Tropoelastin is a Flexible Molecule that Retains its Canonical Shape

Anna Tarakanova, Giselle C. Yeo, Clair Baldock, Anthony S. Weiss, and Markus J. Buehler*

1. Introduction

Tropoelastin is the molecular–precursor to elastin and is a dominant component of elastic fibers, key structural elements in elastic tissues including blood vessels, skin, and lungs.\(^1\)\(^-\)\(^2\) The mature form of the secreted human tropoelastin monomer is 60 kDa\(^3\) and is composed of an alternating pattern of repetitive, hydrophobic glycine-, valine-, and proline-rich domains spread out between lysine-containing cross-linking domains.\(^4\) Traditionally described as a largely disordered molecule, recent studies suggest that tropoelastin is a flexible molecule with a distinct nanostructure\(^5\)\(^-\)\(^7\) that determines its structural and cell-adhesive properties.

Tropoelastin’s distinct nanostructure determines its propensity to assemble into higher-order structures (Figure 1) through elastogenesis.\(^8\) Elastogenesis begins with self-assembly of the \(\approx15\) nm monomers, a process called coacervation.\(^9\) Self-assembly, naturally occurring at physiological pH and salt conditions at 37 °C, results in spherical structures from 200 nm to 6 \(\mu\)m in diameter.\(^9\)\(^-\)\(^11\) Spherule formation through coacervation begins at the cell surface by surface-tethered tropoelastin molecules, until released for integration into the growing elastic fiber. Tropoelastin aggregates adhere to the cell surface through GAG- and integrin–mediated interactions at specific interaction sites.\(^12\)\(^-\)\(^18\) Spherules are deposited on a microfibrillar scaffold where they are stabilized and then cross-linked to form the growing elastic fiber. The mature form of the secreted human tropoelastin monomer is 60 kDa and is composed of an alternating pattern of repetitive, hydrophobic glycine-, valine-, and proline-rich domains spread out between lysine-containing cross-linking domains. Traditionally described as a largely disordered molecule, recent studies suggest that tropoelastin is a flexible molecule with a distinct nanostructure that determines its structural and cell-adhesive properties. Tropoelastin’s distinct nanostructure determines its propensity to assemble into higher-order structures (Figure 1) through elastogenesis. Elastogenesis begins with self-assembly of the \(\approx15\) nm monomers, a process called coacervation. Self-assembly, naturally occurring at physiological pH and salt conditions at 37 °C, results in spherical structures from 200 nm to 6 \(\mu\)m in diameter. Spherule formation through coacervation begins at the cell surface by surface-tethered tropoelastin molecules, until released for integration into the growing elastic fiber. Tropoelastin aggregates adhere to the cell surface through GAG- and integrin–mediated interactions at specific interaction sites. Spherules are deposited on a microfibrillar scaffold where they are stabilized and then cross-linked to form the growing elastic fiber.
Our simulations capture intrinsic conformational characteristics of tropoelastin, suggesting a bias toward tropoelastin's canonical nanostructure. We describe the conformational heterogeneity of tropoelastin within the confines of its broader nanostructure that determines the molecule's wide range of functions. Our results are discussed in the context of previous studies to rationalize functional implications of the conformational ensemble of the molecule.

2. Results and Discussion

We analyze the ensemble of 1000 tropoelastin structures at 37 °C derived from replica exchange molecular dynamics simulation as described previously. The structures considered correspond to the last 2 ns of REMD simulation in explicit water. To characterize the structural variability of tropoelastin (also referred to the motions and dynamics of the molecule throughout the text, not to be confused with time-dependent dynamics not accessible to REMD), we perform principal component analysis on this set of structures. The variance for the first 20 principal component modes is shown in Figure 2b. We note that the first two principal component modes have variances of 23% and 19%, respectively, accounting together for 42% of the variation in the structural ensemble. The variance of higher modes drops considerably after the second mode, indicating a lower contribution to structural variation. As we will show in the forthcoming discussion, higher modes do not affect the relative flexibility of tropoelastin domains.

While the variance of the principal components allows us to identify dominant directional shifts in the distribution of all considered structures, we perform an independent clustering analysis to extract and analyze representative structures within the ensemble. We use k-means clustering, a method that employs a fixed cluster radius based on Cartesian coordinate root mean square deviation (RMSD) between structures, where RMSD is set to 5 Å. There are a total of 39 clusters identified through this approach. A histogram of clusters, ranked from most populated (Cluster 1) to least populated (Cluster 39) is shown in Figure 2b. The lowest energy structure is extracted from each of the 39 clusters. We next consider the distribution...
of the structures within the subspace spanned by the first two principal components (Figure 3a) as a representative spread of the observed structures of tropoelastin at 37 °C. All 1000 structures are projected onto PC1-PC2 space. Thirteen representative structures from k-means clustering analysis are labeled on the principal component map (Figure 3a) according to their cluster affiliation. The specific location of structures 1–10, 17, 20, and 24 is shown within the subspace. Representative structures affiliated with the most populous clusters (1–10) exhibit broad flexibility as observed in structures 2, 17, 20, and 24 are chosen as a means to consider less frequently sampled, yet still accessible states of tropoelastin. The representative structure from the most populous cluster (structure 1) is considered as a reference. Six structures (2, 5, 8, 17, 20, 24) are chosen in more detail to analyze the range of accessible conformations available to the molecule (Figure 3b–g). These six structures are chosen based on their location on the principal component map, as representative of greatest structural variability found in tropoelastin at 37 °C. They are superimposed on structure 1 by considering the minimal RMSD based on all atoms.

