiroz-García for the support conducting NMR experiments. This work was financed by Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México [PAPIIT, IN206221, IN214722, IN209624]. NMR experiments were performed in the UNAM's NMR lab: LURMN at IQ-UNAM which is funded by CONAHCyT (Project 0224747).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ORCID

Luis Fernando Cofas-Vargas  <https://orcid.org/0000-0001-5603-1437>

Federico del Río-Portilla  <https://orcid.org/0000-0001-8672-0567>

Enrique García-Hernández  <https://orcid.org/0000-0003-4561-4973>

REFERENCES

- Altis A, Otten M, Nguyen PH, Hegger R, Stock G. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *J Chem Phys.* 2008;128:245102.
- Apfelthaler C, Anzengruber M, Gabor F, Wirth M. Poly-(L)-glutamic acid drug delivery system for the intravesical therapy of bladder cancer using WGA as targeting moiety. *Eur J Pharm Biopharm.* 2017;115:131–9.
- Arreguín-Espinosa R, Fenton B, Vázquez-Contreras E, Arreguín B, García-Hernández E. PFA, a novel mollusk agglutinin, is structurally related to the ribosome-inactivating protein superfamily. *Arch Biochem Biophys.* 2001;394:151–5.
- Aub JC, Sanford BH, Cote MN. Studies on reactivity of tumor and normal cells to a wheat germ agglutinin. *Proc Natl Acad Sci USA.* 1965;54:396–9.

- Avila-Barrimentos LP, Cofas-Vargas LF, Agüero-Chapin G, Hernández-García E, Ruiz-Carmona S, Valdez-Cruz NA, et al. Computational design of inhibitors targeting the catalytic β subunit of *Escherichia coli* F₀F₁-ATP synthase. *Antibiotics* (Basel). 2022;11:557.
- Balčiūnaitė-Murzienė G, Dzikaš M. Wheat germ agglutinin—from toxicity to biomedical applications. *Appl Sci*. 2021;11:884.
- Case DA, Berryman JT, Betz RM, Cerutti DS, Cheatham III TAD TE, Duke RE, et al. AMBER 2014. San Francisco: University of California; 2014.
- Chavelas EA, Beltrán AP, Pérez-Hernández G, García-Hernández E. Spectroscopic characterization of the thermal unfolding of wheat germ agglutinin. *J Mex Chem Soc*. 2004;48:279–82.
- Chiu HI, Lim V. Wheat germ agglutinin-conjugated disulfide cross-linked alginate nanoparticles as a docetaxel carrier for colon cancer therapy. *Int J Nanomedicine*. 2021;16:2995–3020.
- Cofas-Vargas LF, Mendoza-Espinosa P, Avila-Barrimentos LP, Prada-Gracia D, Riveros-Rosas H, García-Hernández E. Exploring the druggability of the binding site of aurovertin, an exogenous allosteric inhibitor of F₀F₁-ATP synthase. *Front Pharmacol*. 2022;13:1012008.
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995;6:277–93.
- Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, et al. OpenMM 7: rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol*. 2017;13:e1005659.
- Espinosa JF, Asensio JL, García JL, Laynez J, Bruix M, Wright C, et al. NMR investigations of protein-carbohydrate interactions binding studies and refined three-dimensional solution structure of the complex between the B domain of wheat germ agglutinin and N,N',N''-triacetylchitotriose. *Eur J Biochem*. 2000;267:3965–78.
- Gabius H-J. The sugar code: why glycans are so important. *Biosystems*. 2018;164:102–11.
- García-Hernández E, Zubillaga R a RA, Rojo-Domínguez A, Rodríguez-Romero A, Hernández-Arana A. New insights into the molecular basis of lectin-carbohydrate interactions: a calorimetric and structural study of the association of hevein to oligomers of N-acetylglucosamine. *Proteins Struct Funct Genet*. 1997;29:467–77.
- Güntert P. Automated NMR structure calculation with CYANA. In: Clifton NJ, editor. *Methods in molecular biology*. Volume 278. New Jersey: Humana Press; 2004. p. 353–78.
- Hädrich G, Vaz GR, Bidone J, Yurgel VC, Teixeira HF, Gonçalves Dal Bó A, et al. Development of a novel lipid-based nanosystem functionalized with WGA for enhanced intracellular drug delivery. *Pharmaceutics*. 2022;14:2022.
- Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature*. 2020;585:357–62.
- Hopkins CW, Le Grand S, Walker RC, Roitberg AE. Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput*. 2015;11:1864–74.
- Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9:90–5.
- Itakura Y, Nakamura-Tsuruta S, Kominami J, Tateno H, Hirabayashi J. Sugar-binding profiles of chitin-binding lectins from the Hevein family: a comprehensive study. *Int J Mol Sci*. 2017;18:1160.
- Izadi S, Anandakrishnan R, Onufriev AV. Building water models: a different approach. *J Phys Chem Lett*. 2014;5:3863–71.
- Jiménez-Barbero J, Javier Cañada F, Asensio JL, Aboitiz N, Vidal P, Canales A, et al. Hevein domains: an attractive model to study carbohydrate-protein interactions at atomic resolution. *Adv Carbohydr Chem Biochem*. 2006;60:303–54.
- Leyva E, Medrano-Cerano JL, Cano-Sánchez P, López-González I, Gómez-Velasco H, del Río-Portilla F, et al. Bacterial expression, purification and biophysical characterization of wheat germ agglutinin and its four hevein-like domains. *Biopolymers*. 2019;110:e23242.
- Liu Y, Zhao Y, Liu J, Zhang M, Yu M, Feng N, et al. Wheat germ agglutinin modification of lipid-polymer hybrid nanoparticles: enhanced cellular uptake and bioadhesion. *RSC Adv*. 2016;6:36125–35.
- Livingstone CD, Barton GJ. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput Appl Biosci*. 1993;9:745–56.
- Martínez L. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS One*. 2015;10:e0119264.
- McKinney W. Data structures for statistical computing in python. *Proceedings of the 9th python in science conference*; 2010. p. 56–61.
- Miyamoto S, Kollman PA. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem*. 1992;13:952–62.
- Nguyen BD, Meng X, Donovan KJ, Shaka AJ. SOGGY: solvent-optimized double gradient spectroscopy for water suppression. A comparison with some existing techniques. *J Magn Reson*. 2007;184:263–74.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30.
- Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci*. 2021;30:70–82.
- Porter LL, Rose GD. A thermodynamic definition of protein domains. *Proc Natl Acad Sci USA*. 2012;109:9420–5.
- Portillo-Télliz MC, Bello M, Salcedo G, Gutiérrez G, Gómez-Vidales V, García-Hernández E. Folding and homodimerization of wheat germ agglutinin. *Biophys J*. 2011;101:1423–31.
- Raghu HV, Kumar N. Rapid detection of *Listeria monocytogenes* in milk by surface plasmon resonance using wheat germ agglutinin. *Food Anal Methods*. 2020;13:982–91.
- Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput*. 2013;9:3084–95.
- Salomon-Ferrer R, Case DA, Walker RC. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip Rev Comput Mol Sci*. 2013a;3:198–210.
- Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with

- AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J Chem Theory Comput.* 2013b;9:3878–88.
- Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, et al. PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. *J Chem Theory Comput.* 2015;11:5525–42.
- Schneider M, Antes I. SenseNet, a tool for analysis of protein structure networks obtained from molecular dynamics simulations. *PLoS One.* 2022;17:e0265194.
- Schwefel D, Maierhofer C, Beck JG, Seeberger S, Diederichs K, Möller HM, et al. Structural basis of multivalent binding to wheat germ agglutinin. *J Am Chem Soc.* 2010;132:8704–19.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13:2498–504.
- Shirts MR, Klein C, Swails JM, Yin J, Gilson MK, Mobley DL, et al. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J Comput Aided Mol Des.* 2017;31:147–61.
- Slavokhotova AA, Shelenkov AA, Andreev YA, Odintsova TI. Hevein-like antimicrobial peptides of plants. *Biochemistry.* 2017;82:1659–74.
- Solís D, Bovin NV, Davis AP, Jiménez-Barbero J, Romero A, Roy R, et al. A guide into glycosciences: how chemistry, biochemistry and biology cooperate to crack the sugar code. *Biochim Biophys Acta.* 2015;1850:186–235.
- Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Miguez AN, et al. ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J Chem Theory Comput.* 2020;16:528–52.
- Titau-Delgado G, Lopez-Giraldo AE, Carrillo E, Cofas-Vargas LF, Carranza LE, López-Vera E, et al. Beta-KTx14.3, a scorpion toxin, blocks the human potassium channel KCNQ1. *Biochim Biophys Acta, Proteins Proteomics.* 2023;1871:140906.
- Tu Z, Cheng S, Dong H, Wang W, Yang X, Gu B, et al. Universal and ultrasensitive detection of foodborne bacteria on a lateral flow assay strip by using wheat germ agglutinin-modified magnetic SERS nanotags. *RSC Adv.* 2022;12:27344–54.
- Wang W, Hassan MM, Mao G. Colloidal perspective on targeted drug delivery to the central nervous system. *Langmuir.* 2023;39:3235–45.
- Waskom M. seaborn: statistical data visualization. *J Open Source Softw.* 2021;6:3021.
- Wright CS. Crystal structure of a wheat germ agglutinin/glycophorin-sialoglycopeptide receptor complex. Structural basis for cooperative lectin-cell binding. *J Biol Chem.* 1992;267:14345–52.
- Wüthrich K. NMR studies of structure and function of biological macromolecules (Nobel lecture). *Angew Chem Int Ed.* 2003;42:3340–63.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Medrano-Cerano JL, Cofas-Vargas LF, Leyva E, Rauda-Ceja JA, Calderón-Vargas M, Cano-Sánchez P, et al. Decoding the mechanism governing the structural stability of wheat germ agglutinin and its isolated domains: A combined calorimetric, NMR, and MD simulation study. *Protein Science.* 2024;33(6): e5020. <https://doi.org/10.1002/pro.5020>