Published in final edited form as:

J Inorg Biochem. 2018 August ; 185: 86–102. doi:10.1016/j.jinorgbio.2018.04.021.

Structural stabilities of Calcium proteins: human intelectin-1 and frog lectin XEEL

John J. Kozak^a, Harry B. Gray^b, Roberto A. Garza-López^{c,*}, Kittikhun Wangkanont^d

^aDePaul University, 243 South Wabash Ave., Chicago, IL 60604-6116

^bBeckman Institute, California Institute of Technology, Pasadena CA 91125

^cDepartment of Chemistry, Seaver Chemistry Laboratory, Pomona College, Claremont, CA 91125

^dDepartment of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok, Thailand 10330

Abstract

We extend our study of the structural stability of helical and nonhelical regions in chain A of human intelectin-1 to include a second human intelectin (4WMY) and the frog protein “Xenopus embryonic epidermal lectin” (XEEL). These unique lectins have been shown to recognize carbohydrate residues found exclusively in microbes, thus they could potentially be developed into novel microbe detection and sequestration tools. We believe that by studying the structural stability of these proteins we can provide insights on their biological role and activities. Using a geometrical model introduced previously, we perform computational analyses of protein crystal structures that quantify the resiliency of the native state to steric perturbations. Based on these analyses, we conclude that differences in the resiliency of the human and frog proteins can be attributed primarily to differences in non-helical regions and to residues near Ca ions. Since these differences are particularly pronounced in the vicinity of the ligand binding site, they provide an explanation for the finding that human intelectin-1 has a higher affinity for a ligand than XEEL. We also present data on conserved and position-equivalent pairs of residues in 4WMY and XEEL. We identify residue pairs as well as regions in which the influence of neighboring residues is nearly uniform as the parent protein denatures. Since the structural signatures are conserved, this identification provides a basis for understanding why both proteins exhibit trimeric structures despite poor sequence conservation at the interface.

Graphical Abstract

*corresponding author.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Native structure of Human intelectin-1 (PDB # 4WMY) and *Xenopus* embryonic epidermal lectin (PDB # 4WN0) showing helices in orange, strands in purple and coils in gray. PDB listings for the three calcium are: 401 (red), 402 (green) and 403 (blue).

I. Introduction

Intelectins are fibrinogen-like domain-containing macromolecules ubiquitous among chordates [1–2]. It is likely that these proteins are innate immune molecules because they are upregulated upon infections, and the proteins can bind microbe-specific carbohydrate residues [3–4]. Although human intelectin-1 does not bind to any of the 609 glycan molecules on the mammalian glycan array,[5] it recruits microbial glycans containing microbe-specific β -D-galactofuranose (Gal β), D-glycerol-1-phosphate (GroP), D/L-*manno*-heptose, D-glycero-D-talo-oct-2-ulosonic acid or 3-Deoxy-D-manno-oct-2-ulosonic acid residues. Similar to human intelectin-1, a *Xenopus* homologue of human intelectin-1, *Xenopus* embryonic epidermal lectin (XEEL), binds to microbe-specific Gal β and GroP [6]. Of special interest is that both intelectins bind several serotypes of *Streptococcus pneumoniae* that contain appropriate ligands in a calcium ion-dependent manner [5]. Thus, these lectins are capable of interacting with bacterial cells. Owing to a unique architecture, XEEL agglutinates bacteria that contain its ligand, which in turn indicates that the innate immune intelectin function is conserved among chordates. Importantly, these findings suggest that intelectins could be used for microbe detection, sequestration, or targeting.

Crystal structures of human intelectin-1 and XEEL, both with and without ligands, allow direct comparisons of their architectures and functions [6]. Both proteins contain three calcium ions per monomer: two are buried within the trimer, likely enhancing structural stability; and the remaining dispositive ion, which is surface exposed, directly binds the exocyclic 1,2-diol epitope on the ligand. Broader comparisons confirm that amino acid sequences around structural calcium sites in intelectins are highly conserved, in accord with the finding that the proteins have the same ligand specificity.

Alas, the crystal structure analyses of intelectins do not tell us about their structural stabilities. Resilience to unfolding is necessary for proteins because they need to maintain appropriate three-dimensional structure to function. Moreover, protein engineering to enhance stability is essential for producing therapeutic proteins. Because intelectins have potential to become novel antimicrobial agents, we aim to characterize structural stabilities of these proteins at the molecular level. Consider the following unanswered questions: (1) Are the stabilities of the structural calcium sites [3] of human intelectin-1 and XEEL similar (we might think so)? (2) Why are the two trimeric structures so similar (a finding that we find surprising, as the former protein is disulfide-linked, but the carbohydrate recognition domain of XEEL (XEEL-CRD) is not)? (3) Despite the conserved ligand binding site and trimeric architecture, why do these lectins have different ligand binding affinity? We will attempt to answer these questions in the hope that greater understanding of intelectin stabilities will speed the development of protein-based agents for microbial identification and elimination.

II. Methods

We have developed a geometrical model to follow quantitatively the early stages of unfolding of a native protein. In effect, we characterize the resiliency of the native state to structural perturbations that can arise as a result of, for example, thermal fluctuations. Our geometrical model has been elaborated in earlier work. Three recent publications illustrating the method of analysis are noted: our study of the structural stability of intelectin-1 [7], a comparative study of three cytochromes [cyt c, cyt-b562, and cyt c' in Ref. (8)], and a study of the thermal denaturation of amicyanin [9], including comparisons with experimental evidence. We present in the Appendix a description of the method used to calculate the signatures of our geometrical model using the symbolic and numeric programming language MAPLE.

Fundamental to our approach is that first-nearest-neighbor repulsive *and* attractive interactions between each residue j and its two nearest-neighbors are kept intact by introducing a triplet modular unit centered on residue j . The geometry of each and every triplet, determined directly from crystallographic data, remains locally optimized.

Overall, the native state portrayed by the crystallographic evidence is globally optimized. Geometrical constraints (excluded volume effects) limit sterically the number of possible configurations accessible to a polypeptide chain in the native state, the key insight of Ramachandran [10] (as seen in ϕ/ψ plots). In our approach, structural perturbations of the native state are studied by relaxing geometrical constraints *only* between and among *non*-nearest neighbors.

Among the myriad of configurations that can be adopted by a segment of the polypeptide chain as the native state is disrupted, we adopt one configuration as representative of many that can result upon destabilization, namely, a maximally extended, linear sequence of triplets. We consider six such configurations, linear extensions of two triplets ($n=5$ residues), three triplets ($n=7$ residues), ..., seven triplets ($n=15$ residues), each reflecting a stage in the denaturation of the native protein. The changing environment of each residue as the protein

unfolds from the native state to one of these linearly-extended reference states is followed quantitatively via direct application of classical Euclidean geometry and trigonometry [See Refs. (7–9) and Appendix].

As noted, the linear extended state in our geometrical model is representative of one of many possible states in the near neighborhood of the globally-optimized native state that are accessible energetically as the protein is destabilized. The configurations represented by the six stages of unfolding considered here can be realized in a sequence of stepwise transitions starting from native state or (with a lower probability) to an advanced stage in a single transition [9].

By adopting the same reference state for all proteins in a given family [the cytochromes studied in Ref. (8), the blue copper proteins amicyanin in Ref. (9) and azurin in Refs. (11–13), or the proteins studied here], unfolding processes can be systematically compared. Moreover, keeping the reference state invariant allows proteins of different families with very different physiological functions to be compared.

To quantify the structural stability of the native state of a protein, we calculate three signatures (See following section and Appendix). A lateral signature quantifies the representation of $n=5$, $n=7$, ..., $n=15$ residue segments of the polypeptide chain in terms of invariant triplets. This signature is determined directly from crystallographic data, and is not model dependent. A second, spatial signature quantifies the radial displacement of each alpha carbon from the native state to one of the six unfolded states (see above); it is calculated using our geometrical model. The third signature gauges (residue by residue) the importance of down-range angular correlations in the polypeptide chain as the protein unfolds.

The crystallographic structure for each of the proteins studied here is reported in the following PDB files: for human intelectin-1, 4WMQ; for the related human intelectin 4WMY; and, for the frog lectin (XEEL), 4WN0.

Note that, in reporting data, we have converted residue signatures for each protein from their PDB listing to a sequential listing. There is a 30 residue gap between the PDB listing for 4WMQ and 4WN0 and a sequential listing, and a 65 residue gap between the PDB listing for 4WN0 and a sequential listing. This relabeling ensures that all three proteins are placed on the same footing, facilitating comparisons

To complete the specification of our geometrical model, we need to assign one Ca ion to be the center of the coordinate system. There are three calcium ions in each of the structures: 4WMQ, 4WMY and 4WN0. Their distances from the (respective) crystallographic center of each protein are: for intelectin-1 [4WMQ], these distances are 24.97, 31.05 and 45.28 Å for Ca (PDB 401), Ca(PDB 402) and Ca(PDB 403), respectively; for intelectin [4WMY], they are 24.84, 31.05 and 45.28 Å for Ca (PDB 401), Ca(PDB 402) and Ca(PDB 403), respectively; and, for XEEL [4WN0], they are 61.39, 62.28 and 58.81 Å for Ca (PDB 401), Ca(PDB 402) and Ca(PDB 403), respectively. Since the three Ca ions in XEEL are essentially equidistant from the crystallographic origin, we assign Ca (PDB 401) as the

center of the coordinate system for all three proteins. Crystallography of Sm-soaked XEEL is in the supplementary data.