We find that despite an intrinsic flexibility displayed by the molecule, the range of motion accessible by tropoelastin preserves its canonical structure\cite{5,6,54} with an extended molecular body and two protruding legs (Figure 3b–g). The first four domains of the molecule (D2–5) are characteristically positioned at the head of the structure, in the N-terminal region. The location is consistent with earlier studies based on small angle X-ray and neutron scattering experiments\cite{9} that estimate the locations of these domains by comparing full-length and truncated tropoelastin constructs. This region exhibits broad flexibility as observed by its twisted position in structures 2, 5, and 24 (Figure 3b,c,g). A more localized twist of domains 2 and 3 (D2 and D3) is seen in structure 8 (Figure 3d). The repositioning of these domains is consistent with the natural tendency toward a twisting motion in the upper region of the molecule described in earlier elastic network models and normal mode analysis of tropoelastin\cite{6,54}.

The preserved position of the first four domains is supported by adjacent domain 6, which contains aspartate in position 72, the only negatively charged residue in the N-terminal region of the molecule, that may function to stabilize the upper region of the molecule.\cite{54,55} We have shown previously that domain 6 partially forms a stable α-helix that is locked in place by a salt bridge between aspartate 72 and three neighboring lysine residues.\cite{54} In fact, the location of domain 6 (D6) shows regularity across the tropoelastin ensemble. We note that in structure 17 (Figure 3e), domains 6 and 7 (D6 and D7) are shifted, yet the location of D2–5 remains consistent with structure 1. A minor shift in domain 6 is seen also in structure 20 (Figure 3f), again without significant change to the location of the first domains of the molecule.

The central region of the molecule through domain 19, characterized by a conserved arrangement of alternating hydrophobic and hydrophilic domains, is contained within the elongated torso of the molecule. The largest structural oscillation in this region is apparent in structures 5 and 8 (Figure 3c,d), where domains 10–19 and 15–19, respectively, are twisted leftward. This shift is again consistent with the predicted twist in the torso of the molecule.\cite{6,54} More localized shifts are also characteristic of this region. Domain 19 displays flexibility away from structure 1 (in structures 2, 17, 24, Figure 3b,e,g), as do domains 15 and 17 (in structure 20, Figure 3f). The flexibility of the large central region of the molecule suggests a twofold functionality. First, the central spring-like coil region has been previously proposed to account for tropoelastin’s superb elasticity, allowing the molecule to extend to eight times its resting length.\cite{5} A wide structural flexibility available to these domains in the relaxed state is consistent with an entropic model of tropoelastin’s elasticity, where extension restricts the number of conformations available to the molecule which then drives its return to the resting state.\cite{56–58} Second, the variability in domain positioning may promote cellular interaction. Earlier work suggests that the central region of the molecule interacts with cell surface receptors including glycosaminoglycans (GAGs) and integrins, promoting cell binding and cell spreading.\cite{17,18} In particular, domains 17 and 18 contain a glycosaminoglycan-binding

Figure 2. a) Percentage variance for the first 20 modes from principal component analysis. Variance corresponds to the fractional contribution of each mode to structural variation within the ensemble. b) Population of structures in 39 clusters through k-means clustering, with an RMSD of 5 Å, where clusters are ranked from most (1) to least (39) populous.
Figure 3. a) Projection of each of the 1000 structures of tropoelastin at 37 °C onto PC1 and PC2 (white circles). Structures 1–10, 17, 20, 24 represent the lowest energy structures from corresponding clusters (colored circles with red numerical labels). Structure 1 is the reference structure, and structures 2, 5, 8, 17, 20, and 24 are considered for characterizing the spread of structural variation on PC1-PC2 space, indicated by arrows. Structure 1, in blue, overlaid with b) structure 2, c) 5, d) 8, e) 17, f) 20, and g) 24. Specific domain shifts are indicated by arrows in (b–g).
site responsible for cell adhesion. This sequence works synergistically with integrins, where GAG-binding allows the engagement of the $\alpha_v$ integrin family to induce cell spreading.[18] Domains that contribute to integrin binding reside in regions adjacent to the GAG-binding sequence spanning domains 17 and 18.[18] We propose that the flexibility of the central region may therefore facilitate cell–receptor interactions by exposing relevant domains.

