Steered Molecular Dynamics Study of the Age-related Stiffening

of the Crystalline Lens

by Ewa Golas, Leslie Glasser*, and Harold A. Scheraga

Ewa Golas

Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland

Baker Lab of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, U.S.A

Current address: *Institute of Physics, Polish Academy of Sciences*
Al. Lotnikow 32/48, 02-668 Warsaw Poland

Leslie Glasser*

Nanochemistry Research Institute, Department of Chemistry,

Curtin University, GPO Box U1987, Perth, WA 6845, Australia

Harold A. Scheraga

Baker Laboratory of Chemistry and Chemical Biology,

Cornell University, Ithaca, New York 14853-1301, U.S.A.

17 Figures

Keywords: lens, cataract, crystallin, stretch, isomerize, snag-point, mechanical clamp

[*]Corresponding author

Leslie Glasser:

Nanochemistry Research Institute, Department of Chemistry,

Curtin University, GPO Box U1987, Perth, WA 6845, Australia

Telephone: + 61 8 9848-3334

E-mail: leslieglasser@yahoo.co.uk   l.glasser@curtin.edu.au

Abstract

αA-crystallin is a key component of the glassy solution of proteins that constitutes the mammalian lens. It contributes to the refractive and mechanical properties of the lens, and as a member of the small heat shock protein (sHSP) family of chaperones, plays a role in aggregate prevention. Age-dependent L- to D- racemization of amino acids in the sequence of the protein has been implicated in lens stiffening and cataract, and is suspected to interfere with the protein's basic chaperone activity and structural features. This communication investigates the mechanical properties of bovine αA-crystallin and several of its (point) D-isomerized derivatives by way of Steered Molecular Dynamics simulation. In a series of induced unfolding experiments, an external pulling force is applied to the native protein and, independently, to three D-amino acid variants. A principal component-based technique is applied to extract dominant structural and mechanical features from the system variants. The D-isomerization of a single residue in the structure of αA-crystallin alters the protein's unfolding pathway, and changes the mechanical properties of its inherent elements of (secondary) protein structure. The location of the D-substituted residue is critical to defining the extent and nature of the observed effects. The latter are expressed as divergence from the typical native induced unfolding pathway and altered structural element stiffnesses.

Introduction

Presbyopia (Greek "elderly vision") is the age-related loss of near-focussing ability in a normal young eye and is caused by increased stiffness of the lens, while cataract is the accumulation over time of light-scattering aggregates in the eye lens, also accompanied by lens stiffening[1]. These are serious and expensive problems. Presbyopia is an inevitable accompaniment of aging, while an estimated 20 million people in the world suffer from cataracts, which also increase in prevalence with age and are a leading cause of blindness.

The lens contains an extraordinarily high concentration of protein in solution, principally crystallins (the various types are labelled αA and αB, β, and so on). The most prevalent are the α-crystallins which combine to form large complexes.[2] A feature of the crystallins is that they act as chaperones (and are part of the small *Heat Shock Protein* [sHsp] family) thus binding to damaged and unfolded proteins, thereby generally mitigating the effects of the formation of light-scattering aggregates.

Various alterations develop in the crystallins with both age and heat, of which deamidation and racemization appear to be the most important, and it is suspected that it is racemization which may be associated with the lens stiffening.[2] Recent work[3] has identified the principal amino acid residues undergoing racemization as serine (Ser), asparagine (Asp/Asn), and threonine (Thr). Study of the effects of these changes may assist in providing an understanding of the molecular basis of presbyopia and thereby lead to possible treatments. This is the purpose of the current communication.

Materials and Methods

Structure[4] and Contents[5] of the Lens of the Human Eye

The lens consists of a bundle of fibers, lined up closely together and stretching from front to back, in layers like those of an onion. These lens fiber cells are filled with a gel-like concentrated solution of crystallins (~70%) – of which the most common (~50%) is α-crystallin, followed by decreasing proportions of β- and γ-crystallins.

The α-crystallins are very large oligomers assembled from the two types (A and B) of sequence-related 20 kDa subunits. The β-crystallins are also oligomeric but more complex, consisting of up to seven sequence-related polypeptides, ranging in size from about 20 to 30 kDa. The γ-crystallins are a family of 21 kDa monomeric proteins, sharing about 30% sequence identity with the β-crystallins; γ-crystallin acts as a weak glue to gently bind the α-crystallins together.

The initial embryonic and fetal crystallins remain for life, and form the central part of the lens. They are supplemented by crystallins which accumulate over time at the edges of the lens. No other biological components remain in the central lens (nucleus), so that the nucleus is "the realm of chemistry rather than of biochemistry".[2] The chemistry which operates is such as truncation, isomerisation, racemization, phosphorylation, deamidation and oxidation.

These modified proteins would tend to aggregate were it not for the chaperonage of the α-crystallins. With age, however, the available free α-crystallins diminish to zero in the lens

nucleus by about age 40 - after which the initially soluble lens contents accumulate an insoluble fraction of protein aggregates.

Any aggregation brings about increase in the rigidity and opacity of the lens, resulting in presbyopia and cataract formation. Oxidation of the lens proteins also contributes to the darkening of the nuclear lens. An important observed change in the crystallins is L- to D-racemization over time, with the formation of significant proportions of D-Ser, D-Asp/Asn and D-Thr. Larger proportions of the racemized amino acids occur in lenses having cataracts.[6]

However, it is by no means clear whether all the changes in the proteins produce deleterious effects. Indeed, given the high selective pressure of visual acuity, some post-translational modifications are likely to be of adaptive value. The most obvious possibility is that some are required for protein packing in the water-poor center of the lens. Changes that result in the altered behavior of α-crystallin due to the application of force are also significant; in flexing of the lens (as in focal accommodation), proteins within must be capable of functioning under and responding to mechanical stress. Clearly, certain regions of the protein can be envisaged as being either stiffer or more flexible than others, and likewise, as more or less susceptible to producing detrimental effects. As such, a deleterious effect upon D-isomerization within the lens should be highly location-dependent. Following such reasoning, a simple albeit effective method for testing the isomer-dependent mechanical properties of α-crystallin would entail D-amino acid substitution and following it by testing with the application of force. However, performing such an experiment *in-situ* in the lens is not feasible, due to excessive crowding and an inability to control the numerous proteins and the environmental conditions within. Nonetheless, our study

of the behavior of simple models may be used to assess the mechanical features of α-crystallin from a fundamental perspective.[7,8] The latter is a more specific procedure and is implemented in the present work.

αA-crystallin is simulated here *in silico* by application of an external stretching force by Steered Molecular Dynamics (SMD). Steered molecular dynamics simulations, or force probe simulations, apply forces to a protein in order to manipulate its structure by pulling it along desired degrees of freedom. These experiments can be used to reveal structural changes in a protein at the atomic level. SMD is often used to simulate events such as mechanical unfolding or stretching.[9-11]

First, the native structure is stretched. This is followed by stretching versions of the native structure, each of which has been modified by altering ('racemizing') single Ser, Asp, or Thr L-amino acid residues to their D-versions. The detailed differences in the stretching pathways are examined in order to elucidate the consequences of the 'racemizations'. For this purpose, a principal component-based technique has been developed to correlate alterations in the secondary structural elements in response to the applied force during the SMD. Links between the amino acid residues are identified, enabling a quantitative description of the α-crystallin organization. The 'stiffness' of particular zones and structures within α-crystallin are assessed. Furthermore, the induced unfolding of α-crystallin carries with it an unfolding pathway, that allows for an investigation of the reproducibility of the protein's behavior in its native state and following D-isomerization. The potential for reproducible behavior is an important aspect in the lens of the eye, where coherence at some level must operate to ensure recovery and continued

functionality. The current study, therefore, exposes mechanical phenomena at a localized level, and pathway-dependent behaviors that are D-isomer dependent.

Computational Methods

The crystal structure of αA-crystallin from *Bos tauru*s (pdb code 3L1E) was used as an initial template to form the four systems investigated; this structure is a fragment, and describes αA-crystallin over residues 59-163. These generated systems encompass: (i) the **native** protein without any single-point amino-acid isomerizations; (ii) the **S111** system, which bore a racemized D-serine residue at position 111; (iii) the **S130** system, which was similarly modified to carry D-serine at residue 130; and (iv) system **T140**, which contained a racemized D-threonine at position 140. The locations of these residues in the protein are shown in Figure 1.

Figure 1

Each native and racemized system was modeled and energy-minimized using the AMBER12 all-atom model and force field,[12,13] and all calculations were executed on the Beowulf cluster at the University of Gdansk, Poland. Each system was simulated under conditions that employed a Generalized Born implicit solvent,[10] at a salt concentration of 0.150 M. An initial 5 000 steps in the minimisation employed a steepest descent algorithm,[14] while a further 2 000 steps used the conjugate gradient scheme.[15] Minimized systems were initially heated to the desired temperature over a sequence of 5 000 molecular dynamics (MD) steps, after which thermal equilibration ensued for another 210 000 steps. Subsequent Langevin MD simulations[16] employed a 2 fs time step executed at a temperature of 300 K. At this point, each of the four system types was branched to produce twelve independent trajectories, obtained by

resetting system velocities and running 100 000 steps of MD at a fixed temperature of 25°C. Thus, 48 trajectories were prepared in total.

Steered Molecular Dynamics simulations[12] were then performed for each of the 48 trajectories. An external pulling force was applied that stretched the initial $C^{\alpha \cdots} C^{\alpha}$ distance of ~30 Å between residues 60 (Gly) and 148 (Ser). This occurred via AMBER12's constant velocity SMD algorithm, stretching at a rate of 1 mÅ/fs to a total final stretch distance of 144 Å over 80 000 steps. Applied force, total work, and distance information were recorded every 25 iterations; system conformations were collected every 100 steps (generating a total of 800 snapshots per trajectory). Force curves were smoothed via a tandem pair of Savitzky-Golay filters [17] of order 2 and 3, with an averaging window span of 100 data points.

Analysis Methods I

A novel matrix-based Principal Component Analysis (PCA) technique was developed to identify 'structural links' as described below. It can be thought of as a cross between Essential Dynamics and a traditional PCA (where Essential Dynamics is the application of PCA to the time-dependent coordinates of a biomolecule). Traditional Essential Dynamics was not used because the dominant principal components in this case delineate the obvious trend of elongation along the stretch axis.

Figure 2

A 'structural link' is defined as a pair of C^{α} atoms whose distance of separation is less than 6 Å, but omitting adjacent C^{α} atoms (which are covalently bonded). This distance reflects elements of secondary structure or organization within the protein structure, and so retrieves C^{α}

atoms whose residues are involved in α-helices, β-sheets, and/or other residue-residue interactions. Figure 2 portrays the structural links found for the native αA-crystallin; an inset in the Figure focuses on a case exemplifying a region of β-sheet (B), where it can be seen that the use of structural links effectively locates the C^α pairs that constitute the sheet, producing a 'ladder-like' pattern of links along the long axis of the β-sheet. Likewise, residues interacting to form the turn of the α-helix forming at the terminus of the protein are also identified in this figure (C) by appropriate links.

For a given trajectory, the structural links present in each conformational snap-shot are recorded. The unfolding process, as induced by mechanical stretching, results in the breaking of structural links, but also produces new structural links. Although the number of possible structural links over the course of a single trajectory depends on its individual progression, the systems simulated generally bore approximately 250-300 structural links. For an individual trajectory, *P*, a reference database of non-repeated possible structural links is prepared. The structure of the protein in a snap-shot extracted from the simulation can therefore be summarized by a one-dimensional vector, whose individual entries are either 1 or 0, and reflect the presence or absence, respectively, of a particular structural link from the reference set. The inclusion of an entry, expressing the force for a given frame in effect at a given time, produces a **snapshot vector** of the form:

$$\vec{v_a}(P) = \{f_a, i_1, i_2, i_3, \dots, i_n\} \qquad [1]$$

where f_a represents the standardized force(which ensures that the magnitude of the force is comparable to the 0 or 1 binary link values) recorded for the ath snapshot, and each subsequent

term is the binary value (0 or 1) for each of the n links of the reference set. Accordingly, the whole trajectory, P, can be described by a set of snapshot vectors:

$$\vec{V}(P) = \{ \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_z \} \qquad [2]$$

In this case, z is the total number of conformational frames recorded during the trajectory, and is equal to 800 for all systems.

For the trajectory P, a covariance matrix, C_P, is formed with dimensions $\overrightarrow{v_a}(P) \times \overrightarrow{v_a}(P)$ or equivalently, $(1 + S_P) \times (1 + S_P)$, where S_P is the length of the link reference set of P (see Figure 3).

$$C_P = cov(\vec{\imath}) = \langle (\vec{\imath}_m - \langle \vec{\imath} \rangle)(\vec{\imath}_n - \langle \vec{\imath} \rangle)^T \rangle \qquad [3]$$

Figure 3

Here, averaging occurs over all time frames of $\vec{V}(P)$, with $\vec{\imath}$ representing either a link or force column vector, and with n being equal to m, the number of elements, along the diagonal. C_P presents the system's variation in link identity and force (see Fig. 3). The eigenvectors of the matrix C_P provide the principal components or dominant trends in the system over the course of the simulation, in terms of correlated links and associated force. Eigenvectors are of the same length as the snapshot vectors, $\vec{v}_\imath$, used to generate C_P, and their coefficients address the contribution of the force and each subsequent structural link to the principal component inspected. Eigenvalues reflect the importance of a trend vector in describing the system.

From a mechanical perspective, a correlation is sought between the structure of the protein and high force (i. e., moments of maximal force over the course of the simulation). Components of interest should therefore bear a significant force coefficient; eigenvectors from C_P are extracted for inspection according to the **force criterion**, namely that the force coefficient

for a selected vector lies at or outside the standard deviation for the distribution of force coefficients from the components of C_P that account for 70% variance. Accordingly, the subset of eigenvectors satisfying the force criterion (g_a, g_b, …, g_z) from trajectory P form the set:

$$G_P = \{g_a, g_b, …, g_z\} \qquad [4]$$

Corroborating the significance of this set to a description of the behavior of trajectory P, and to the importance of force as an element therein, is the observation that the members of G_P generally identify with the first 1 to 10 components of C_P (where the first ten vectors account for over 70% of the system variance).

By analogy, the structural links most associated with stretch conditions of high force are then selected from the member vectors, g_i, of G_P by the **hot criterion**, which identifies a structural link as 'hot' if its coefficient lies at or outside the standard deviation for the distribution of the coefficients in g_i. The set of hot links from vector g_a forms the set:

$$h_P(g_a) = \{l_1, l_2, …, l_n\} \qquad [5]$$

And accordingly, all hot links across all the g_z vectors of G_P form a union:

$$H_P = h_P(g_a) \cup h_P(g_b) … \cup h_P(g_z) \qquad [6]$$

A collection based on system type forms:

$$H_S = H_1(g_a) \cup H_2(g_b) … \cup H_{12}(g_z) \qquad [7]$$

S signifies one of the four system types (native, S111, S130, or S140), and the subscripts 1 to 12 include the 12 simulated trajectories.

Collecting the 'hot' links across all trajectories for a given system type, as per H_S above, is a useful exercise, and enables a general picture to be drawn of the dominant links for each of

the four system types. By considering only those links whose frequency across H_S is significant, each system of αA-crystallin may be expressed in terms of a characteristic set of hot links:

$$L_S = \{w_1 l_1, w_2 l_2, \dots, w_n l_n\} \qquad [8]$$

The weight w_i reflects the frequency of hot link l_i in the system, and is expressed as $w_i = \frac{\sum n_i}{Max(\sum N)}$, where $\sum n_i$ is the number of times the hot link appears in H_S, and $Max(\sum N)$ is the maximal count of any hotlink in H_S. A hot link l_i was considered only as a set member of L_s if $w_i \geq 0.50$. This latter restriction, in fact, encompassed 50-60% of the hot links from H_S, corroborating the suitability and potential of L_s as a descriptive and characteristic tool. The response of a particular system to high force can therefore be expressed in terms of link descriptors which are, in turn, associated with particular structural elements in the protein.

The **lifetime** of a particular hot link for a trajectory P may be defined as the number of 0 to 1 or vice-versa existence transitions that occur as a fraction of 400, which represents the total possible maximal number of exchanges should the link exist over every other frame for the 800 total frames recorded. The hot links of L_s may therefore also be expressed in terms of the (weighted) mean frequency with which they break or form during a trajectory. The weighted link lifetimes from system S, T_S , may be expressed as follows.

$$T_S = \{w'_1 l_1, w'_2 l_2, \dots, w'_n l_n\} \qquad [9]$$

In this case, the weight w'_i represents the average link lifetime over all the n_i occurences of link i in H_S. The lifetime contributes to a description of the degree of 'stiffness' for a given hot link, and expresses whether the link has a tendency to re-iteratively form or to be cleanly broken.

Analysis Methods II

In addition to the above-mentioned matrix method, trajectories were also analyzed by other tools, as described here. Instead of a trend-based technique (and hence one that is based on eigenvectors), the current section addresses time-dependent strategies that focus on the unfolding pathway.

Each trajectory was aligned in a sequential algorithm that was dependent on the structure of the frame that preceded it. In this scheme, the stretch center (the midpoint of C^α atoms from residue 60 and 148) between two adjacent frames is superimposed, and their stretch axes (the axis formed by the C^α atoms of residues 60 and 148) are aligned. Rotation along the stretch axis of the later structure with respect to its predecessor is applied so as to minimize the RMSD between the two species. This alignment technique was employed since it both conserves the stretch axis and disfavors alignment to conserved elements of secondary structure in cases where these remain intact despite significant re-orientation or displacement (e.g. detachment) along the axis during stretching.

As a stretching trajectory P proceeded, the change in RMSD of the structure of αA-crystallin was monitored, producing a plot as in Figure 4.

Figure 4

Two Savitzky-Golay filters were applied in tandem (of polynomial order 2 and 4, respectively), using a window width of 100 frames. The minima are considered as quasi-states or 'snag-points' of the structure along the pathway (elsewhere described as "mechanical clamps"[8]) ,

13

and their coordinates were recorded. The resultant collection of snag-point structures, across all trajectories and systems, was analyzed by two methods.

Unfolding Pathway Analysis I

All snag-point structures were aligned to each other in a pairwise fashion. Those with an RMSD of less than or equal to 4.0 Å were selected for clustering, which occurred in a hierarchical fashion. Clusters were classed as native-, S111-, S130-, or T140-like if $\geq$ 60% of its members originated from the same system type; failing this, mixed clusters were identified as being of 'mutual-type'. A 'null' cluster was formed to classify unrelated structures that did not belong to any cluster. Each trajectory P was then expressed as a string of classified quasi-states:

$$Q_P = \{q_1, q_2, \dots, q_n\} \qquad [10]$$

where q_i represents a species belonging to one of the null, mutual, native-, S111-, S130, or T140-like clusters.

Unfolding Pathway Analysis II:

A structural-link based analysis was employed to compare the quasi-states across all trajectories and systems. To this end, each quasi-state n from trajectory P was described by the vector:

$$\vec{r}_{P_n} = \{1\ i_1, 2\ i_2, 3\ i_3, \dots, m\ i_m\} \qquad [11]$$

where m is the total number of unique possible structural links over all 48 trajectories simulated, and i takes on a binary value (either 0 or 1) for either the presence or absence, respectively, of the ith structural link. Quasi-state vectors, $\vec{r}_n$, were then clustered according to the distance function:

$$dist = \frac{|\vec{r}_{P_a} - \vec{r}_{P_b}|}{m^2}$$ [12]

Clusters were classified as either native-, S111-, S130-, or T140-like if $\geq$ 60% of cluster members originated from the same system type. The mutual-type classification encompassed mixed clusters. As in the previous section, each trajectory P was then expressed as a string of classified quasi-states:

$$Q_P = \{q_1, q_2, \ldots, q_n\}$$ [13]

where species q_i belonged to one of the mutual, native-, S111-, S130, or T140-like clusters.

The Energetics of the Stretching Process

The unfolding pathway for the stretching of αA-crystallin presents a 'reaction coordinate', whose variable is elongation by 144 Å (as the protein was almost completely unfolded at this point) along the linear path separating residues C^{α}-60 and C^{α}-148. A constrained free energy profile (PMF, a potential of mean force) can be generated for this process by means of Jarzynski's equality:[18-20]

$$F_a - F_{a_0} = -\frac{1}{\beta}\log\langle e^{-\beta W(a)}\rangle$$ [14]

F_a and Fa_0 are the Helmholtz free energies for the system at states a and a_0, respectively. $\beta = \frac{1}{k_B T}$, where T is the temperature, and k_B the Boltzmann constant. $W(a)$ is the work performed by an external force on the system for its transformation from a_0 to a. Averaging occurs over an ensemble of $W(a)$ values for replicate essays of this process.[18] Furthermore, equation 11 is both path- and rate-independent[18] which enables equilibrium information to be derived from a non-equilibrium transformation.[18,19] Due to difficulty in working with the exponential average, a cumulant expansion approximation[19] was employed:

$$F_a - F_{a_0} = \langle W(a) \rangle + \frac{\beta}{2}(\langle W(a)^2 \rangle - \langle W(a) \rangle^2)$$ [15]

Ensemble averaging occurred over the twelve trajectories that constituted a particular system.

Results I

An inspection of Fig. 5 reveals that a simple correlation between the force curve obtained for a given trajectory and the type and site of racemization is not directly apparent. This highlights the unique and independent nature of each molecular dynamics trajectory. Clearly, although largely directed by the external stretching force, local atomic details propagate through phase-space along a highly unique and trajectory-dependent path. The stretching process, compounded by the cumulative consideration of all such local interactions, produces the nuances of each force curve. This suggests that the force response includes mechanical trends that occur at both global and finer levels, which cannot be decomposed from one another by visual inspection. Negative forces arise as the SMD algorithm adjusts the force so as to maintain stretching at a constant velocity.

Figure 5

The consideration of structural links acts as a filter which omits the treatment of local details and allows concentration on elements of protein structure and organization. A principal component analysis that includes both force and structural links enables the drawing of relationships between protein structure and force. It should be noted that a given set of hot links originating from an eigenvector g_a portrays a trend in structural links that entails both link making and breaking, as Fig. 6 exemplifies.

Figure 6

Examining the sets of hot links, L_s, for each system type reveals that, relative to the native system, isomerization induces a unique set of system-specific links. The beta sheets of αA-crystallin have been highlighted in Figure 7 for reference.

Figure 7

Figures 8 to 11 show each system type and its set of hot links.

Figure 8

Figure 9

Figure 10

Figure 11

The native set, L_N, of hot links (Fig. 8) reveals that the application of high force is correlated with a series of structural links; stress propagates through the structure, and in particular, targets certain structural elements. This dismisses the notion that high force results in the concentration of stress in one particular structural element. Stress propagates through the structure, and in particular, targets certain structural elements. It is also clear that two hot links mark the native trajectory: the first entails the hot link at the junction of β-sheets 2 and 3, and an entering loop/strand of the polypeptide chain; the second, the lower interstice of strand 4 and the exiting polypeptide chain (Fig. 12, shown in red). The short lifetime of the former indicates that this link is broken swiftly in a high-force operation; this effectively frees strands 4 and 5, which pivot on the latter hot-link for the remainder of the trajectory, breaking and re-forming to produce its longer lifetime.

Figure 12

Following the S111 set, L_{S111}` (Fig. 9), one observes that isomerization of residue S111 causes a clear change in the hot link patterning and link lifetime of the native protein. Most

17

affected is β-sheet 5, which has effectively lost all its hot-links. L- to D- conversion therefore produces local effects which weaken the sheet and mechanically uncouple its structure from the applied force (as hot links reflect correlation/sensitivity to high force). Hot links to β-sheet 4 are also missing, suggesting that the region's capacity for continuous sheet formation is hindered. Moreover, the two notable hot links of the native system (Fig. 12, shown in red) no longer dominate the S111 variant; this is particularly true for the lower hot link at the base of sheet 4. Such an observation is in accord with the destabilization of the continuous β-sheet formed by strands 4 and 5, which causes pivot point variability (the region *circa* residues 93 and 120 that mitigates the separation between the upper beta (strands 4-5) and the lower (strands 1-3) sheets), and hence a decline in hot link frequency (Fig.9, blue coloring of these hotlinks). The lifetimes of the hot links in strands 1 and 3 also diminish; this mutant is readily 'unzipped' with the application of force. This observation indicates that isomerization not only produces local effects but also changes the mechanical features of distant structural elements. That the species can be readily 'unzipped' is coherent with its constrained free energy profile (see later text).

The L_{S130} hot link set (Fig. 10) reflects both local and long-range effects of the L- to D-isomerization of Ser residue 130. Locally, the hot link profile of sheet 3 is heavily modified: the life-times and frequencies of links are altered; mutation causes the hydrogen-bonding of this sheet to re-organize, favoring non-native inter-strand interactions. As in the S111 variant, the two notable hot links of the native system have been lost; apparently, interactions within sheet 3 favor internal association over the propagation of the β-sheet across to sheet 2. In compensation, the link density increases in sheet 1. Interestingly, the interstice of sheets 4 and 5 is fortified by long-lived dominant hot-links, which reflect a structural stiffening of this region.

The L_{T140} hot link set (Fig. 11) portrays the most distinct set of links. Mutation in sheet 3 causes significant stiffening of this structural element: native and non-native hot link frequencies and notably increased lifetimes occur along the length of the sheet, further revealing a preference for non-native interactions (as native links give way to non-standard interactions). Unlike the S130 variant, this favors both inter-strand interaction and sheet continuation. As a result, the hot link density increases in sheet 1 (and through transmission, sheet 2). The two key hot links (Fig. 12) of the native system are likewise missing. Link formation in the terminus of the protein shows a tendency towards α-helix formation in preference over interaction with the β-sheet sandwich of the core. The T140 isomerization produces the most pronounced degree of structural element stiffening, both locally in the pertinent β-sheet, and further in the structure via sheets 1 and 2, and the terminal tail.

Across all trajectories, characteristic hot-link sets correlate structural regions to moments of high force, and thus pinpoint mechanically decisive locations. The existence of hot-links reflecting non-native (diagonal) hydrogen-bonding in beta-sheets supports a shearing mechanism for protein unfolding upon the application of force—a result coherent with experimental and *in silico* stretching studies of scaffoldins.[21] Beta-sheet hydrogen-bonding interactions, representing 'mechanical-clamps', are shear-ruptured.[21] Indeed, such mechanical clamps have been identified as recurring key elements in mechanical proteins,[21] with α-crystallin behaving in the same fashion. Thus, the extent and efficiency of hydrogen-bonding within a beta-sheet following an L- to D- isomerization, will determine the strength of its mechanical clamps[21] and so play a critical role in defining the structural stability of the protein.

Results II

Each trajectory was translated into a series of transient unfolding species or 'snag-points', found as minima in dRMSD over the stretch timeline. In general, 7-12 such points were identified per run. Although each unfolding trajectory, by virtue of its molecular dynamics evolution from an independent (although analogous) ensemble, traces a unique path in conformational space, of critical importance is the similarity in snag-point structure as a function of system type. From a mechanical perspective, several questions need to be addressed regarding the nature of snag-points over a stretching simulation of αA-crystallin. Is the unfolding pathway of αA-crystallin marked by snag-points that are of a relatively similar structure and/or similar link pattern? Does L- to D-isomerization change the probability with which certain such structures occur, and furthermore, does it result in the appearance of non-native snag-points or link sets? In the following, we attempt to answer these questions.

Snag-point analysis method I:

Fig. 13 portrays each trajectory simulated as a string of snag-points, which are further classified as belonging to a system-specific cluster (one of either native-, S111-, S130-, or T140-, a mutual-type cluster inherent to αA-crystallin regardless of the system-type, or the null-cluster, which reflects an independent structure not belonging to any of the above mentioned groups).

Figure 13

What is immediately apparent is that all trajectories commence from mutual-class structure(s). These species reflect fundamentally plausible structures that αA-crystallin assumes, regardless of its isomerization state, which suggests that unfolding occurs against a network of cooperative structural interactions. As this network becomes depleted, trajectories diverge and,

for the most part, take on unique snag-point structures. As per the native system, once past this initial 'mutual' stage in the unfolding pathway, 40% of snag-points take on unique structures (Figure 14), which underlines that the unfolding pathway is highly sensitive to the on-the-fly MD evolution of the system.

Figure 14

The remaining 60% of structures fall into the native or mutual classes, with a predominating fraction of mutual-type members. Point isomerization at any of the investigated residues increases the probability for unique snag-point structures to 50%; a coherent unfolding pathway of the protein becomes less probable. Most importantly, the allocation of the remaining 50% of snag-points into the mutual and other classes depends on the mutation. The S111 isomerization alters the unfolding pathway significantly; the mutual fraction is suppressed in favor of characteristic S111-class species. Mutation in the beta-sheet of αA-crystallin (S130, T140) maintains or augments the mutual fraction, and causes a decline of the native fraction (Figure 14).

Snag-point analysis method II

Figure 15 reproduces the snag-points for each trajectory, and colors them according to cluster, as obtained through the clustering of structural link patterns. (A cluster here is defined as a family of structures with similar quasi-state vectors, and hence structural link patterns).

Figure 15

It is clear that clustering by this method is more effective; trajectories are by majority homogenous in regard to cluster type, and the clustering method by design allows for the elimination of 'unique' structures. The former indicates that structural link patterns are superior

21

to RMSD similarity as a structural comparison tool. Indeed, as an example, the sequential detachment of intact structural elements would produce a large change in RMSD for two hypothetical structures, although structurally, these could in principal differ by the absence of only a few connecting links between protein units. Accordingly, the RMSD tracks spatial similarities, while the use of structural links captures organizational patterns. Fig. 15 indicates that trajectory species do not necessarily start with a member of the mutual class; this is especially evident for the S111 mutation; and so structural link patterns are often perturbed in the protein even before it is stretched, despite a normal 3D appearance by RMSD inspection. As Figure 16 reveals, native snag-points follow two patterning behaviors during the stretching process.

Figure 16

50% of snag-point species follow characteristic native-type link patterns, and 40% assume structures belonging to the mutual cluster, which expresses patterning trends inherent to αA-crystallin, regardless of its isomerization state. The S111 isomerization favors majority formation of S111-class structures, at the expense of the native fraction (Fig. 16).

That both this and the previous clustering method are coherent with this observation indicates that the S111 mutation causes changes to the structure of αA-crystallin from both a spatial (RMSD) and link perspective. The S130 mutation causes a significant change in structural link pattern—over 60% of snag-points are of the S130 cluster type (Fig. 16). This suggests that this mutation causes significant rearrangement in the β-sheet, and favors retention of the new link pattern during the stretching process. This link arrangement is not readily visible with only an RMSD cluster analysis, as the latter identifies only a small fraction of the snag-points as being

22

of the S130 type. Applying the adjacently located T140 mutation causes an unstable link patterning, and results in the splitting/scrambling of the snag points among the various classes (and even the S130 pattern) with a moderate preference for the T140 class (Fig.16).

Snag-Point Summary

From a link perspective, the unfolding pathway passes through a series of consecutive snag-points that fall into two main categories: i) those that take on link patterns reflective of the inherent nature of αA-crystallin, and ii) patterns that are isomerization-dependent. For the native system, snag-points fall into either group with a preference for native-states. D-isomerization of a point residue results in the suppression of native-like snag-points and involves the introduction of a new D-isomer-class of structures. The location of the mutation determines the degree to which the probability of native snag-points is diminished, and the degree to which the new isomer-dependent class is favored. The S111 and S130 mutations produce the dominant link classes in the unfolding pathway, while the D-T140 substitution provides a transient/unstable class of link patterns. From a spatial and hence RMSD perspective, the native and S111 system types produce snag-points that spatially resemble each other. Despite their appearance, all system types focus on the preservation of link-patterning. The structural-link coherence of trajectories on a per-system basis confirms that αA-crystallin unfolds by the sequential detachment of organized protein-units.

Results III:

The Jarzynski equality enables a potential of mean force to be drawn for the stretching process (Fig. 17).

Figure 17

23

As expected from the earlier analysis of structural links as indicators of fragment stiffness, the T140 mutation; which produced the greatest concentration of hot links in the beta-sheet region; renders the highest energy PMF. The neighboring S130 mutation is more moderate in nature, and produces a lower energy trace. As predicted earlier, the S111 mutation allows the protein to be "unzipped", and embodies an unfolding pathway at a lower energy than the native species.

Discussion

The D-isomerization of a select residue in the structure of αA-crystallin causes changes to both the protein's force-induced unfolding pathway, and to the mechanical properties of its structural elements. The effects rendered depend on the location of the D-substitution, and manifest themselves as divergence from the probable native unfolding pathway, and in changes to the stiffness of particular structural elements.

Structural links allow protein structure and organization to be studied in a fashion that filters out the (variable) local atomic details of MD. Moreover, a principal component analysis of time-dependent structural link patterns that includes the on-the-fly stretching force enables the mechanical properties of structural elements to be assessed.

The native protein responds to force by propagating the resultant stress through the β-sheets of αA-crystallin; these are the stiffest structural elements in the protein. The unfolding pathway occurs with passage through intermediate 'snag-points.' Although the progression of each trajectory is unique, the structural organization of these intermediates falls with significant probability into a mutually coherent native-like structural class. Structural similarity may be

assessed by comparing either the RMSD or link patterns; the latter offers a simple and unambiguous method for discerning the degree of similarity between snag-points across different trajectories.

The D-S111 mutation causes the core beta-sheet structure of αA-crystallin to be weakened, which allows the protein to be more readily unfolded than its native precursor; indeed, a PMF for the stretching process confirms a lower energy pathway. Unfolding snag-points diverge away from the native-like family in preference for structures that are characteristic of D-S111 substitution, from both an RMSD and (especially) link perspective. This means that the probability of a native-like snag-point drops significantly, while the probability of a featured D-S111 structure increases. On the other hand, direct substitution in the beta-sheet itself causes two effects: the formation and stabilization of a new family of structures bearing a distinct non-native link patterning within the beta-sheet, indicative of non-native H-bonding, or class scrambling. In both cases, the effect is best represented by link pattern comparison (as the RMSD shows a tendency for structural broadening/divergence, despite a potentially high degree of conserved links). The D-S130 substitution favors the formation of species that reflect the D-S130 mutation; the core beta-sheet becomes stiffer than in the native protein. The D-T140 substitution produces a non-stable D-T140 family of structures, whose members favor run-scrambling: D-T140 trajectories adopt unique link-patterns or express a snag-point from any of the other classes (native, D-S111, D-S130). Indeed, the D-T140 PMF offers the highest-energy profile for the stretching process, while visualization of hotlinks reflects the formation of local stiff zones across the beta-sheet and tail of αA-crystallin.

The behavior of αA-crystallin as both a chaperone and structural protein depends on both its ability to form and recover from protein-protein interactions. Clearly, the present study reveals that the isomerization of a single amino-acid can have drastic effects on its mechanical and unfolding properties, substantiating a direct correlation between the experimentally noted parallel increases in lens stiffness and D-amino acid content[22,23]. Moreover, previous concern that diminished chaperone capacity can render the interception of lens unfolding intermediates by αA-crystallin inefficient[22,23] is coherent with and further compounded by the inferior ability of αA-crystallin to recover from its own unfolded states, as is demonstrated by a drop in the probability of native snag-points under all of the studied D-systems.

Conclusion

Although the effects of a given isomerization are site-specific, the present study concludes that L- to D-isomerization of a single residue in the sequence of αA-crystallin alters both its mechanical stiffness and its force-induced unfolding pathway, regardless of its position along the sequence of the protein. An increase in mechanical stiffness assumes lens hardening, while the reverse implies lower stability, and hence susceptibility towards agglomeration. Both scenarios are coherent with the phenomena of lens hardening as a function of age and, further, as a function of concomitant D-amino acid content. The speed with which compromised accommodation and insoluble plaques appear likely depends on the potential of the substituted D-bearing sites for exhibiting non-native properties. In summary, L- to D-amino acid conversion

can be viewed as a detrimental part of the ageing process, being critically involved in progressive

cataract (agglomerate) formation and lens hardening.

References

1. Glasser A, C.W. Campbell M. Biometric, optical and physical changes in the isolated human crystalline lens with age in relation to presbyopia. Vision Research 1999;39(11):1991-2015.
2. Truscott RJW, Zhu X. Presbyopia and cataract: A question of heat and time. Progress in Retinal and Eye Research 2010;29:487-499.
3. Hooi MYS, Truscott RJW. Racemisation and human cataract. AGE 2011;33:131-141.
4. Sharma K, Santhoshkumar P. Lens aging: effects of crystallins. Biochim Biophys Acta 2009;1790(10):1095-1108.
5. Goodsell DS. Molecule of the Month: Crystallins. RCSB PDB Molecule of the Month, www.pdb.org: RCSB Protein Data Bank; 2010.
6. Hooi MYS, Raftery MJ, Truscott RJW. Age-dependent racemization of serine residues in a human chaperone protein. Protein Science 2013;22(1):93-100.
7. Galera-Prat A, Gomez-Sicilia A, Oberhauser AF, Cieplak M, Carrion-Vazquez M. Understanding biology by stretching proteins: recent progress. Current Opinion in Structural Biology 2010;20(1):63-69.
8. Sikora M, Cieplak M. Mechanical stability of multidomain proteins and novel mechanical clamps. Proteins 2011;79:1786–1799.
9. Lu H, Schulten K. Steered Molecular Dynamics Simulations of Force-Induced Protein Domain Unfolding. Proteins 1999;35:453-463.
10. Best RB, Li B, Steward A, Daggett V, Clarke J. Can Non-Mechanical Proteins Withstand Force? Stretching Barnase by Atomic Force Microscopy and Molecular Dynamics Simulation. Biophys J 2001;81:2344-2356.
11. Wikipedia. Molecular Dynamics.
12. Case DA, Darden TA, III TEC, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts B, Hayik S, Roitberg A, Seabra G, Swails J, Götz AW, Kolossváry I, K.F.Wong, Paesani F, Vanicek J, R.M.Wolf, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Salomon-Ferrer R, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA. Amber 12. San Francisco: University of California; 2012.
13. Case DA, Cheatham I, T. E. , Darden T, Gohlke H, R. Luo, Merz J, K. M. , Onufriev A, Simmerling C, Wang B, Woods R. The Amber biomolecular simulation programs. Journal of Computational Chemistry 2005;26:1668-1688.
14. Levitt M, S. L. Refinement of Protein Conformations using a Macromolecular Energy Minimization Procedure. Journal of Molecular Biology 1969;46:269-279.
15. R. Fletcher, Reeves CM. Function minimization by conjugate gradients. The Computer Journal 1964;7(2):149-154.
16. Mey Khalili, Adam Liwo, Anna Jagielska, Scheraga HA. Molecular Dynamics with the United-Residue Model of Polypeptide Chains. II. Langevin and Berendsen-Bath Dynamics and Tests on Model α-Helical Systems. J Phys Chem B 2005;109:13798-13810.
17. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. Numerical Recipes. 3rd ed: Cambridge University Press, Cambridge, UK; 2007.
18. Jarzynski C. Nonequilibrium Equality for Free Energy Differences. Phys Rev Lett 1997;78(14):2690-2693.

19. Park S, Schulten K. Calculating potentials of mean force from steered molecular dynamics simulations. J Chem Phys 2004;120:5946-5961.
20. Ozer G, Quirk S, Hernandez R. Thermodynamics of Decaalanine Stretching in Water Obtained by Adaptive Steered Molecular Dynamics Simulations. 2012;8:4837−4844.
21. Valbuena A, Oroz J, Hervás R, Vera AM, Rodríguez D, Menéndez M, Sulkowska JI, Cieplak M, Carrión-Vázquez M. On the remarkable mechanostability of scaffoldins and the mechanical clamp motif. Proceedings of the National Academy of Sciences 2009;106(33):13791-13796.
22. Hooi MY, Raftery MJ, Truscott RJ. Age-dependent racemization of serine residues in a human chaperone protein. Protein Sci 2013;22(1):93-100.
23. Heys KR, Cram SL, Truscott RJW. Massive increase in the stiffness of the human lens nucleus with age: the basis for presbyopia? Molecular Vision 2004;10:956-963.

Figure Legends

Figure 1: Crystal structure of αA-crystallin (*Bos taurus*, PDB: 3L1E). The residues (Gly 60 and Ser 148) used as reference points in the SMD stretching process are shown in blue. The location of candidate residues for D-isomerization are identified in yellow (Ser 111 and 130) and purple (Thr 140).

Figure 2: A. An illustration of the structural links, identified by orange bars, present in the native structure of αA-crystallin. **B.** Expanded view of the structural links as they appear in a fragment of a β-sheet. **C.** Expanded view of the structural links as they occur within a fragment containing a short stretch of α-helix.

Figure 3: An illustration of the construction of the covariance matrix, C_P, for trajectory P. Matrix columns and rows consist of the force ($\vec{f}$) and n link ($\vec{\iota}_a$) vectors of P ($n \approx 250\text{-}300$). Individual terms consist of the covariance between corresponding entries.

Figure 4: Graphical representation of the change in root-mean-square deviation (dRMSD) for consecutively aligned frames of a (native) trajectory over the course of a simulation. Numbered points indicate 'snag-points' – that is, sequences of minimal structural change, found as minima in the function after smoothing.

Figure 5: Force (pN) as a function of time over the course of the SMD simulation. The stretching force is plotted in a different color for each individual trajectory. A. Native system

trajectories, B. D-S111 system trajectories, C. D-S130 system trajectories, D. D-T140 system trajectories. (1 frame = 1 600 fs)

Figure 6: A sample eigenvector g_a and its associated hot links (left) for the native structure. The life-time plot (right) of all structural links reveals that the trend described by the eigenvector entails both link-making and -breaking; a colored point indicates that the link is present at that time-frame in the simulation. Different colours are used to distinguish adjacent links.

Figure 7: Schematic division of the β-sheets of αA-crystallin for reference purposes. Yellow spheres: serine residues; purple spheres: threonine residues; blue spheres: pull points. Large spheres: serine 111 (red), serine 130 (yellow), threonine 140 (purple). [Front: strands 1-3; Back: strands 4-5]

Figure 8: The characteristic native set, L_N, of hot links. The hotlink frequency (set weight, w_i) is indicated by color (red being most hot/frequent). The lifetime of a hotlink is depicted by the length of the central barrel (increased making/breaking frequency has a longer barrel). Left-side and the right-side views of the molecule are provided.

Figure 9: The characteristic D-S111 set, L_{S111}, of hot links. The hotlink frequency (set weight, w_i) is indicated by color (red being most hot/frequent). The lifetime of a hotlink is depicted by the length of the central barrel (increased making/breaking frequency has a longer barrel). Left-side and the right-side views of the molecule are provided.

Figure 10: The characteristic D-S130 set, L_{S130}, of hot links. The hotlink frequency (set weight, w_i) is indicated by color (red being most hot/frequent). The lifetime of a hotlink is depicted by the length of the central barrel (increased making/breaking frequency has a longer barrel). Left-side and the right-side views of the molecule are provided.

Figure 11: The characteristic D-T140 set, L_{T140}, of hot links. The hotlink frequency (set weight, w_i) is indicated by color (red being most hot/frequent). The lifetime of a hotlink is depicted by the length of the central barrel (increased making/breaking frequency has a longer barrel). Left-side and the right-side views of the molecule are provided.

Figure 12: Two major hotlinks from the native set, L_N, of hot links. Residues (spheres) for reference: orange (S111), yellow (S130), purple (T140). Blue spheres: pull points.

Figure 13: RMSD classification of snag-point structures for each trajectory into the Native, D-S111, D-S130, D-T140, and Mutual classes. A unique (non-classed) species is indicated by a black arrow, while a classified structure by a colored square. The 12 rows refer to the 12 trajectories. Snag-points (time) progress(es) along each row, from left to right. The number of snag-points is run-specific (hence the recorded length of structures per run).

Figure 14: Classification of snag-point structures based on RMSD for each trajectory into the Native, D-S111, D-S130, D-T140, and Mutual classes.

Figure 15: Fraction of snag-point structures falling into each of the Native, D-S111, D-S130, D-T140, and Mutual classes. Clustering occurs by structural link similarity. A classified structure is shown by a colored square. The 12 rows refer to the 12 trajectories. Snag-points (time) progress(es) along each row, from left to right. The number of snag-points is run-specific (hence the recorded length of structures per run).

Figure 16: Fraction of snag-point structures falling into each of the Native, D-S111, D-S130, D-T140, and Mutual classes. Clustering by link-pattern similarity.

Figure 17: PMF profiles (kcal/mol) for the stretching reaction coordinate (Å) of αA-crystallin. Green: native, Orange: D-S111, Yellow: D-S130, and Purple: D-T140 systems, respectively.

Figures

Figure 1:

Figure 2

Figure 3

Covariance Matrix C_P

$\vec{i}_n = n^{th}$ link vector

$\vec{f} = $ trajectory force vector
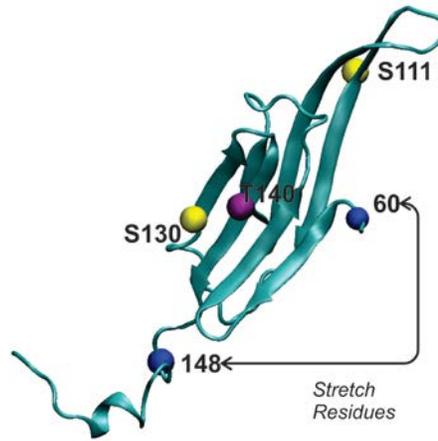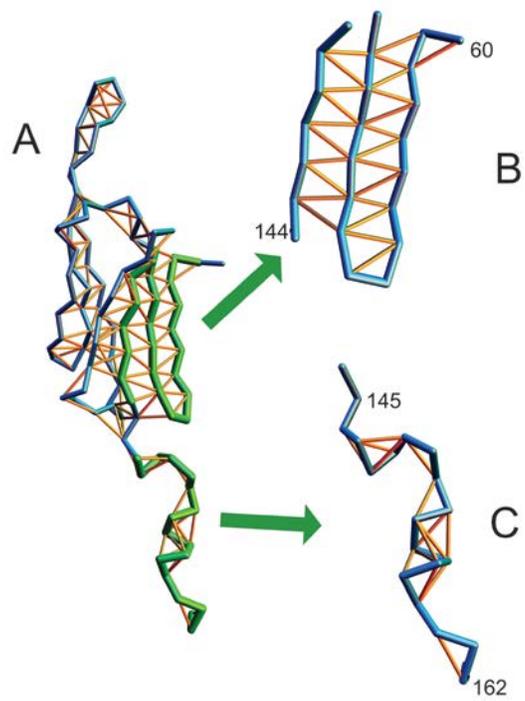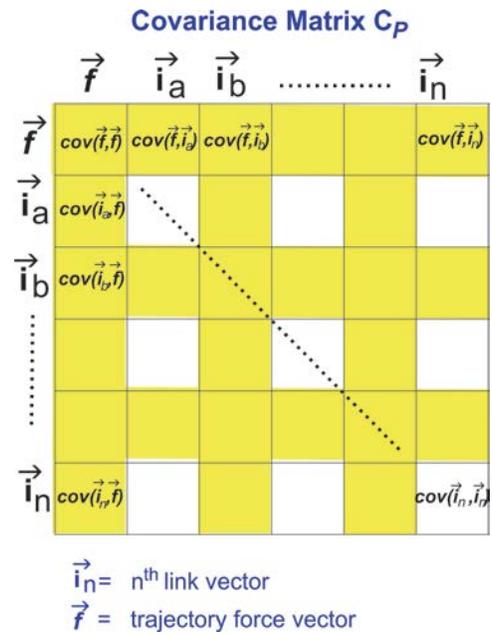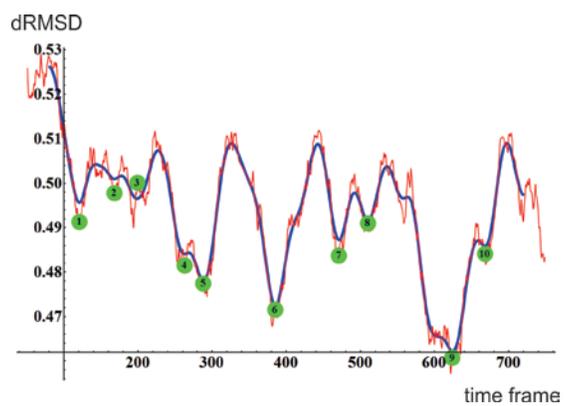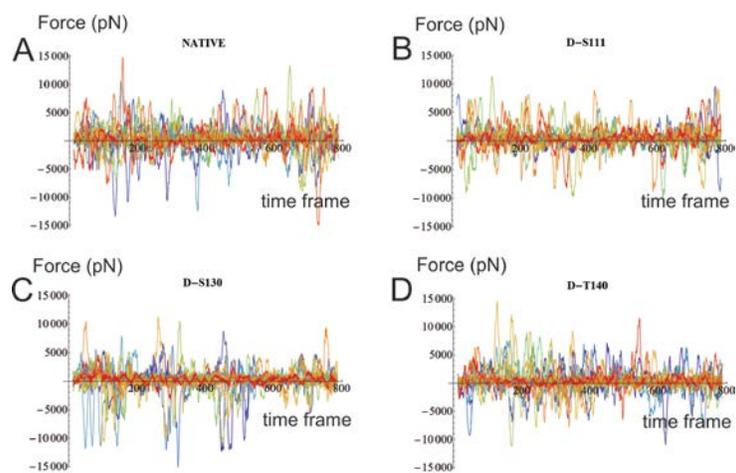
Figure 4

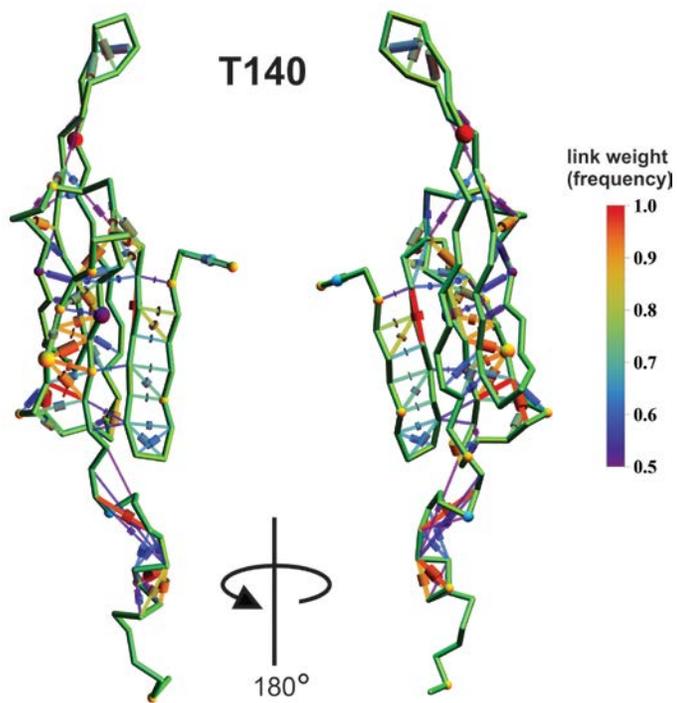**Figure 5**

Figure 6
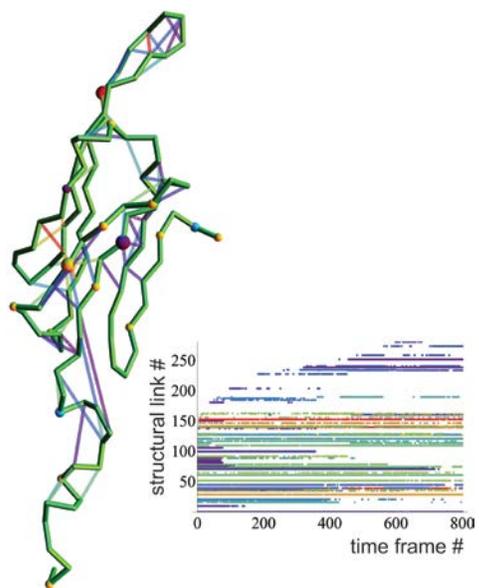
Figure 7
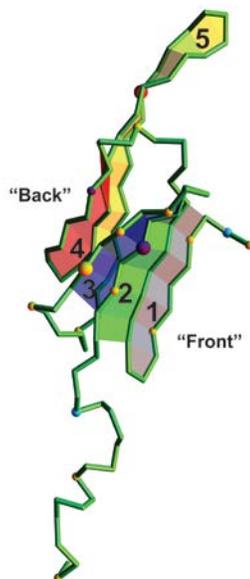
Figure 8
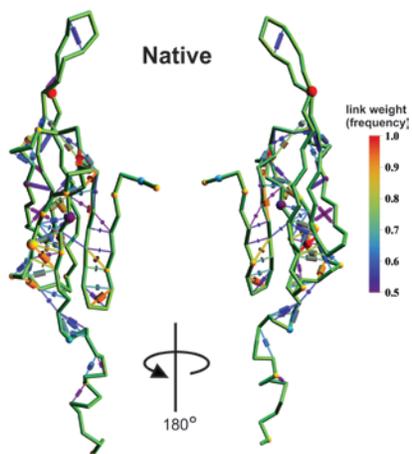
Figure 9
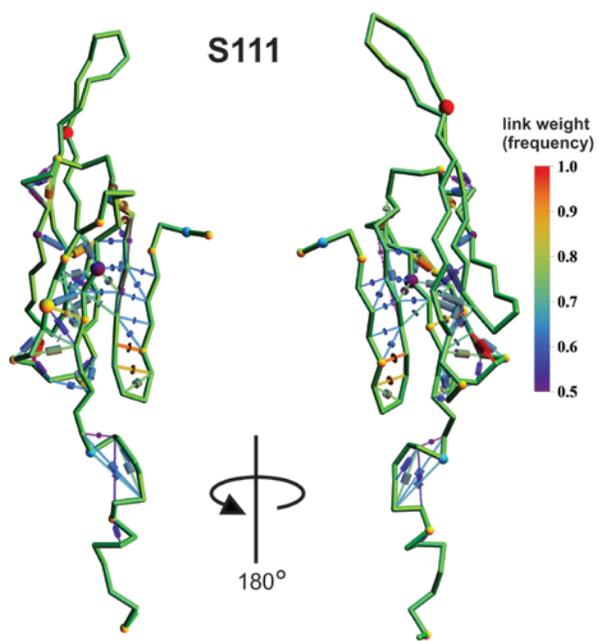
Figure 10
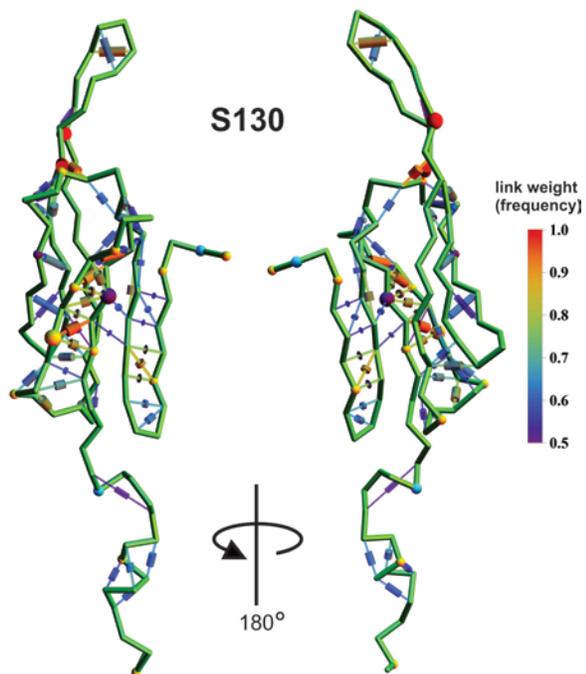
Figure 11

Figure 12
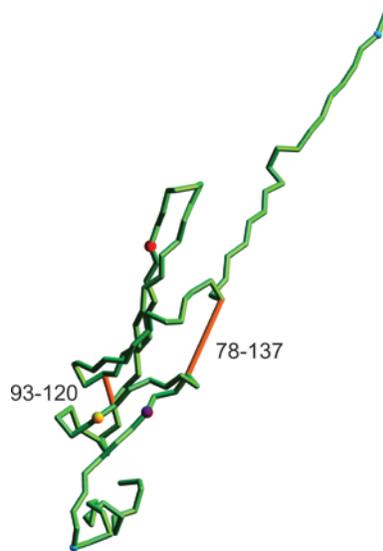
Figure 13
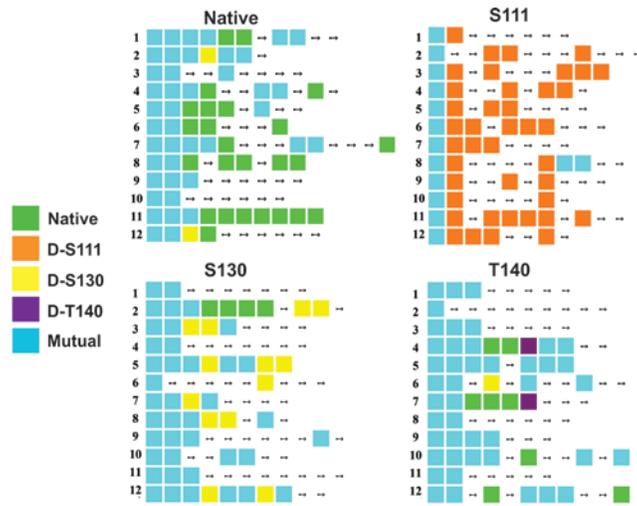
Figure 14
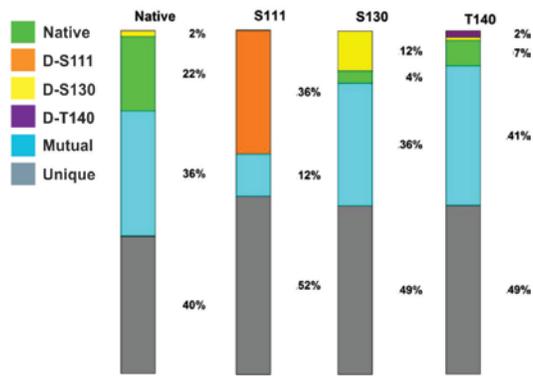
Figure 15

Figure 16

Figure 17