III. Global Comparisons

We present in Tables 1(a), (b) and (c), respectively, the overall (all-residue) average lateral signature $\langle T/T_n \rangle$, radial signature $\langle f \rangle$ and angular spread signature $\langle \beta \rangle$ for the three proteins, 4WMQ, 4WMY and 4WN0. The signature $\langle T/T_n \rangle$ is the ratio of the lateral extension of an n -residue segment centered on the α -carbon of residue i in the unfolded state, to the lateral extension of a triplet centered on the α -carbon of residue i in the native state, summed over all residues, and normalized by the number of residues in the particular protein. By construction, values of $\langle T/T_n \rangle$ at each stage of unfolding are (near) integers, reflecting the number of triplets comprising the n -residue segment. See Table 1(a).

The position of the α -carbon of each residue, relative to the Ca ion assigned as the center of the coordinate system (see below), is calculated at each stage of unfolding using our geometrical model. The signature $\langle f \rangle$ is the ratio of the radial extension of an n -residue extension centered on residue i in the unfolded state, to the radial extension of the triplet centered on residue i in the native state, summed over all residues, and normalized by the number of residues of the particular protein.

By construction, values of $\langle f \rangle$ greater than “1” gauge the spatial displacement of the α -carbon of a residue from the native state as the protein unfolds. As seen in Table 1(b), there is a gradual increase in the value of $\langle f \rangle$ as each protein unfolds. Note that $\langle f \rangle \approx 1.4$ for all three proteins; typical values of the spatial displacement corresponding to this value of $\langle f \rangle$ are near 3 Å.

The calculated value of the signature f for each residue in the first extended state of 4WMY is displayed in Fig. 1(a). A visual representation of this state of 4WMY is presented in Fig. 1(b). The color and intensity of individual residues shown in Fig. 1(b) are determined by the value of $\langle f \rangle$ reported in Fig. 1(a). For example, a dark / light green shading of a residue denotes a value of $\langle f \rangle$ in the low / high end of the color range. Presented in Figs. 2(a) and 2(b) are analogous plots for the first extended state in XEEL (4WN0).

The signature β is the difference in degrees between terminal residues in an n -residue segment centered on residue i , and terminal residues in the triplet centered on residue i in the native state, summed over all residues and normalized by the number of residues in the particular protein. By construction, the native state value is $\langle \beta \rangle = 0$, so values of $\langle \beta \rangle$ greater than zero correspond to an angular configuration more extended than that in the native state; values smaller than zero signal a configuration more localized than in the native state.

As noted in the Appendix, two angular spread signatures can be calculated, one for the native state, and a second for a specified extended state. Both will now be illustrated. Displayed in Figures 3–5 are the angular spread signatures for all residues in 4WMY, XEEL (4WN0) and intelectin-1, first for the native state and then for the third extended state. For each protein, the profiles for the native and third extended states are qualitatively similar, but

quantitatively different. Note the similarity between the profiles for 4WMY and intelectin-1. Note further the quantitative difference between the angular signatures of these two proteins and XEEL in the third extended state.

The overall (all residue) average value of the native-state angular spread signature for the three lectin proteins increases systematically as each protein unfolds. See Table 1. In contrast to the modest changes in the radial displacement of α -carbons for individual residues, changes in the signature β for individual residues are much more pronounced [Compare Figs. 1(a) and 2(a) with Figs. 3–5].

In comparing the two human lectins [4WMQ and 4WMY], we find that the overall signatures calculated are essentially the same, and different from those calculated for the frog protein XEEL (4WN0). As we shall find later in this study, this observation, contrasting and distinguishing 4WMQ and 4WMY from 4WN0, will be sustained when considering subsets of residues of these proteins.

As an illustration, we present in Table 2(a) data on the region bracketing the (sequential) residues 25 to 125, for each protein. Table 2(b) focuses on the region bracketed by residues 155–255. These two regions are displayed in Figs. (6) and (7) for the native state of 4WMY and XEEL (4WN0), respectively.

Comparing residues in these two regions, the region defined by residues 25–125 includes ten residues identified as “conserved,” and eleven residues identified as position-equivalent in these two proteins (See Section V.). The second region, residues 155–255, was chosen to explore the (possibly) different resiliency of residues relatively removed from the fixed point, Ca [PDB 401]. Table 2(c) presents a comparison of 4WMQ and 4WMY for the region 213–267 [PDB 243–297], highlighted in Fig. (3) in Kiessling et al., [5–6] and analyzed using our geometrical model [7].

To facilitate comparisons, data in these tables are normalized with respect to the all-residue averages reported in Tables 1(a), 1(b) and 1(c), respectively. Thus, values in the near neighborhood of “1” suggest that the signature being reported is not very different from the all-residue average at that stage of unfolding for the particular protein being considered. This situation describes the behavior of the spatial signatures $\langle T/T_n \rangle$ and $\langle f \rangle$.

Importantly, as seen in Table 2(c), values both greater than and less than “1” can be found for the signature $\langle \beta \rangle$, reflecting the fact that the angular configuration of a segment or region can be significantly different from the average, all-residue angular spread.

IV. Regional Comparisons

In this section, we present a detailed comparison of the human intelectin proteins 4WMQ and 4WMY with the frog protein, XEEL [4WN0] by quantifying the behavior of four distinct structural moieties as the protein unfolds: α -helices, β -sheets, extended non-helical regions, and residue segments within 10 Å of the Ca ion closest to the crystallographic origin.

The metric of 10 Å, noted above, is totally arbitrary. It is the same metric we used in our studies of the cytochromes [8] and amicyanin [9], thereby allowing comparisons of the relative stabilization of residue segments in the neighborhood of the metal ion(s) in different proteins in the same family and in different families. Thus, we can assess the relative influence of the iron ion in the cytochromes, the Cu ion in amicyanin, as well as the Ca ions in the intelectin family.

The core data on which this segment of our study is based are presented in Tables S1(a), S1(b) and S1(c) through Tables S12(a), S12(b) and S12(c). It is impractical to present graphical representations for every α -helical, β -helical and non-helical region reported in these tables. Hence, we shall focus on polypeptide sequences for each structural unit with the largest number of residues: for the α -helix, 14 residues; for β -helices, 7 residues; and, for the non-helical regions, 28 residues.

To facilitate comparisons, the data in these Tables S1(a) thru S12(c) are normalized as described for the data presented in Tables 2(a), 2(b) and 2(c). Before proceeding with the discussion of data on α -helices, β -sheets, etc., we note that in collapsing the sample size from the global regions considered in the previous section to regions in which (here, at least) five residues comprise the structural unit being considered, the computed averages are less regular, reflecting the fact that, owing to nearest-neighbor steric interactions (excluded volume effects), the “personalities” of residues in the near neighborhood of a given residue are (typically) very different from one residue to the next. For the structural moieties considered in this section, overall trends can be discerned. But, as we shall see later when comparing conserved residues in 4WMY and 4WN0, only approximate trends can be identified in studying the destabilization of individual residues, and only for certain signatures.

Data for the α -helices for all three proteins are presented in Tables S1(a), S1(b) and (c) for the signature $\langle T/T_n \rangle$, in Tables S2(a), S2(b) and S2(c) for $\langle f \rangle$, and in Tables S3(a), S3(b) and S3(c) for the native-state angular spread signature β . For the signature $\langle T/T_n \rangle$, the data show only marginal differences between each of the helical regions and the all-residue averages for all three proteins.

With respect to the signature $\langle f \rangle$, wherever there is a direct correspondence between an α -helix in 4WMQ and 4WMY, the results are essentially the same. See, for example, the helical segments 10–18, 43–47, 95–99, 107–112, 127–135, 144–148, 152–160, 230–234 and 270–275. A similar correlation *cannot* be made with the helical segments in XEEL. For the 66 possible states comprising the “phase space” of XEEL, the most one can say is that there is no significant statistical difference between the number of states for which the signature $\langle f \rangle$ is greater than “1”, signaling a state for which the displacement of the helix alpha-carbons is greater than the all-residue average, viz., 42% for XEEL versus 39% for 4WMY (and 4WMQ).

To give a visual sense of the similarities/differences between 4WMY and XEEL, a comparison is presented in Fig. 8 of the signature $\langle f \rangle$ for the helical region 185–199 in

4WMY versus the region 179–193 in XEEL, both segments comprising 14 residues. Figure 9 displays the signature $\langle\beta\rangle$ for the native state for the same helical regions.

Figures 10 and 11 display the similarities/differences between 4WMY and XEEL(4WN0) for residues comprising (one leg of) a β -sheet. The segments 116–123 in 4WMY and 110–117 in XEEL each have 7 residues.

As is evident from the representative cases plotted in Figs. 8–13, whereas *quantitative* differences are found when comparing profiles of the radial signature versus extension stage for 4WMY and 4WN0, the profiles of the angular spread signature versus extension stage for the two proteins can be *qualitatively* different. Examining the complete data set for all α -helical regions and β -sheets for all three proteins, 95% of the possible states have values of $\langle f \rangle$ greater than the all-residue average, i.e., reflect states more displaced, on average, than the totality of residues comprising each protein. As for the angular spread signature, 88% of the states accessible to intelectin-1 (4WMQ) and 4WMY are characterized by $\langle\beta\rangle$ values greater than “1”, indicating that these states are more extended than the all-residue average. The corresponding percentage for XEEL is 82%, a small difference that may contribute to, but is not responsible for, the difference in resiliency observed when considering the protein as a whole.

For non-helical regions of five residues or greater, data on the signatures $\langle T/T_n \rangle$ in Tables S7(a) through S7(c), data on $\langle f \rangle$ in Tables S8(a) through S8(c), and data on $\langle\beta\rangle$ Tables S9(a) through S9(c) show only small differences between the signature $\langle T/T_n \rangle$ and the signature $\langle f \rangle$ for 4WMY and XEEL. But, significant differences *are* found in the values of the angular spread signature for the two proteins. Placing in (approximate) correspondence the residue segment 56–84 in 4WMY with 50–78 in XEEL, down-range angular correlations in XEEL are larger than in 4WMY for shorter residues segments, smaller for larger segments. See Figs. 12 and 13. Placing in (approximate) correspondence the residue segments 161–165 in 4WMY and 155–169 in XEEL, the values and the trends are virtually the same.

The most significant difference is found comparing the >30 residue segments, 236–269 in 4WMY versus 233–263 in XEEL, with the former region much more localized than the latter as the two proteins evolve through six stages of unfolding. Kiessling et al.[6]. highlight the importance of segment 213 to 267 (PDB 243–297) in intelectin-1 in the vicinity of the Ca ion (PDB 403) and the binding site. See their Figure 3 and our subsequent analysis of this region in intelectin-1 using the signatures calculated using the geometrical model [7]. A significant fraction ($\approx 60\%$) of residues in region 213–267 is within region 236–269 in 4WMQ (and 4WMY). Residues 213–267 are in the neighborhood of the dipositive Ca ion (PDB 403) that is surface exposed and directly binds the exocyclic 1,2-diol epitope on the ligand. Given the quantitatively different behavior exhibited by 4WMY and 4WN0 in the region 236–269 [Tables S9(a)–S9(c)], we suggest that these differences may explain why human intelectin-1 has a higher affinity for the ligand ($K_d = 85 \pm 14$ nM) compared to the equivalent carbohydrate recognition domain of XEEL ($K_d = 4.1 \pm 0.5$ μ M). Explaining ligand affinity differences in the past has been problematic, because the trimeric state of the carbohydrate recognition domain and the amino acids in the binding site are conserved, but

the affinities are not similar. The higher resiliency of the human intelectin-1 ligand binding site suggests a more stable, well-formed binding site that can bind the ligand more tightly than the equivalent site in XEEL.

Finally, when comparing the residue segments (five residues or more) within 10 Å of Ca (PDB 401), the scenario described in the preceding paragraph is further corroborated. Placing in (approximate) correspondence the residue region 54–73 in 4WMY and 50–69 in XEEL, and 251–254 and 241–250 in XEEL, we find that both regions in 4WMY are more localized than the corresponding regions in XEEL.

In summary, it seems reasonable to conclude that differences in the overall resiliencies of the human vs frog proteins may be attributed mainly to differences in the non-helical regions, as well as to differences exhibited by residues in the near neighborhood of the Ca ions.

V. Conserved and Position-Equivalent Residues

In this Section, we focus on the resiliency of 10 conserved and/or position-equivalent residues among the total set of conserved residues in XEEL and 4WMY. In 4WMY we identify residues 56(H), 57(E), 58(N), 59(D), 62(G), 65(T), 67(G), 68(D), 103(D), and 252(D). In XEEL we identify residues 50(H), 51(E), 52(N), 53(N), 56(G), 59(T), 61(G), 62(D), 97(D) and 246(D). In both residue sets, the residue number is the sequential assignment. Note that only one pair is not a perfect match [Wangkanont et al. [6] report that “XEEL is homologous (64% identical) to human intelectin-1”. Note also that all residues in both sets, except the last, are in the region 25–125 studied earlier. Data for individual residues in Tables 5(a) and 5(b) for the signature T/T_n , in Tables 6(a) and 6(b) for f , and in Table 7(a) and 7(b) for β , are normalized with respect to the native state triplet, as in Tables 1(a) through 1(c)

As evident from data in Tables 5(a) and 5(b), and Tables 6(a) and 6(b), those for T/T_n and f establish a reasonable correspondence between paired, conserved residues in 4WMY and XEEL. Given the definition of the lateral signature, and the modest radial displacement of each residue as each protein unfolds, this correspondence is a necessary but not sufficient condition to establish the usefulness of our geometrical model in correlating conserved residues.

Finding correspondence between conserved residues is more problematic when the angular spread signature is considered [see Tables 7(a) and 7(b)]. Choosing a “sample size” of one residue can lead to (apparently) anomalous values of β . These anomalies are a consequence of the fact that the near neighborhood of a given residue is different for every residue. Excluded volume effects arising from the stereochemistry of adjacent functional groups can influence the angular spread in a very different manner relative to the native state triplet. This influence is amplified in our model in progressing from five residue segments to fifteen residue segments. Moreover, organizing values of β , smallest to largest, for conserved residues in a given set, we find the ordering does not remain invariant as one considers successively longer-range angular correlations.

Despite these complications, it is reassuring (but not surprising) that when larger sample sets are considered, for example, the residue region 25–125 that encompasses all the residues in the above two sets, local deviations arising from excluded volume effects in a residue sequence are averaged out, and we are left with a final statistical average that behaves sensibly as the protein unfolds.

In conclusion, using our model to establish a correspondence between two conserved residues is rigorously possible only if the functional groups of residues that are *n*th nearest neighbors to the central residue in each modular unit are similar. This condition is approximately satisfied by the conserved pair 57(E) in 4WMY and 51(E) in XEEL; see Table 7(c). In the listing below, the ratio *R* is the value $\beta[57(E)] / \beta[51(E)]$ calculated for the native state.

$$n = 5 [57(E), 51(E); R] = [48.37, 53.11; 1.10]$$

$$n = 7 [57(E), 51(E); R] = [49.58, 58.82; 1.19]$$

$$n = 9 [57(E), 51(E); R] = [78.90, 106.80; 1.35]$$

$$n = 11 [57(E), 51(E); R] = [95.07, 116.10; 1.22]$$

$$n = 13 [57(E), 51(E); R] = [80.68, 103.00; 1.28]$$

$$n = 15 [57(E), 51(E); R] = [69.44, 62.08; 0.89]$$

A value of $\beta[57(E)] / \beta[51(E)] = 1$ would represent a perfect match between the conserved residues 57(E) in 4WMY and 51(E) in 4WN0. Departures from unity reflect the degree to which differences in next nearest-neighbors of the two residues influence the calculated *R* values. The interesting question is whether values calculated for the ratio *R* are (reasonably) uniform. If so, this would suggest that the influence of residues in the extended neighborhood of a given residue is uniform. In the above case, the average value of *R* is $\langle R \rangle = 1.2 \pm 0.2$.

We next study the (possible) correlation of 11 *position-equivalent* residues. We extract from Tables 10, and present in Table 11, data for residue pairs. The data for the first of these pairs, residue 34(V) in 4WMY and 28(E) in XEEL (4WN0), and their respective ratios are:

$$n = 5 [34(V), 28(E); R] = [12.88, 10.64; 0.82]$$

$$n = 7 [34(V), 28(E); R] = [20.44, 16.56; 0.81]$$

$$n = 9 [34(V), 28(E); R] = [17.81, 14.19; 0.80]$$

$$n = 11 [34(V), 28(E); R] = [20.22, 15.78; 0.78]$$

$$n = 13 [34(V), 28(E); R] = [13.57, 10.70; 0.79]$$

$$n = 15 [34(V), 28(E); R] = [21.56, 17.00; 0.79]$$

In contrast to the case of conserved residues, departure from unity of the ratio

$\beta[34(V)] / \beta[28(E)]$ is effectively uniform, with an overall average $\langle R \rangle = 0.80 (\pm) 0.02$. Residue pairs highlighted in red in Table 11 are similarly correlated. Given the caveat stated in our discussion of conserved residues, it is remarkable that we can differentiate and quantify down-range angular correlations in pairs of position-equivalent residues using crystallographic data on the native state. Inasmuch as the structural signatures are conserved, it is perhaps not surprising that both proteins are trimeric, even though their interfacial sequences are different.

VI. Discussion

We have undertaken a comprehensive study of the structural stability of human intelectin [4WMY] and frog [4WN0 (XEEL)] proteins to complement our earlier analysis of intelectin-1 (4WMQ). The intent here is to explore whether these proteins can be distinguished and differences quantified using the geometrical model presented in earlier work. We find that we are able to document, for example, that the higher resiliency of the human intelectin-1 ligand binding suggests a more stable, well-formed binding site that can bind the ligand more tightly than the equivalent site in XEEL.

From analysis of the data generated using our model, we conclude that differences in the overall resiliency of human intelectin from that of the frog protein may be attributed primarily to differences in the non-helical regions, and to residues in the near neighborhood of the Ca ions. When XEEL crystals were soaked in a SmCl_3 solution prior to diffraction data collection, strong anomalous difference peaks were observed. One such peak coincides with the location of the calcium ion in the ligand binding site (Figure 14). These results indicate that Sm ion can compete with the Ca ion in the ligand binding site, but not compete with calcium ions in the structural calcium site. Indeed after occupancy refinement, replacement of the Ca ion with Sm ion in the ligand binding site yielded Sm ion occupancy of 1, indicating complete substitution of Ca ion with Sm ion. This observation supports the

hypothesis that the structural calcium site is more stable and the ligand binding site of XEEL is more labile.

We present data on conserved and position-equivalent pairs of residues. Of the three signatures T/T_n , f and β calculated for 10 conserved residue pairs and 11 position-equivalent pairs in 4WMY and 4WN0, we find the native-state angular spread signature β is more sensitive in establishing a correlation (or not) between residue pairs. Values of the ratio, $\beta[4WMY \text{ residue}] / \beta[4WN0 \text{ residue}]$, different from unity reflect the stereochemistry of different neighborhoods of a given residue in the two proteins 4WMY and 4WN0. The value of the ratio gauges the influence of down-range angular correlations on the (invariant) triplet centered on each residue. We identify in Section V for both conserved position-equivalent residues the ones for which the influence of neighboring residues is nearly uniform as the parent protein denatures.

The statistical reliability of the predictions of our geometrical model depends on the number of residues in the region being considered. The model is most reliable in describing the all-residue, global behavior of 4WMQ, 4WMY and 4WN0 and the extended sub-regions (25–125, 155–255 and 213–267). Signatures calculated for *individual* residues can be irregular. For example, in a comparison of paired residues, the stereochemistry of side chains of nearby residues can influence the calculated values of the three signatures, particularly the angular spread signature β . Given this limitation, it is noteworthy that for several of the conserved and position-equivalent residues in the human and frog proteins, the ratio $\beta[4WMY \text{ residue}] / \beta[4WN0 \text{ residue}]$ is nearly constant, suggesting that down-range angular correlations for the residue considered are similar in both proteins. In light of our results on conserved and position-equivalent residues in 4WMY and 4WN0, it is perhaps not surprising that both proteins are trimeric, even though their interfacial sequences are different.

A natural question is whether or not there are other proteins with the same conserved and position-equivalent residues surrounding Ca^{+2} regulated ligand binding. Intelectin is a unique family of lectins, as is the environment around the ligand binding site [see Refs.5 and 6]. Our study focuses on structured (α -helical and β -helical) regions and unique non-structured regions of intelectins. While the number of intelectin structures presently available is limited, we believe as crystallographic data on new structures become available, our current work will lay the foundation for further analysis of these unique regions.

In conclusion, we have shown that our geometrical model can be used to quantify the loss of structural stability of the native state and early stages of denaturation of human and frog lectins. Once the model is defined, no further approximations are introduced in applying the model, and the quality of results generated depends only on the quality of crystallographic data and the number of residues used in constructing the averages $\langle T/T_n \rangle$, $\langle f \rangle$ and $\langle \beta \rangle$. Since the only functions needed to implement the theory are sines and cosines, all calculations can be performed on a standard workstation. Most importantly, as with ϕ/ψ plots, the model can be used to identify sensitive regions of configuration space that can then be explored systematically using more sophisticated computational methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

We thank NIH (R01 DK019038 to HBG) for support of work performed at the California Institute of Technology. Financial support for R.A.G.-L. was provided in part by grant # 52007555 to Pomona College from the Howard Hughes Medical Institute through the Precollege and Undergraduate Science Education Program. KW was supported in part by Grants for Development of New Faculty Staff, Ratchadaphiseksomphot Endowment Fund (Grant no. GDNS 59-059-23-020 and DNS 61-011-23-003-2), Chulalongkorn University and the Research Fund for DPST Graduate with First Placement [Grant no. 018/2559] from the Institute for the Promotion of Teaching Science and Technology (IPST), Thailand. Molecular graphics images were produced using the Chimera package from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH P41RR-01081).

Appendix.: Specification of the geometrical model, illustrated using Maplesoft.

Following the recommendation of a reviewer, we present here, step-by-step, the specification and calculation of the signatures of our geometrical model, using the symbolic and numeric programming language MAPLE. At each step we record the MAPLE command used to calculate the quantity being studied.

To link the presentation to our earlier work on intelectin-1, we consider the nine residue segment Glu(61), Asn(62), Gly(63), Val(64), Ile(65), Tyr(66), Gln(67), Thr(68), Phe(69). We choose one of the calcium ions to be the origin of the coordinate system, specifically the Ca ion with signature PDB H401. Relative to the crystallographic origin, the coordinates of this Ca ion in units of Ångstroms are:

$$H401x := 6.842;$$

$$H401y := 21.716;$$

$$H401z := 10.239;$$

A Chimera diagram for this segment of the polypeptide chain is shown in Figure A1. A diagram showing the native-state distances and angles relating residues 61 to 69 to the Ca ion is shown in Figure A2. The calculation of these distances and angles is given below.

Using the reported crystallographic data, we calculate the following distances:

$$R61 := \sqrt{(R61x - Hx)^2 + (R61y - Hy)^2 + (R61z - Hz)^2};$$

$$R69 := \sqrt{(R69x - Hx)^2 + (R69y - Hy)^2 + (R69z - Hz)^2};$$

$$R_{61to69} := \sqrt{(R_{69x} - R_{61x})^2 + (R_{69y} - R_{61y})^2 + (R_{69z} - R_{61z})^2};$$

Evaluation yields the following: $R_{61} = 27.648 \text{ \AA}$, $R_{69} = 19.002 \text{ \AA}$, $R_{61} \text{ to } R_{69} = 18.733 \text{ \AA}$

Note that the Theorem of Pythagoras is *not* satisfied $[(27.65)^2 - (19.00)^2 + (18.73)^2]$, since the theorem requires that the triangle have a right angle. However, the calculated lengths *do* satisfy the generalization of Pythagoras Theorem called the Pappus's Area Theorem which describes the relationship between the areas of three parallelograms attached to three sides of an arbitrary triangle.

The distance between the Ca ion and residue Ile(65), the midpoint residue in the nine residue segment, in the native state is given by:

$$R_{65} := \sqrt{(R_{65x} - H_x)^2 + (R_{65y} - H_y)^2 + (R_{65z} - H_z)^2};$$

Figure A1. Intelectin -1 (PDB ID: 4WMQ). Nine-residue segment Glu (61) to Phe (69)

Figure A2. Native State, third extension. Nine-residue segment Glu(61) to Phe (69) in intelectin-1.

The value is: $R_{65} = 23.957 \text{ \AA}$

We next calculate the three angles, α , β , γ using the Law of Cosines (Proposition 12 in Euclid's *Geometry*), valid for *any* triangle. The commands in Maplesoft are:

β is the angle between residue 61 and residue 69.

$$\text{CBETA61to69} := (R_{61}^2 + R_{69}^2 - R_{61\text{to}69}^2) * (2 * R_{61} * R_{69})^{(-1)};$$

$$\text{BETA61to69} := \text{RAD} * \arccos(\text{CBETA61to69});$$

α is the angle between Ca ion and residue 61:

$$\text{CALPHA61to69} := (R_{69}^2 + R_{61\text{to}69}^2 - R_{61}^2) * (2 * R_{69} * R_{61\text{to}69})^{(-1)};$$

$$\text{ALPHA61to69} := \text{RAD} * \arccos(\text{CALPHA61to69});$$

γ is the angle between residue 69 and the Ca ion:

$$\text{CGAMMA61to69} := (R_{61}^2 + R_{61\text{to}69}^2 - R_{69}^2) * (2 * R_{61} * R_{61\text{to}69})^{-1};$$

$$\text{GAMMA61to69} := \text{RAD} * \arccos(\text{CGAMMA61to69});$$

Here, as must be the case,

$$\alpha + \beta + \gamma = 94.221^\circ + 42.510^\circ + 43.269^\circ = 180^\circ$$

All of the above information is extracted *directly* from the reported data on the crystal structure of intelectin-1.

To determine the change in distance of the alpha carbon of the midpoint residue Ile(65) with respect to the Ca ion as the protein denatures, we adopt as representative of the myriad of states accessible to the native protein one configuration, a fully-extended sequence of nine residues, the third in the six stages of unfolding considered in our study.

The essence of our geometrical model is that we focus on triplet modules. The crystal structure for intelectin-1 encodes exactly all (locally-optimized) attractive and repulsive interactions characterizing each triplet.

The overall length of the extended state is designated T65, and is specified in Maplesoft as:

$$\text{T65} := \text{R61to63} + \text{R63to65} + \text{R65to67} + \text{R67to69};$$

The value is: $\text{T65} = 25.928 \text{ \AA}$

which may be compared with the value calculated for the native state n:

$$\text{R61toR69} = 18.733 \text{ \AA}$$

If one divides T65 by the corresponding length of a triplet centered on residue 65, the resulting ratio is 3.979, a value close to the integer 4, which is the number of triplets in a nine residue segment. This correlation is sustained for all residues, taken one at a time, in the polypeptide chain. [The value of the ratio reported in Ref. [7] for Ile(44) is incorrect; the correct value is 4.027].

The overall average value of the ratio for all residues in intelectin-1 is 4.032. Similar results are calculated at each stage of (linear) extension. That is, the averages calculated are always near integers, reflecting the number of triplet modular units in a polypeptide segment at the given stage of unfolding. Using the notation T_i for the extended state and T_{n_i} for the native state, both metric calculated directly from the crystallographic data, the ratio T_i/T_{n_i} defines the “*lateral signature*” reported in our study.

We have already determined the distance between the Ca ion and residue Ile(65) in the native state, $R_{65} = 23.957 \text{ \AA}$. We next determine the distance between the Ca ion and residue 65 in the third extended state. For this we need the Law of Sines;

$$f_{65} := T_{65} * \sin(\text{ALPHA}_{61\text{to}69}/\text{RAD})/\sin(\text{BETA}_{61\text{to}69}/\text{RAD})/R_{61}.$$

By construction, the value of f calculated for a triplet centered on residue Ile(65) is exactly “one.”

The f_{65} calculated for the nine residue, maximally-extended segment illustrated in Fig.A1 is a length normalized with respect to the native state triplet. The value calculated for f in the third stage of extension is $f_{65} = 1.384$. We note in passing that if, instead of angles $[\beta, \alpha]$, we use the angles $[\beta, \gamma]$ to calculate f_{65}

$$f_{6S} := T_{65} * \sin(\text{GAMMA}_{61\text{to}69}/\text{RAD})/\sin(\text{BETA}_{669}/\text{RAD})/R_{G9};$$

the result obtained is the same, a test of the internal consistency of the method of calculation.

The value of f increases systematically as the protein unfolds. Focusing on Ile(65), the displacement changes from $f=1$ (native state), to 1.030, 1.112, 1.384, 2.034, 2.297 and 3.017 for the first through sixth stages of unfolding, respectively.

The distances corresponding to these displacements are: 24.676 $\text{\AA}$, 26.640 $\text{\AA}$, 33.156 $\text{\AA}$, 48.729 $\text{\AA}$, 55.029 $\text{\AA}$ and 72.278 $\text{\AA}$.

With respect to the displacement of Ile(65) relative to the *native* state, the differences are: 0.72 $\text{\AA}$, 2.6 $\text{\AA}$, 9.2 $\text{\AA}$, 24.8 $\text{\AA}$, 31.1 $\text{\AA}$ and 48.3 $\text{\AA}$.

One of the diagnostics used to assess crystallographic data for a given protein is the α -carbon to α -carbon distance between adjacent amino acids in the polypeptide chain. Averaging over *all* residues in intelectin-1, we calculate for the native state of intelectin-1 the distance 3.807 $\text{\AA}$, a value only slightly different from the theoretical value of 3.8 $\text{\AA}$. When considering stages in the unfolding of the native protein, larger values of the nearest-neighbor, α -carbon to α -carbon distance can occur. Rather than adjust individual residue-to-residue distances to the theoretical value of 3.80 $\text{\AA}$, we adjust the (overall) average α -carbon to α -carbon distance for all nearest-neighbor residues to be 3.8 $\text{\AA}$ at each stage of unfolding.

The adjusted coordinates are denoted by:

$$FA_{65} := [FA_{65x}, FA_{65y}, FA_{65z}];$$

We are interested primarily in the displacement of individual residues in the unfolded state *relative to the native state*. For the midpoint residue Ile(65), this relative metric is defined as

$$fd_{65} := ((FD_x - H_x)^2 + (FD_y - H_y)^2 + (FD_z - H_z)^2)^{1/2}/R_{65};$$

where,

$$FD65 := [R65x - (FA65x - R65x), R65y - (FA65y - R65y), R65z - (FA65z - R65z)];$$

Values of fd for residue Ile(65) at each of the six stages of unfolding considered in this study are given by: 1.52, 1.56, 1.54, 1.45, 1.47, 1.40. As expected, there is a gradual decrease in the value of fd as the protein begins to approximate a “random coil.” In presenting our results, the value calculated for fd is referred to as the “*radial signature*.”

The angular extension β of the sequence, residues 61–69, in the *native* state is, as reported above, 42.510° . The β -angle between the two residues flanking residue Ile(65) in the triplet module, residue Val(64) to residue Tyr(66), is 14.546° . The difference in angular extension of the segment, residues 61–69, in the native state relative to the angular extension of the triplet centered on Ile(65) is given by the Maplesoft command:

$$\text{Beta } 65 := \text{BETA61to69} - \text{BETA64to66};$$

By subtracting out the contribution from the invariant triplet, residues Val(64) to Tyr(66), the difference, $\text{Beta}65 = 27.964^\circ$, isolates the contribution of *down-range* angular correlations in constraining the triplet in the *native* state.

The angular extension of the sequence, residues 61–69, in the third *extended* state, can be determined directly from the crystallographic data as follows. A “new” angle β is calculated using data for the four modular triplet units comprising the fully-extended, nine residue segment

$$\text{ABETA61to69} := \text{BETA61to63} + \text{BETA63to65} + \text{BETA65to67} + \text{BETA67to69};$$

Here, $\text{ABETA}65 = 59.284^\circ$. An expression formally analogous to the above expression for $\text{Beta}65$ can be defined.

$$\text{ABeta}65 := \text{ABETA61to69} - \text{ABETA64to66};$$

Note that the second term on the right-hand side $\text{ABETA}64to66$ is just $\text{BETA}64to66$, since the triplet modular unit in our model remains invariant. Hence, the difference in angular extension between the fully-extended state and the native state triplet, both centered on Ile(65), is

$$\text{ABeta}65 := \text{ABETA61to69} - \text{BETA64to66};$$

The value of $\text{ABeta}65$ calculated for the *extended* state is 44.737° , which can be compared with the difference calculated for the *native* state where $\text{Beta}65$ is 27.964° . In the text, $\text{Beta}35$ (for the *native* state) and $\text{ABeta}65$ (for the *extended* state) are referred to as *angular spread signatures*.

When the first term on the right-hand side of either Beta65 or ABeta65 is specified to be the triplet centered on Ile(65), $\text{Beta65} = \text{ABeta65} \equiv \langle \beta \rangle = 0$. In plots of $\langle \beta \rangle$ versus residue number, the value $\langle \beta \rangle = 0$ references the *native* state triplet, the same reference state that is assigned in calculating the radial signature f (where, for the native state triplet, $f=1$).

Represented in Figure A3 is the extended state for the segment, residues Glu(61) to Phe(69), drawn using the following specifications. As reported above, $\text{ABETA65} = 59.284^\circ$, $T65 = 25.928 \text{ \AA}$; then, $R(\text{Ca to Glu}(61)) = 21.283 \text{ \AA}$ and $R(\text{Ca to Phe}(69)) = 30.809 \text{ \AA}$. Dividing the distance $R(\text{Ca to Ile}(65)) = 36.821 \text{ \AA}$ by $R65 = 23.957 \text{ \AA}$, recovers the *relative extension* of Ile(65) in the third stage of unfolding, $f = 1.54$. Importantly, Figures 2 and 3 represent the exact specification of our geometrical model for Ile(65), in contrast to Fig. 1 of Ref. [7] which is simply a schematic drawing of the relationship between the Ca ion and the residues in the sequence Cys(31) to Arg(39) centered on Leu(35) (and misidentified as the helical segment, Ser(40) to Cys(48), in [7]).

The above Maplesoft commands are implemented for *each* residue at *each* stage in the extension of the polypeptide chain. As is the case for the radial and angular spread signatures calculated for Ile(65) in the native versus third extended state, similar qualitative differences are found between the two signatures for *all* residues in intelectin-1, and in *every* stage of unfolding. The same general conclusions were reached in our earlier studies on the cytochromes and blue copper proteins.

Figure A3. Unfolded State, third extension. Nine-residue segment Glu(61) to Phe (69) in intelectin-1.

REFERENCES

- [1]. Yan J, Xu L, Zhang Y, Zhang C, Zhang Ch., Zhao F, Feng L, Comparative Genomic and Phylogenetic Analyses of the Intelectin Gene Family: Implications for their Origin and Evolution, *Dev Comp Immunol* 41 (2013) 189–199. [PubMed: 23643964]
- [2]. Lee JK, Baum LG, Moremen K, Pierce M, The X-lectins: A New Family with Homology to the *Xenopus Laevis* Oocyte Lectin XL-35, *Glycoconj. J* 21 (2004) 443–450. [PubMed: 15750785]
- [3]. Tsuji S, Uehori J, Matsumoto M, Suzuki Y, Matsuhisa A, Toyoshima K, Seya T, Human Intelectin Is a Novel Soluble Lectin that Recognizes Galactofuranose in Carbohydrate Chains of Bacterial Cell Wall. *J. Biol. Chem* 276(2001) 23456–23463. [PubMed: 11313366]
- [4]. Tsuji S, Yamashita M, Hoffman D, Nishiyama A, Shinohara T, Ohtsu T, Shibata Y, Capture of Heat-Killed *Mycobacterium Bovis* Bacillus Calmette-Guérin by Intelectin-1 Deposited on Cell Surfaces, *Glycobiology* 19 (2009) 518–526. [PubMed: 19179460]
- [5]. Wesener DA, Wangkanont K, McBride R, Song X, Kraft MB, Hodges HL, Zarling LC, Splain RA, Smith DF, Cummings RD, Paulson JC, Forest KT, Kiessling LL, Recognition of Microbial Glycans by Human Intelectin-1, *Nat Struct Mol Biol* 22 (2015) 603–610. [PubMed: 26148048]
- [6]. Wangkanont K, Wesener DA, Vidani JA, Kiessling LL, Forest KT, Structures of *Xenopus* Embryonic Epidermal Lectin Reveal a Conserved Mechanism of Microbial Glycan Recognition, *J Biol. Chem* 291 (2016) 5596–5610. [PubMed: 26755729]
- [7]. Kozak JJ, Gray HB, Garza-López RA, Structural Stability of Intelectin-1, *J. Phys. Chem. B* 120 (2016) 11888–11896. [PubMed: 27784157]
- [8]. Kozak JJ, Gray HB, Garza-López RA, Cytochrome Unfolding Pathways from Computational Analysis of Crystal Structures, *J. Inorg. Biochem* 155 (2016) 44–55. [PubMed: 26606292]

- [9]. Kozak JJ, Gray HB, Garza-López RA, Relaxation of Structural Constraints During Amicyanin Unfolding, *J. Inorg. Biochem* 179 (2018) 135–145. [PubMed: 29222970]
- [10]. Ramachandran GN, Ramakrishnan C, Sasisekharan V, Stereochemistry of Polypeptide Chain Configurations, *J. Mol. Biol* 7 (1963) 95–99. [PubMed: 13990617]
- [11]. Warren JJ, Gray HB, Winkler JR, Kozak JJ, A Euclidean Perspective on the Unfolding of Azurin: Spatial Correlations, *Mol. Phys* 111 (2013) 922–929. [PubMed: 23853392]
- [12]. Warren JJ, Gray HB, Winkler JR, Kozak JJ, A Euclidean Perspective on the Unfolding of Azurin: Angular Correlations, *Mol. Phys* 111 (2013) 3762–3729.
- [13]. Gray HB, Warren JJ, Winkler JR, Kozak JJ, A Euclidean Perspective on the Unfolding of Azurin: Chain Motion, *J. Biol. Inorg. Chem* 19 (2014) 555–563. [PubMed: 24378983]

HIGHLIGHTS

- Study of factors on unfolding of intelectin-1 and *Xenopus* embryonic epidermal lectin
- Predictions of model correlate with reported experimental data
- Geometrical model quantifies the resiliency of the native state to steric perturbations
- Human intelectin-1 has a higher ligand affinity than *Xenopus* embryonic epidermal lectin

Fig. 1a. Radial signature f for each residue in 4WMY in the first extended state, calculated using the geometrical model. Residue listing is sequential from 3 to 275. Alpha-helical residues are in blue, beta helical residues in red, non-helical residues in green, and disulfide bonds in yellow.

Fig. 1b.

Chimera graphic corresponding to Fig. 1a. First extended state of 4WMY. PDB listing for the three calcium are: 401 (red), 402 (green) and 403 (blue), sequential listing 284, 285 and 286, respectively. Residues 25–125 and 155–255 (sequential listing) are in cyan and yellow, respectively. Bar gives the scaling determined from the data in Fig. 2b.

Fig. 2a. Radial signature f for each residue in 4WN0 in the first extended state, calculated using the geometrical model. Residue listing is sequential from 3 to 275. Alpha-helical residues are in blue, beta-helical residues in red, nonhelical residues in green, and disulfide bonds in yellow.

Fig. 2b.
Chimera graphic corresponding to Fig. 2a.

Fig. 3a.
Native-state angular spread signature β for each residue in 4WMY. Color code is the same as Fig. 1a.

Fig. 3b.
Chimera graphic corresponding to Fig. 3a.

Fig. 3c.

Angular spread signature β in the third extension for each residue in 4WMY. Color code is the same as Fig. 1a.

Fig. 4a.
Native-state angular spread signature β for each residue in 4WN0. Color code is the same as Fig. 2a.

Fig. 4b.
Chimera graphic corresponding to Fig. 4a.

Fig. 4c.
Angular spread signature β in the third extension for each residue in XEEL. Color code is the same as Fig. 1a.

Fig. 5a.

An gular spread signature for the native state (top) and third extended state (bottom) for intelectin-1. Color coding is the same as for Fig. 1a. The horizontal line in each is the all-residue average, 27.96 and 71.32, for the native and extended stat respectively.

Fig. 6. Native structure of 4WMY highlighting the (sequential) regions 25–125 (in cyan) and 155–255 (in yellow), other regions in gray. Calcium ions specified in Fig.1a.

Fig. 7. Native structure of 4WN0 highlighting the (sequential) regions 25–125 (in cyan) and 155–255 (in yellow), other regions in gray. Calcium ions specified in Fig.1a.

Fig. 8.
All-residue radial signature $\langle f \rangle$ for residues 185–100 in 4WMY (blue) and residues 179–183 in XEEL (in red). See text.

Fig. 9. All-residue angular spread signature $\langle \beta \rangle$ for residues 185–199 in 4WMY (blue) and residues 179–193 in XEEL (in red). see text.

Fig. 10. All-residue radial signature $\langle f \rangle$ for residues 116–123 in 4WMY (blue) and residues 110–117 in XEEL (in red). See text.

Fig. 11.

All-residue angular spread signature $\langle \beta \rangle$ for residues 116–123 in 4WMY (blue) and residues 110–117 in XEEL (in red). See text.

Fig. 12.

All-residue radial signature $\langle f \rangle$ for residues 56–84 in 4WMY (blue) and residues 50–78 in XEEL (in red). See text.

Fig. 13. All-residue angular spread signature $\langle \beta \rangle$ for residues 56–84 in 4WMY (blue) and residues 50–78 in XEEL (in red). See text.

Figure 14.

A) Structure of XEEL in complex with Ca ions (green). B) Anomalous difference map (magenta mesh) at 6σ of SmCl_3 -soaked XEEL showing a peak in the ligand binding site (arrowhead).

Table 1(a).

Overall (all residue) average of the lateral signature $\langle T/T_n \rangle$ for stages $n=5$ to $n=15$ in the unfolding of three lectins: 4WMQ, 4WMY and 4WN0. Extensions are normalized with respect to the native state triplet (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	283	2.014	3.016	4.032	5.038	6.057	7.016
4WMY	283	2.014	2.997	3.994	4.991	5.988	6.988
4WN0	277	2.014	2.997	4.037	5.043	6.061	7.000

Table 1(b).

Overall (all residue) average of the radial signature $\langle r \rangle$ for stages $n=5$ to $n=15$ in the unfolding of three lectins: 4WMQ, 4WMY and 4WN0. Extensions are normalized with respect to the native state triplet (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	283	1.366	1.366	1.368	1.407	1.403	1.420
4WMY	283	1.364	1.365	1.367	1.407	1.402	1.418
4WN0	277	1.392	1.413	1.387	1.430	1.426	1.446

Table 1(c).

Overall (all residue) average of the angular spread signature $\langle\beta\rangle$ for the native state of three lectins: 4WMQ, 4WMY and 4WN0. Reported is the difference in degrees between the n-residue native state and the triplet native state (See Appendix).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	283	10.09	19.05	25.44	31.10	35.02	37.99
4WMY	283	10.11	19.07	25.46	31.12	35.05	38.03
4WN0	277	10.20	19.06	23.94	27.60	29.52	31.44

Table 2(a).

Average of the lateral signature $\langle T/T_n \rangle$ for stages $n=5$ to $n=15$ in the unfolding of the region defined by residues 25 – 125 in 4WMQ, 4WMY and 4WN0. Extensions are normalized with respect to the all-residue average [Table 1a]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	25–125	1.001	1.001	1.001	1.000	1.000	1.004
4WMY		1.001	1.000	1.001	1.000	0.999	1.005
4WN0		1.001	1.001	1.000	1.002	1.001	1.002

Table 2(b).

Average of the radial signature $\langle f \rangle$ for stages $n=5$ to $n=15$ in the unfolding of the region defined by residues 25–125 in 4WMQ, 4WMY and 4WN0. Extensions are normalized with respect to the all-residue average [Table 1b]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	25–125	1.018	1.016	1.005	0.998	1.002	1.011
4WMY		1.019	1.016	1.005	0.998	1.002	1.011
4WN0		1.013	1.008	0.997	0.985	0.987	0.996

Table 2(c).

Overall (all residue) average of the angular spread signature $\langle\beta\rangle$ for the native state of three lectins, 4WMQ, 4WMY and 4WN0, residues 25–125. Extensions are normalized with respect to the all-residue average [Table 1c]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	25–125	1.313	1.298	1.318	1.329	1.331	1.283
4WMY		1.313	1.299	1.319	1.330	1.332	1.284
4WN0		1.227	1.190	1.178	1.129	1.126	1.087

Table 3(a).

Average of the lateral signature $\langle T/T_n \rangle$ for stages $n=5$ to $n=15$ in the unfolding of the region defined by residues 155–255 in 4WMQ, 4WMY and 4WN0. Extensions are normalized with respect to the all-residue average [Table 1a]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	155–255	0.999	0.999	1.000	0.999	1.000	1.006
4WMY		0.999	0.999	1.000	0.999	1.000	1.006
4WN0		0.999	0.999	0.997	0.997	0.997	0.997

Table 3(b).

Average of the radial signature $\langle f \rangle$ for stages $n=5$ to $n=15$ in the unfolding of the region defined by residues 155–255 in 4WMQ, 4WMY and 4WN0. Extensions are normalized with respect to the all-residue average [Table 1b]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	155–255	0.994	0.992	1.005	0.997	0.988	0.967
4WMY		0.994	0.994	1.004	0.997	0.988	0.966
4WN0		1.000	1.003	1.003	1.003	0.986	0.966

Table 3(c).

Overall (all residue) average of the angular spread signature $\langle\beta\rangle$ for the native state of three lectins, 4WMQ, 4WMY and 4WN0, residues 155–255. Extensions are normalized with respect to the all-residue average [Table 1c]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	155–255	0.892	0.868	0.835	0.795	0.781	0.776
4WMY		0.891	0.867	0.834	0.794	0.779	0.775
4WN0		0.990	0.999	0.950	0.943	0.888	0.880

Table 4(a).

Average of the lateral signature $\langle T/T_n \rangle$ for stages $n=5$ to $n=15$ in the unfolding of the region defined by residues 213–267 in 4WMQ and 4WMY. Extensions are normalized with respect to the all-residue average [Table 1a]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	213–267	1.004	1.004	1.006	1.005	1.006	1.012
4WMY		1.004	1.004	1.005	1.005	1.005	1.012

Table 4(b).

Average of the radial signature $\langle f \rangle$ for stages $n=5$ to $n=15$ in the unfolding of the region defined by residues 213–267 in 4WMQ and 4WMY. Extensions are normalized with respect to the all-residue average [Table 1b]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	213–267	0.982	0.935	0.913	0.899	0.891	0.865
4WMY		0.982	0.935	0.913	0.899	0.891	0.864

Table 4(c).

Overall (all residue) average of the angular spread signature $\langle\beta\rangle$ for the native state of three lectins, 4WMQ, 4WMY and 4WN0, residues 213–267. Extensions are normalized with respect to the all-residue average [Table 1c]. Residues are identified by their sequential listing (see text).

Protein	Residues	n=5	n=7	n=9	n=11	n=13	n=15
4WMQ	155–255	0.863	0.763	0.664	0.592	0.593	0.599
4WMY		0.862	0.762	0.663	0.592	0.593	0.598

Table 5(a).

Lateral signature T/Tn for conserved residues for stages n=5 to n=15 in the unfolding of intelectin (4WMY). Extensions are normalized with respect to the native state triplet. Residues are identified by their sequential listing (difference of 30 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WMY	56(H)	1.807	2.847	3.855	4.508	5.539	6.326
	57(E)	2.019	3.049	3.972	5.092	5.873	7.040
	58(N)	2.241	3.119	4.244	5.315	6.568	7.314
	59(D)	1.662	2.749	3.560	4.574	5.412	6.720
	62(G)	1.970	2.854	4.312	4.894	6.352	6.984
	65(T)	1.664	2.678	3.265	4.225	4.803	6.128
	67(G)	2.029	3.207	3.955	5.423	6.214	7.314
	68(D)	2.087	3.157	4.602	5.480	6.718	7.624
	103(D)	2.150	2.922	4.064	4.582	5.720	6.216
	252(D)	2.271	3.504	4.349	5.320	6.318	7.428

Table 5(b).

Lateral signature T/Tn for conserved residues for stages n=5 to n=15 in the unfolding of XEEL (4WN0). Extensions are normalized with respect to the native state triplet. Residues are identified by their sequential listing (difference of 65 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WN0	50(H)	1.82	2.829	3.849	4.454	5.522	6.284
	51(E)	1.992	3.038	3.902	5.067	5.801	6.982
	52(N)	2.277	3.116	4.337	5.377	6.689	7.380
	53(N)	1.633	2.778	3.566	4.627	5.413	6.758
	56(G)	1.941	2.767	4.203	4.749	6.196	6.803
	59(T)	1.681	2.702	3.338	4.326	4.873	6.192
	61(G)	2.063	3.232	3.984	5.418	6.321	7.353
	62(D)	2.081	3.132	4.491	5.447	6.602	7.569
	97(D)	2.174	2.946	4.094	4.640	5.763	6.277
	246(D)	2.298	3.342	4.369	5.328	6.367	7.406

Table 6(a).

Radial signature f for conserved residues for stages $n=5$ to $n=15$ in the unfolding of intelectin (4WMY). Extensions are normalized with respect to the native state triplet. Residues are identified by their sequential listing (difference of 30 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WMY	56(H)	1.523	1.582	1.620	1.700	1.708	1.733
	57(E)	1.523	1.530	1.635	1.658	1.682	1.690
	58(N)	1.412	1.560	1.556	1.613	1.610	1.667
	59(D)	1.477	1.433	1.479	1.517	1.569	1.579
	62(G)	1.415	1.204	1.024	0.994	0.952	1.053
	65(T)	1.457	1.094	1.482	1.388	1.387	1.335
	67(G)	1.372	1.213	1.223	1.228	1.367	1.449
	68(D)	1.512	1.271	0.309	1.049	1.232	1.317
	103(D)	1.488	1.471	1.384	1.508	1.469	1.493
	252(D)	0.825	0.269	1.182	1.359	1.305	1.351

Table 6(b).

Radial signature f for conserved residues for stages $n=5$ to $n=15$ in the unfolding of XEEL (4WN0). Extensions are normalized with respect to the native state triplet. Residues are identified by their sequential listing (difference of 65 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WN0	50(H)	1.544	1.616	1.632	1.713	1.721	1.749
	51(E)	1.551	1.565	1.650	1.669	1.697	1.706
	52(N)	1.434	1.597	1.567	1.627	1.627	1.688
	53(N)	1.505	1.469	1.495	1.531	1.593	1.603
	56(G)	1.449	1.263	1.049	1.022	1.002	1.102
	59(T)	1.480	1.165	1.496	1.416	1.421	1.371
	61(G)	1.403	1.279	1.250	1.279	1.398	1.496
	62(D)	1.535	1.341	0.478	1.088	1.296	1.368
	97(D)	1.509	1.509	1.404	1.527	1.502	1.528
	246(D)	0.912	0.459	1.223	1.393	1.331	1.353

Table 7(a).

Angular spread signature β for conserved residues for the native state of intelectin (4WMY). Reported is the difference in degrees between the n-residue native state and the triplet native state (See Appendix). Residues are identified by their sequential listing (difference of 30 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WMY	56(H)	36.08	70.40	63.58	68.34	86.63	68.85
	57(E)	48.37	49.58	78.90	95.07	80.68	69.44
	58(N)	41.68	79.75	114.3	107.3	95.20	116.6
	59(D)	5.881	73.52	85.21	105.0	107.2	91.52
	62(G)	29.65	28.77	47.35	86.57	125.7	73.40
	65(T)	18.89	2.705	63.34	65.24	57.69	70.27
	67(G)	25.57	29.83	18.33	25.26	32.51	41.71
	68(D)	29.98	-1.926	-35.84	-18.48	-11.72	-10.45
	103(D)	28.69	18.56	4.610	9.756	12.44	5.678
	252(D)	-15.06	-23.03	-1.680	6.347	-2.820	-4.987

Table 7(b).

Angular spread signature β for conserved residues for the native state of XEEL (4WN0). Reported is the difference in degrees between the n-residue native state and the triplet native state (See Appendix). Residues are identified by their sequential listing (difference of 30 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WN0	50(H)	23.87	68.57	80.99	111.1	130.9	112.9
	51(E)	53.11	58.82	106.8	116.1	103.0	62.08
	52(N)	23.06	73.57	94.17	72.64	32.13	44.28
	53(N)	34.00	72.87	63.77	30.56	26.66	2.980
	56(G)	73.94	42.67	12.27	7.000	-3.114	-18.16
	59(T)	56.03	62.99	107.3	69.04	90.90	54.54
	61(G)	19.22	20.01	35.55	59.77	80.80	105.8
	62(D)	39.31	40.24	-6.231	35.10	43.77	63.16
	97(D)	23.38	19.29	12.78	17.73	21.54	17.65
	246(D)	-13.51	-27.08	5.417	18.13	-0.939	-7.902

Table 8(a).

Lateral signature T/Tn for position-equivalent residues for stages n=5 to n=15 in the unfolding of intelectin (4WMY). Extensions are normalized with respect to the native state triplet. Residues are identified by their sequential listing (difference of 30 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WMY	34(V)	2.006	3.020	3.942	5.279	6.094	7.459
	37(O)	2.141	2.946	4.051	4.706	5.697	6.455
	59(D)	1.662	2.749	3.560	4.574	5.412	6.720
	61(R)	2.056	3.346	4.129	5.784	6.519	7.955
	74(O)	2.114	3.325	4.497	5.238	6.403	7.209
	79(V)	2.125	3.394	4.307	5.858	6.357	7.886
	90(Y)	2.349	3.264	4.544	5.427	6.738	7.321
	97(E)	2.114	3.359	4.453	5.837	7.091	8.282
	101(S)	2.157	3.038	4.286	5.076	6.422	7.192
	110(Y)	1.937	2.975	3.809	5.338	5.938	7.705
	125(K)	2.246	3.332	4.523	5.368	6.781	7.646

Table 8(b).

Lateral signature T/Tn for position-equivalent residues for stages n=5 to n=15 in the unfolding of XEEL(4WN0). Extensions are normalized with respect to the native state triplet. Residues are identified by their sequential listing (difference of 65 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WN0	28(E)	2.089	3.082	4.062	5.340	6.253	7.554
	31(Q)	2.105	2.963	3.975	4.686	5.618	6.412
	53(N)	1.633	2.778	3.566	4.627	5.413	6.758
	55(F)	2.021	3.310	4.031	5.700	6.384	7.842
	68(Q)	2.153	3.365	4.503	5.314	6.500	7.358
	73(Q)	2.093	3.311	4.220	5.717	6.240	7.742
	84(Y)	2.405	3.336	4.600	5.586	6.765	7.546
	91(E)	2.167	3.350	4.582	5.828	7.294	8.309
	95(S)	2.185	3.086	4.353	5.170	6.545	7.258
	104(Y)	1.936	3.014	3.875	5.416	6.060	7.813
	119(K)	2.245	3.304	4.453	5.305	6.679	7.536

Table 9(a).

Radial signature $\langle f \rangle$ for position-equivalent residues for stages $n=5$ to $n=15$ in the unfolding of intelectin (4WMY). Extensions are normalized with respect to the native state triplet. Residues are identified by their sequential listing (difference of 30 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WMY	34(V)	1.516	1.497	1.338	1.135	1.135	1.311
	37(O)	1.513	1.547	1.596	1.658	1.648	1.627
	59(D)	1.477	1.433	1.479	1.517	1.569	1.579
	61(R)	1.346	1.314	1.301	1.306	1.196	1.276
	74(O)	1.216	1.342	1.486	1.530	1.391	1.535
	79(V)	1.246	1.321	1.118	0.946	1.242	1.081
	90(Y)	1.437	1.386	1.321	1.565	1.571	1.680
	97(E)	1.160	1.295	1.347	1.334	1.267	1.310
	101(S)	1.018	1.041	1.289	1.055	1.224	1.364
	110(Y)	0.942	1.403	1.477	1.538	1.570	1.487
	125(K)	1.376	1.245	1.423	1.343	1.261	1.333

Table 9(b).

Radial signature $\langle f \rangle$ for position-equivalent residues for stages $n=5$ to $n=15$ in the unfolding of XEEL (4WN0). Extensions are normalized with respect to the native state triplet. Residues are identified by their sequential listing (difference of 65 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WN0	28(E)	1.533	1.530	1.351	1.373	1.172	1.355
	31(Q)	1.535	1.583	1.614	1.667	1.666	1.653
	53(N)	1.505	1.469	1.495	1.531	1.594	1.603
	55(F)	1.374	1.363	1.323	1.326	1.235	1.320
	68(Q)	1.312	1.393	1.515	1.555	1.438	1.566
	73(Q)	1.287	1.394	1.162	0.967	1.291	1.12
	84(Y)	1.451	1.438	1.358	1.583	1.591	1.693
	91(E)	1.250	1.380	1.368	1.368	1.311	1.371
	95(S)	1.056	1.117	1.313	1.091	1.269	1.393
	104(Y)	0.994	1.445	1.500	1.554	1.586	1.515
	119(K)	1.414	1.308	1.449	1.360	1.292	1.354

Table 10(a).

Angular spread signature β for position-equivalent residues for the native state of intelectin-1 (4WMY). Reported is the difference in degrees between the n-residue native state and the triplet native state (See Appendix). Residues are identified by their sequential listing (difference of 30 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WMY	34(V)	12.88	20.44	17.81	20.22	13.57	21.56
	37(O)	16.54	25.87	34.11	36.70	36.65	36.22
	59(D)	5.881	73.52	85.21	105.0	107.2	91.52
	61(R)	14.81	56.01	66.00	121.2	130.3	110.3
	74(O)	0.595	6.203	22.96	14.32	2.388	60.57
	79(V)	9.033	26.49	25.42	24.13	39.16	39.10
	90(Y)	23.74	43.09	54.07	75.15	89.57	116.5
	97(E)	6.924	38.61	55.19	52.81	35.70	26.80
	101(S)	-7.472	7.633	19.99	3.214	8.739	12.36
	110(Y)	2.642	18.53	27.40	30.97	42.38	50.08
	125(K)	16.86	17.22	26.52	11.99	24.48	21.44

Table 10(b).

Angular spread signature β for position-equivalent residues for the native state of XEEL (4WN0). Reported is the difference in degrees between the n-residue native state and the triplet native state (See Appendix). Residues are identified by their sequential listing (difference of 65 from the PDB listing).

Protein	Residue	n=5	n=7	n=9	n=11	n=13	n=15
4WN0	28(E)	10.64	16.56	14.19	15.78	10.70	17.00
	31(Q)	12.49	20.72	27.20	29.75	30.68	30.61
	53(N)	34.00	72.88	63.77	30.57	26.66	2.981
	55(F)	41.88	61.21	54.28	14.62	-14.44	-28.19
	68(Q)	8.455	15.70	33.46	27.54	17.07	50.24
	73(Q)	6.487	22.83	23.98	18.14	20.34	13.04
	84(Y)	17.02	27.55	31.33	49.36	55.75	80.68
	91(E)	4.850	24.32	33.79	28.16	24.00	22.12
	95(S)	4.873	4.612	13.92	3.255	8.387	8.694
	104(Y)	2.253	14.91	22.02	26.29	33.10	35.36
	119(K)	17.38	16.38	24.27	21.76	24.33	22.33

Table 11.

Ratio R of the angular spread signature β for position-equivalent residues for the native state of 4WMY and XEEL (4WN0). Extensions are normalized. Residues are identified by their sequential listing.

Residue Pair	n=5	n=7	n=9	n=11	n=13	n=15
4WMY 34(V)/XEEL 28(E)	0.82	0.81	0.80	0.78	0.79	0.79
4WMY 37(Q)/XEEL 31(Q)	0.76	0.80	0.80	0.81	0.84	0.85
4WMY 59(D)/XEEL 53(N)	5.78	0.99	0.75	0.29	0.25	0.03
4WMY 61(R)/XEEL 55(F)	2.83	1.09	0.82	0.12	-0.11	-0.26
4WMY 74(Q)/XEEL 68(Q)	14.1	2.53	1.46	1.92	7.14	0.83
4WMY 79(V)/XEEL 73(Q)	0.72	0.86	0.94	0.75	0.52	0.33
4WMY 90(Y)/XEEL 84(Y)	0.72	0.64	0.58	0.66	0.62	0.69
4WMY 97(E)/XEEL 91(E)	0.70	0.63	0.61	0.53	0.67	0.83
4WMY 101(S)/XEEL 95(S)	0.65	0.60	0.70	1.02	0.96	0.70
4WMY 110(Y)/XEEL 104(Y)	0.85	0.80	0.80	0.85	0.78	0.71
4WMY 125(V)/XEEL 119(E)	1.03	0.95	0.92	0.53	0.99	1.04