Continuing down the length of the molecule, a variability in the location of domains 21 and/or 23 is observed in five of the six structures considered (Figure 3b–e,g). Due to alternative splicing of tropoelastin mRNA, domain 22 is excluded from human tropoelastin so that two hydrophilic domains 21 and 23 are adjoined, forming a hinge region,[59] previously identified computationally and empirically through NMR and small-angle X-ray scattering.[60–62] Elastic network models of domains 21–23 suggest a flexible peptide capable of existing in closed and open conformations.[6] The high mobility of this region is supported by the findings in the present analysis and suggests that domains 21 and 23 contribute to tropoelastin’s elasticity, as flanking domains to the central elastic region of the molecule. The scissors-like bending of domains 21–23 is consistent with the dynamic models,[6,7] suggesting an additional role in multimeric assembly via domain repositioning and conformational space sampling to facilitate cross-linking.

The C-terminus is highly conserved across mammalian species, terminating with a charged RKKK sequence.[11] It is key for tropoelastin incorporation into elastic fibers through elastogenesis, as well as GAG- and integrin-mediated cell adhesion.[12–16] In tropoelastin’s structural ensemble, the C-terminal domain 36 shows significant flexibility in structures 2, 5, 8, and 24 (Figure 3b–d,g). Adjacent domains, in particular domains 32 and 29 in structures 17 and 20, respectively (Figure 3e,f) can also undergo displacement. The flexibility of the C-terminus has been predicted previously through elastic network models and normal mode analysis.[6,54] This flexibility suggests functionality by widening the conformational space available to this region for improved interactions with cell-surface receptors and elastic fiber-associated proteins.

The displacement profile of tropoelastin domains confirms the observations noted by analyzing specific representative structures. Figure 4a shows the displacement profile as a weighted contribution of movements along the top two PCA modes (black curve) and the top six PCA modes (red curve). The top two PCA modes account for 42% of structural variation, while the top six PCA modes account for 70% of the structural variation. The similar trend for two and six modes supports the dominance of the first two modes, validating our approach to consider the structural projection onto PC1-PC2 space. Largest displacements are found in domains 2–5 (residues 1–51), 10–19.
(residues 133–357), 21–23 (residues 413–445), and domain 36 (residues 685–697) (Figure 4a). Lysine residues, marked as blue dots on Figure 4a, are scattered throughout the tropoelastin structure and involved in inter- and intramolecular cross-links. They are almost uniformly found in highly mobile domains on the molecule. In fact, 83% of the 35 lysine residues found in tropoelastin reside at the high peaks of relative displacement among their nearest neighbor residues. The large displacements of the lysine residues suggest a mechanism for high conformational space sampling toward efficient cross-linking during elastogenesis.

Finally, large domain dynamics of the molecule are considered in Figure 4b. Snapshots (i–iv) in Figure 4b show the motion associated with the linear combination of the top six PCA modes, scaled by their variance. The molecule tends motion associated with the linear combination of the top sixPCA modes, scaled by their variance. The molecule tends to be a dynamic model to drive multimolecular interactions.

3. Conclusions

The distribution of structures accessible to human tropoelastin is characterized via principal component analysis. The distribution preserves the canonical tropoelastin structure, while oscillations in structure give rise to predicted molecular motions driving a twist at the N-terminal region and a scissors-like bending in the more C-terminal regions of the molecule. Enhanced fluctuation of specific domains, and more specifically lysine residues within the molecule, is discussed to explain regional functionality and implications for elastogenesis. The current model supports a paradigm shift for considering tropoelastin’s structure. Despite a high degree of disorder, tropoelastin maintains a uniform nanostructure, and local dynamics are tied to specific regional functions of the molecule.

4. Experimental Section

The REMD-based methods for the development of the structural ensemble of tropoelastin were previously discussed.[7] The ensemble at 37 °C was characterized using the MMTSB toolset[83] and ProDy PCA analysis tools.[64]

Acknowledgements

Support from NIHU01 HS 4976 and ONRN00014-16-1-2333 is acknowledged. This work utilized the Extreme Science and Engineering Discovery Environment (XSEDE) [83] which is supported by National Science Foundation grant number ACI-1053575. XSEDE resources Stampede 2 and Ranch at the Texas Advanced Computing Center and Comet at the San Diego Supercomputing Center through allocations TG-MS5090007 and TG-MCB180008 were used. C.B. acknowledges grant support from the BBSRC (BB/N015398/1 and BB/R008221/1).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

disordered protein, elastic fiber, principal component analysis, structural ensemble, structural protein, tropoelastin

Received: July 3, 2018
Revised: September 3, 2018
Published online: