

MOLECULAR DYNAMICS SIMULATIONS OF THE FULL TRIPLE HELICAL REGION OF COLLAGEN TYPE I PROVIDE AN ATOMIC SCALE VIEW OF THE PROTEIN'S REGIONAL HETEROGENEITY

DALE L BODIAN, RANDALL J RADMER[†], SEAN HOLBERT[‡], TERI E KLEIN

*Department of Genetics, Stanford University, 1501 S. California Avenue
Palo Alto, CA 94304*

Email: teri.klein@stanford.edu

Collagen is a ubiquitous extracellular matrix protein. Its biological functions, including maintenance of the structural integrity of tissues, depend on its multiscale, hierarchical structure. Three elongated, twisted peptide chains of >1000 amino acids each assemble into trimeric proteins characterized by the defining triple helical domain. The trimers associate into fibrils, which pack into fibers. We conducted a 10 ns molecular dynamics simulation of the full-length triple helical domain, which was made computationally feasible by segmenting the protein into overlapping fragments. The calculation included ~1.8 million atoms, including solvent, and took approximately 11 months using the CPUs of over a quarter of a million computers. Specialized analysis protocols and a relational database were developed to process the large amounts of data, which are publicly available. The simulated structures exhibit heterogeneity in the triple helical domain consistent with experimental results but at higher resolution. The structures serve as the foundation for studies of higher order forms of the protein and for modeling the effects of disease-associated mutations.

1. Introduction

Collagen is a ubiquitous protein found in all multicellular organisms. Type I collagen, the most abundant protein in mammals, provides structural and functional integrity to bones, tendons, blood vessels, and other tissues. The function of collagen relies on its multiscale, hierarchical structure. The type I collagen protein is composed of two $\alpha 1(I)$ and one $\alpha 2(I)$ peptide chains, each composed of >1000 amino acids. The three chains, each with a left-handed polyproline II-like twist, associate into a supercoiled right-handed triple helical structure nearly 3000 Å long.¹ The heterotrimers assemble into fibrils with a characteristic packing arrangement that is observable in electron micrographs as a 670 Å repeating pattern of gaps and bands. The fibrils in turn pack into fibers or other suprafibrillar architectures in the extracellular matrix whose configuration varies with the biological tissue in which it is found.^{2,3}

The primary sequence of collagen proteins is characterized by multiple repeats of the Gly-X-Y triplet, where X and Y can be any amino acid but are often proline and hydroxyproline. The type I collagen peptide chains $\alpha 1(I)$ and $\alpha 2(I)$ each contain 338 uninterrupted copies of this repeat which form the triple helical domain in the heterotrimer. Disruption of the structure of the triple helical domain by mutation, particularly of any of the invariant glycine residues, is associated with disease, most commonly osteogenesis imperfecta, a set of disorders characterized by brittle bones.

Experimental studies⁴⁻⁷ and computational models^{8,9} have revealed that, despite the repetitive sequence, the triple helical domain is not homogeneous and has structural and biological

[†] Current address: Department of Bioengineering, Stanford University, Stanford, CA 94305.

[‡] Current address: Department of Computer Science, Stanford University, Stanford, CA 94305.

properties that vary throughout its length. Results from multiscale modeling suggest that variation in the sequence repeat impacts collagen's mechanical properties.^{10, 11} Regional models have been proposed to delineate the heterogeneity of the observed properties, but the boundaries of the regions are still incompletely defined.^{6, 12, 13} Structural studies have been performed to characterize the sequence-dependence of triple helix properties in detail.¹⁴ However, because of collagen's size and fibrous structure, atomic-level analysis of the full-length protein has been difficult. The crystal structure of rat tail tendon collagen determined by fiber diffraction¹⁵ provides a low resolution view of the trimer within a native fibril. However, the C α -only structure does not permit detailed analysis of the atomic interactions, and does not provide a picture of the dynamic properties of the triple helix. Crystallographic and NMR studies provided atomic level detail of the triple helical conformation and NMR and molecular dynamics have revealed aspects of collagen's dynamic behavior.¹⁴ One such finding is that less stable regions of the triple helix may be less tightly wound on average, with longer intermolecular backbone hydrogen bonds contributing to the decreased stability.^{16, 17} However, the sequences investigated in these studies are limited to those of model peptides, most of which are repeating Gly-Pro-Hyp sequences, rather than native collagen sequence.

To further our understanding of the structure and dynamics of the native collagen, we carried out a 10 ns molecular dynamics simulation of the complete, heterotrimeric, triple helical domain of human type I collagen. We designed a relational database to store the raw and derived data and developed specialized analyses protocols to handle analysis of the large volume of data generated. The simulated structures capture structural variation consistent with experimental structural and biophysical studies. Our simulations support hypotheses about collagen's dynamic heterogeneity and lend insights into the properties and regions of the native collagen triple helix. Finally, the simulation structures can serve as the basis for future molecular dynamics studies on the effects of mutations and polymorphisms on collagen structural properties and their relationship to the higher order forms of the protein.

2. Materials and Methods

2.1. Software and parameters

Molecular dynamics simulations were performed with GROMACS version 3.3¹⁸ and managed by the Folding@home distributed computing servers.¹⁹ Parameters are from the AMBER-99 force field²⁰ supplemented by published values for hydroxyproline.²¹ Data are stored in a postgres version 8.1 relational database.

2.2. Construction of the starting structures

The 1014 amino acid-long sequences of the peptide chains comprising the triple helical region of human type I collagen were taken from residues 179-1192 of GenBank entry NP_000079 for α 1(I) and residues 91-1104 of AAB59374 for α 2(I). The chains were assembled into a heterotrimer with chain order α 1(I)- α 2(I)- α 1(I). Coordinates of an idealized triple helix were generated with GENCOLLAGEN²² using default parameters, with all Y-position prolines converted to 4'-

hydroxyproline. The initial helical symmetry parameters for the Gly-X-Y triplets were: $\phi = -74^\circ$, $\psi = 170^\circ$, $\omega = 180^\circ$ for glycine, $\phi = -75^\circ$, $\psi = 168^\circ$, $\omega = 180^\circ$ for the residue in the X position, and $\phi = -75^\circ$, $\psi = 153^\circ$, $\omega = 180^\circ$ for Y-position residues.

The full triple helical region was split into 24 overlapping fragments to facilitate parallelization of the computation (Figure 1). The termini of all fragments were capped with neutral acetyl and N-methylamine end groups. All fragments include 85 residues per chain except the C-terminal fragment which spans 48 triple helix positions, and overlap their adjacent fragments by 42 residues. The starting residue of each fragment is offset from the previous fragment by 42 residues, giving an additional overlap of one helix position between every other fragment. For example, fragment 0 is residues 1-85, fragment 1 is 43-127, and fragment 2 is 85-169.

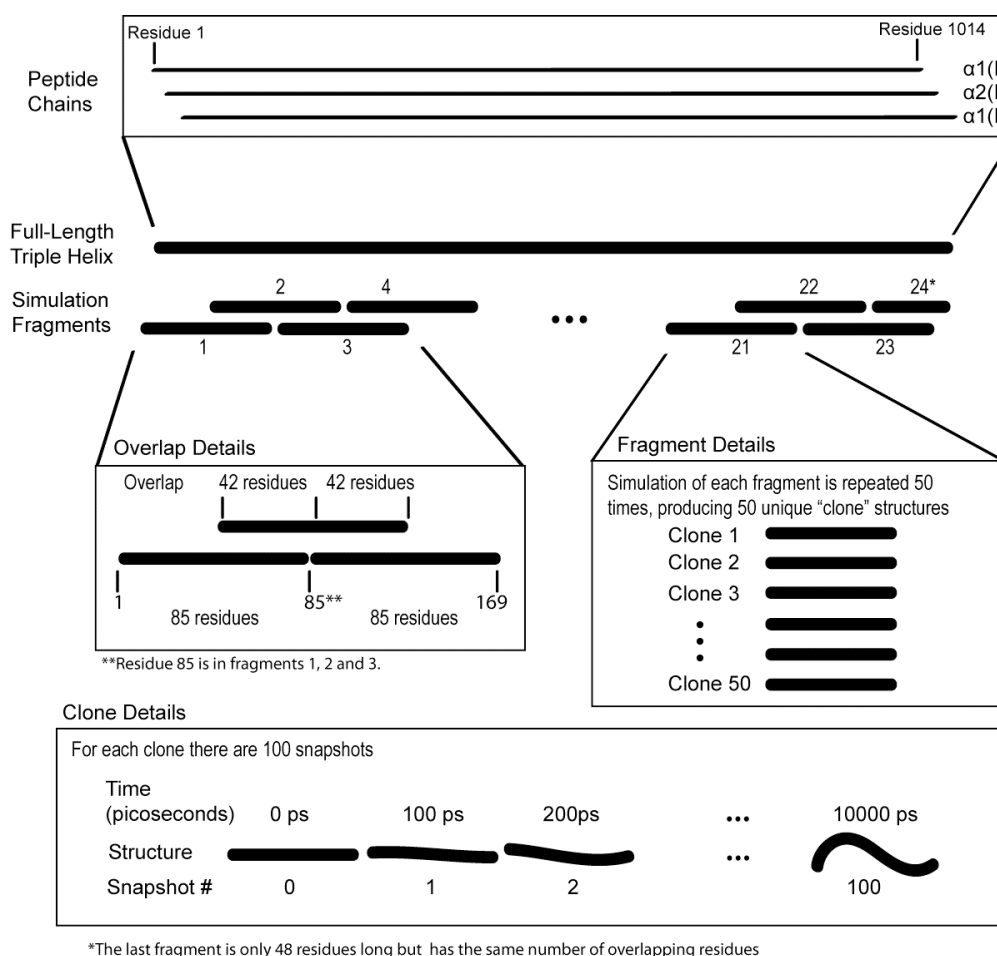


Fig. 1. Schematic representation of the molecular dynamics simulation. The full triple helical region, comprised of 1014 residues per polypeptide chain with chain order $\alpha 1(I) - \alpha 2(I) - \alpha 1(I)$, was modeled as 24 overlapping fragments. Each fragment is associated 50 clones, trajectories of 50 simulations performed for that fragment. Simulated structures are sampled every 100 ps, yielding 101 snapshots for each clone. Snapshots are numbered sequentially with snapshot 0 the starting structure for that clone.

2.3. Molecular Dynamics Simulations

Periodic boundary conditions were applied and each fragment was aligned along the Z-axis and then solvated in a box of TIP3P water.²³ A box size of 45 x 45 x 380 Å was selected for all fragments, giving at least 12 Å water boundary on all sides. To equilibrate the water, position restraints applied to protein heavy atoms and simulations were performed for 100 ps at constant temperature (300°K) and pressure (1 atm).

After equilibration, the simulation was conducted at a constant temperature of 300 °K and continuous pressure of 1 atm for 10 nanoseconds with no restraints. The Nose-Hoover thermostat²⁴ and isotropic Berendsen barostat²⁵ were used for temperature and pressure control, respectively. All covalent bonds that involve hydrogen atoms were constrained with the LINCS algorithm.²⁶ A 2 fs time step was used for all simulations. The total linear and angular momentum were removed at every time step, for protein and water separately. Electrostatic forces were calculated using reaction field with a cutoff distance of 12 Å.

Simulations for each fragment were repeated 50 times starting from the same initial coordinates but with different initial random velocities. Each of the fifty simulations per fragment represents one 10 ns trajectory or clone (Figure 1). Coordinates were written every 100 ps for a total of 101 snapshots for each trajectory including the starting structure.

2.4. Derived Structural Metrics

Helical radius was defined for each residue *i* as the distance from its C α atom to the centroid defined by the C α of residues *i*, *i*-1 and *i*-2 on chains A, B and C respectively. Hydrogen bond lengths were calculated as the distance between the amide hydrogen atom of glycine in one chain and the carbonyl oxygen atom of the X-position residue in an adjacent chain for each Gly–X–Y triplet. Since the large values observed in some structures suggest disruption of hydrogen bonding, this measure will be referred to as the H_N-O_C distance. Backbone dihedral angles were calculated by computing the three vectors *v*₁, *v*₂, and *v*₃ between C-N-C α -C for ϕ and N-C α -C-N for ψ and then taking the arc tangent of $|v_2| v_1 \cdot [v_2 \times v_3]$, $[v_1 \times v_3] \cdot [v_2 \times v_3]$.

Average interchain H_N-O_C distance, ϕ angle, and ψ angle were calculated over all simulations, using ~7,500 structures from the last 5 ns of each 10 ns simulation. Autocorrelations were computed using the mean values over each of the three chains computed as a function of residue offset.

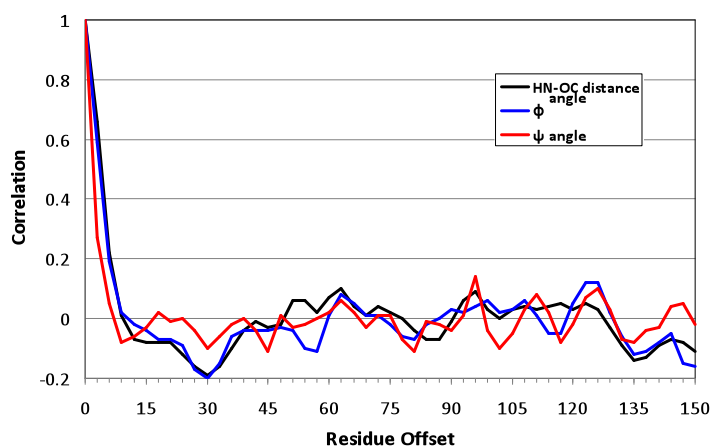


Fig. 2 Autocorrelation of glycine H_N-O_C distance, ϕ angle, and ψ angle as a function of residue offset. The maximum offset shown is 150 residues, which is a region spanned by approximately three simulated fragments.

3. Results

3.1. *Molecular Dynamics Simulation*

The complete triple helical region of heterotrimeric collagen type I, with 1014 amino acids per chain, was simulated for 10 ns. The simulation included ~1.8 million atoms, including solvent, and took approximately 11 months using the CPUs of over a quarter of a million computers. The calculation was parallelized by fragmenting the protein into 24 overlapping triple helical segments (Figure 1). Each fragment included 85 residues per chain, the shortest length needed to minimize end effect issues.²⁷ Fifty 10-ns trajectories (clones) were computed for each fragment, for a total of 1200 simulations.

The overlap in the fragments means that most residues occur in two different sets of 50 clones. To generate a composite set of structures for subsequent analysis in which each residue is represented once, for each residue we used only the simulation with that residue closer to the center of the fragment. This strategy eliminated potential artifacts that might have resulted from fraying at the ends of fragments, which had been observed previously in simulations of triple helical structures.²⁸ Angles at the overlap boundaries were calculated using residues from each side of the boundary.

To facilitate analysis and interpretation of the results, raw and summarized data were stored in a relational database along with biological annotations retrieved from the collagen database COLdb.²⁹ Stored data include 5,072,422 H_N-O_C distances, 15,247,566 each of ϕ and ψ backbone dihedral angles, 15,217,266 helical radii, and 1,615 biological features. The biological data encompass information from multiple scales, including: (1) thermostability of Gly-X-Y triplets, (2) features of the procollagen trimer, such as experimentally observed folding domains, (3) fibril-level features, e.g. ligand interaction sites, (4) characteristics of assembled fibers, such as gaps and bands visible in electron micrographs, and (5) patient phenotypes associated with specific mutation sites. The database and other supplementary materials are available online at simtk.org.

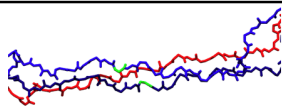

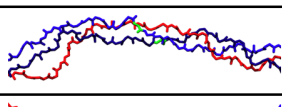
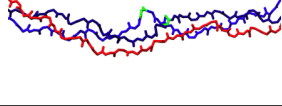
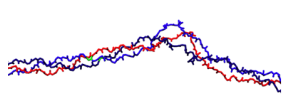
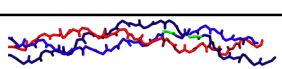
3.2. *Validation of the Simulated Structures*

Several tests were performed to ensure the validity of the resulting composite structures. First, visual inspection of the 1,200 trajectories confirmed the absence of artifacts resulting from use of periodic boundary conditions. Significantly, the solute did not directly interact with its periodic image. Second, analysis of the potential energies of the structures showed that the system was stable over the course of the simulation. The potential energy averaged over all simulations for all fragments equilibrated at approximately -980,000 KJ/mol. Third, the autocorrelations of H_N-O_C distance, ϕ angle, and ψ angle were calculated for each glycine to identify any periodicity resulting from fragmenting the protein at fixed locations (Figure 2). No significant peak was found at 42, suggesting that the fragmentation did not introduce periodicity artifacts. Interestingly, there is a small negative correlation between glycines separated by 30 residues. The source of this correlation is unknown since known repetitive features present in the structures, such as super-helical turns and typical fragment length, are either greater or less than 30 residues.

3.3. Analysis of the Simulation Structures

Visual inspection revealed that the composite structures generally maintained a triple helical conformation throughout their length. However, the triple helical structure was not uniform, with some regions more tightly wound on average than others. For example, Table 1 shows sample regions of type I collagen with disease-associated glycine mutations that differ in the position of the mutation on the $\alpha 1$ chain, whether multiple distinct substitutions have been observed at that position, and the location in the triple helix with respect to known ligand-binding sites.

Table 1. Sample simulation structures of regions of native collagen encompassing observed glycine mutation sites in the $\alpha 1(I)$ chain. The mutation sites are highlighted in green.

| Chain Location | Gly \Rightarrow Mutation | Structural Description around mutation | Representative Model of ~42 residues near mutation |
|-----------------------|--|---|--|
| 106 | Ala, Ser | Unstable region; very loose region |  |
| 220 | Ala, Asp, Cys | Moderately loose region; bending observed |  |
| 382 | Arg, Cys, Ser | Normal region with mild variation; helix loosens and tightens |  |
| 415 | Cys, Asp, Ser, Ser | Very tight region with little variation; region becomes unstable near residue 437 |  |
| 844 | Ser, Val | Unstable region; very loose region with high variation (842-858); tight helix C-terminus to residue 858 |  |
| 1009 | Ser, Val | Very tight region with little variation |  |

The conformations were characterized by mean ϕ and ψ angles (Figure 3). The average ϕ angle ranged between -66.1° and -87.6° and the average ψ angle between 151.5° and 160° . We further characterized the structural variation by computing three additional measures: H_N-O_C distance between glycines on adjacent chains, the number of residues per turn, and helical radius. Interchain H_N-O_C distance averaged 2.6 \AA , with a minimum of 1.5 \AA . Average helix radius measured at glycine positions ranged from 2.8 \AA to 7.3 \AA (Figure 3). Visual inspection of the structures revealed that the sharp peaks in the figure correspond to regions of unwinding of the triple helix. The number of residues per turn ranged from 4.0 to 5.1, with an average of 4.9. The number of residues per turn inversely correlated with helix radius and H_N-O_C distance, and deviations from the average angles corresponded to increased helix radius and hydrogen bond length. Detailed data for these four measurement types are available for download at simtk.org. Although these measures were uniform in the ideal starting conformation, they vary along the

length of the triple helix as a result of the simulation, in agreement with results from experimental studies revealing local variations in helical twist.^{30, 31}

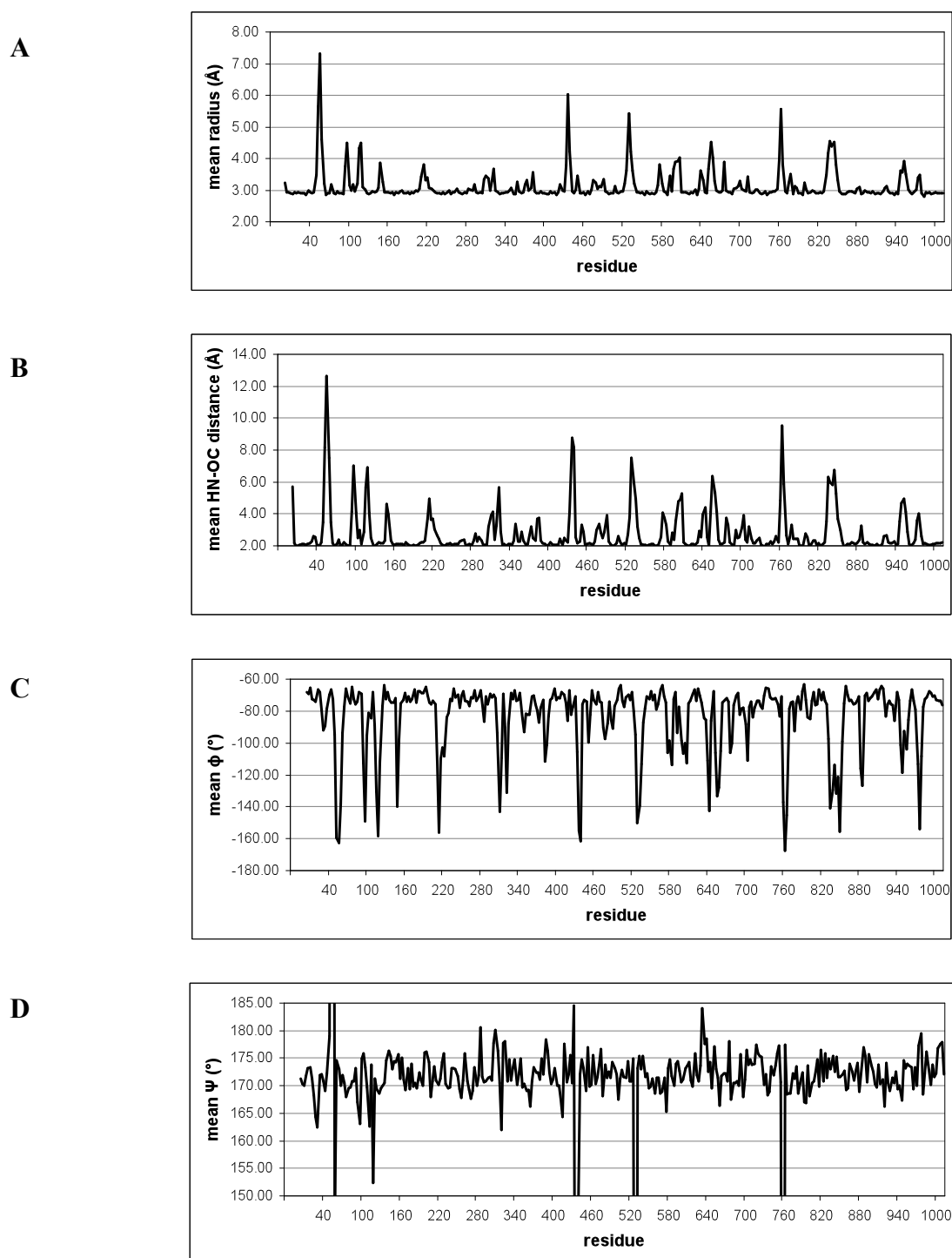


Fig. 3. Average properties at each glycine position from the second half of the simulation. (A) radius. (B) H_N-O_C distance. Long distances indicate the loss of the hydrogen bond in at least a subset of structures (next page). (C) ϕ angle. (D) ψ angle.

Table 2 contrasts ϕ and ψ angles calculated from our simulations with the corresponding angles calculated from crystal structures of model collagen peptides. The angles are organized based on the position of the residue within a triplet (Gly, X, or Y) and by their residue's type (imino, meaning the triplet contains a proline or hydroxyproline, or amino, meaning the triplet contains no proline or hydroxyproline). For most angles (10/12), the crystal structure values are well within ten degrees of the simulation values, and were not examined in more detail. The two most interesting values are for the ϕ angles of the residues in the Gly and X locations, where the values differ by 13 and 16 degrees between the simulation and the crystal structure. These differences correspond to opening of the canonical triple helical structure.

Table 2. Φ and Ψ angles for Amino Residue Pairs and Imino Residue Pairs

| Angle | Residue Position | Residue Type | Crystal Structure ^a | Simulation | Difference |
|--------|------------------|--------------|--------------------------------|------------|------------|
| ϕ | Gly | Amino Pair | -68 | -81 | -13 |
| | | Imino Pair | -72 | -74 | -2 |
| | X | Amino Pair | -71 | -87 | -16 |
| | | Imino Pair | -74 | -67 | 7 |
| | Y | Amino Pair | -66 | -69 | -3 |
| | | Imino Pair | -60 | -55 | 5 |
| ψ | Gly | Amino Pair | 167 | 173 | 6 |
| | | Imino Pair | 176 | 172 | -4 |
| | X | Amino Pair | 160 | 155 | -5 |
| | | Imino Pair | 163 | 156 | -7 |
| | Y | Amino Pair | 148 | 146 | -2 |
| | | Imino Pair | 152 | 151 | -1 |

Table 3 shows details for the ϕ angles for residues in the Gly and X positions, based on the type of residue occupying the Y position. The residues in the Y position are split into two groups. The first group consists of all γ -branched residues (Asp, Phe, His, Leu, Asn) plus glutamic acid and lysine (henceforth identified using their concatenated one-letter-codes, "DEFHKLN"). The second group consists of all other observed residues that are not part of the first group (identified using their one letter codes: "AGIMOQRSTV"). The ϕ angles for the DEFHKLN group show significant deviation from the values derived from the crystal structures.

^a From Rainey and Goh ³²

Table 3. Φ angle for Helical and Helix-Breaking Residues

| Angle | Residue Position | Y Residue Type | Crystal Structure ^b | Simulation | Difference |
|--------|------------------|-------------------------|--------------------------------|------------|------------|
| ϕ | Gly | D,E,F,H,K,L, or N | -68 | -87 | -19 |
| | | A,G,I,M,O,Q,R,S,T, or V | -68 | -77 | -9 |
| | X | D,E,F,H,K,L, or N | -71 | -88 | -17 |
| | | A,G,I,M,O,Q,R,S,T, or V | -71 | -76 | -5 |

Figure 4 shows the frequency of observed ϕ angles for all residues in the Gly position, for each observed residue type in the Y position. The seven dotted lines indicate residues types that are part of the DEFHKLN group described above. The other ten lines represent members of the group that shows a preference for triple helices (the AGIMOQRSTV group). Recall from Table 2 that the mean ϕ angle for non-imino residues in the Gly position is -68° .

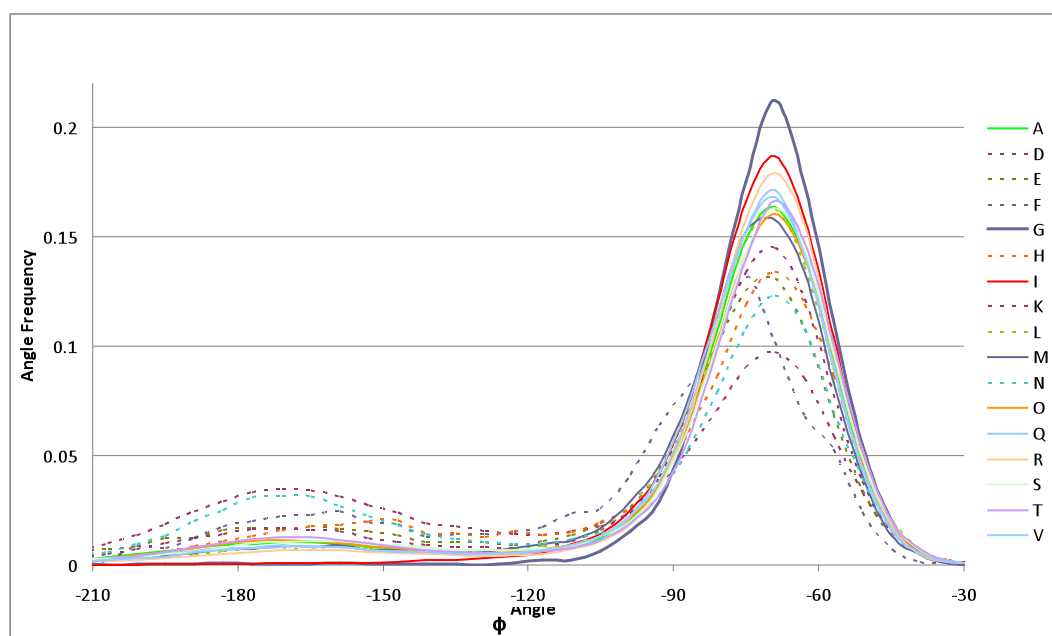


Fig. 4. Frequency of glycine ϕ angles graphed by the residue in the Y position.

Figure 5 is similar to Figure 4, but shows the frequency of observed ϕ angles for all residues in the X position. The mean ϕ angle for non-imino residues in the X position is -71° .

^b From Rainey and Goh³²

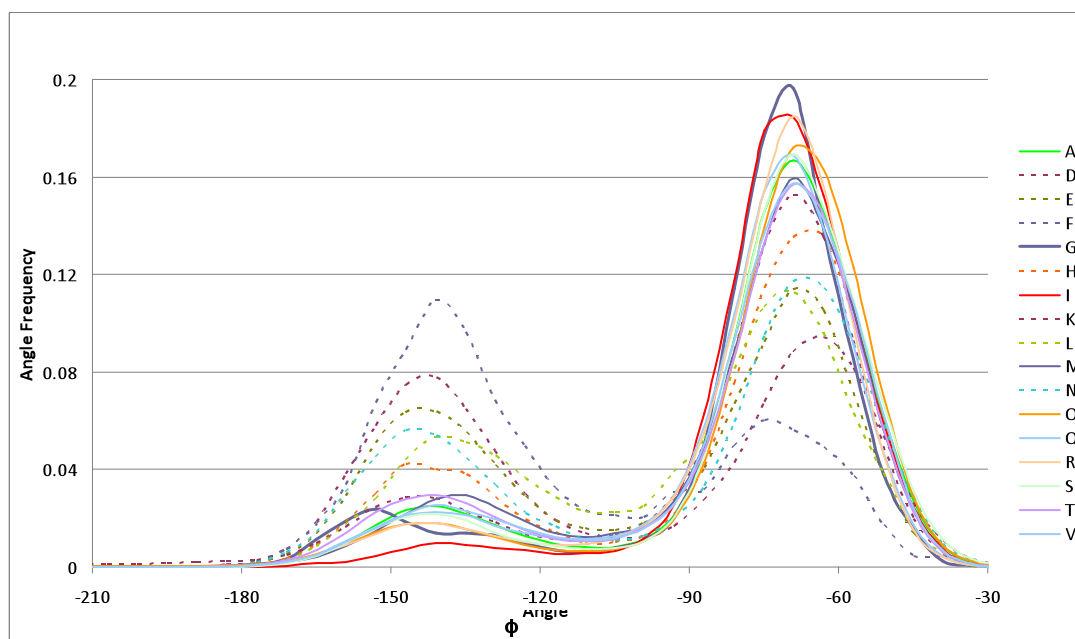


Fig. 5. Frequency of observed ϕ angles of X-position amino acids.

4. Discussion

We have conducted a 10 ns molecular dynamics simulation of the complete triple helical domain of human type I collagen as a first step towards understanding the dynamics of this critical region, its sequence dependence, and its relationship to higher order forms of the protein. Due to the large size of the protein, the calculation was made feasible by splitting the protein into overlapping fragments and the analyses were facilitated by storing the results in a customized relational database.

The simulated molecules capture structural variation in the triple helical domain consistent with experimental structural and biophysical studies and the atomic resolution of the simulations enables more precise definition of the region boundaries. Residues 1-85, identified by Makareeva, et. al.,⁶ as the high stability N-anchor region, have short radius and H_N-O_C distance in our simulations. However, this tightly wound region is interrupted at residue 55 in our simulations, where we see an average radius of 7.3 Å, the highest in the protein (Figure 3). The molecular basis for the implied decrease in stability is unknown but may be due to electrostatic repulsion in a cluster of Asp residues at positions 53 and 54 in $\alpha 1(I)$ and position 54 in $\alpha 2(I)$. The results also suggest that the high stability region may extend to glycine 91, and that the proposed adjoining microunfolded region, captured by long H_N-O_C distances, may span glycine residues 94 - 121. The next two areas of largest unfolding in our simulations, with maximum unwinding at glycines 436 and 763, are contained within Makareeva, et. al.'s mid-flex and C-flex low stability regions. These residues may represent positions at which unfolding initiates in these areas.

Tables 2 and 3 show how dynamic, solvated collagen molecules differ from model collagen crystal structures examined by Rainey and Goh.³² It is interesting that there is little difference

between the ϕ and ψ angles shown in Table 2, despite the difference in environment and the difference in sequence. The only significant difference is in the ϕ angles for residues in the Gly and X positions that are part of triplets containing no proline or hydroxyproline residues. Data mining showed a strong correlation between these ϕ angles and the type of residue occupying the Y position of the triplet (Table 3). Five of the residues in the DEFHKLN group are γ -branched (Asp, Phe, His, Leu, Asn), which are known to be destabilizing in model collagen systems,³³ so their connection with non-helical structures is not surprising. The reason why glutamic acid and lysine also correlate with non-helical structures is under investigation.

Although the calculated structures are similar to experimentally determined structures of collagen-like peptides, the simulation did not reproduce the microfibril conformation of the triple helices observed in the crystal structure conformation of rat tail tendon collagen.¹⁵ This is not surprising due to differences between the simulated and crystallized collagens in: primary sequence, scale (trimer and fibril), environment (water vs in situ), tissue source, and post-translational modifications. This is consistent with there being important structural differences between isolated (solvated) heterotrimers and the trimers in the more complex fibril structure, in which the collagen proteins are closely packed and associated with proteoglycans and other factors.

The complexity of the hierarchical conformations of collagen has made it difficult to determine experimentally the structure of native collagen at high resolution, and the large size of the protein has previously prohibited full atomic modeling of its structure and dynamics. We were able to accomplish molecular dynamics simulation of the full-length triple helix through technological improvements dependent on the accessibility of hundreds of thousands of computers. The resulting models are an important starting point for investigating the unique hierarchical conformations of collagen and for studying the effects of disease-associated mutations on collagen structure.

5. Acknowledgments

This work was supported by NIH grant AR051582 (to TEK) and the Osteogenesis Imperfecta Foundation and the Children's Brittle Bone Foundation (to DLB). We thank Sanghyun Park and Vijay Pande for running the calculations on Folding@Home, the Folding@Home folders for donating CPU time, and Simbios for simtk.org (GM072970).

References

1. P. H. Byers, *Online Metabolic and Molecular Bases of Inherited Disease* **22**, 5241 (2001).
2. T. J. Wess, *Adv Protein Chem* **70**, 341 (2005).
3. C. M. Kielty and M. E. Grant in *Connective Tissue and Its Heritable Disorders*, P. M. Royce and B. Steinmann, Editors. Wiley-Liss 159 (2002).
4. W. V. Arnold, A. Fertala, A. L. Sieron, H. Hattori, D. Mechling, H. P. Bachinger and D. J. Prockop, *J Biol Chem* **273**, 31822 (1998).
5. E. Makareeva, W. A. Cabral, J. C. Marini and S. Leikin, *J Biol Chem* **281**, 6463 (2006).

6. E. Makareeva, E. L. Mertz, N. V. Kuznetsova, M. B. Sutter, A. M. DeRidder, W. A. Cabral, A. M. Barnes, D. J. McBride, J. C. Marini and S. Leikin, *J Biol Chem* **283**, 4787 (2008).
7. A. Steplewski, I. Majsterek, E. McAdams, E. Rucker, R. J. Brittingham, H. Ito, K. Hirai, E. Adachi, S. A. Jimenez and A. Fertala, *J Mol Biol* **338**, 989 (2004).
8. D. L. Bodian, B. Madhan, B. Brodsky and T. E. Klein, *Biochemistry* **47**, 5424 (2008).
9. G. A. Di Lullo, S. M. Sweeney, J. Korkko, L. Ala-Kokko and J. D. San Antonio, *J Biol Chem* **277**, 4223 (2002).
10. A. Gautieri, S. Uzel, S. Vesentini, A. Redaelli and M. J. Buehler, *Biophys J* **97**, 857 (2009).
11. S. G. Uzel and M. J. Buehler, *Integr Biol (Camb)* **1**, 452 (2009).
12. H. P. Bachinger, N. P. Morris and J. M. Davis, *Am J Med Genet* **45**, 152 (1993).
13. J. C. Marini, A. Forlino, W. A. Cabral, A. M. Barnes, J. D. San Antonio, S. Milgrom, J. C. Hyland, J. Korkko, D. J. Prockop, A. De Paepe, P. Coucke, S. Symoens, F. H. Glorieux, P. J. Roughley, A. M. Lund, K. Kuurila-Svahn, H. Hartikka, D. H. Cohn, D. Krakow, M. Mottes, U. Schwarze, D. Chen, K. Yang, C. Kuslich, J. Troendle, R. Dalgleish and P. H. Byers, *Hum Mutat* **28**, 209 (2007).
14. B. Brodsky and A. V. Persikov, *Adv Protein Chem* **70**, 301 (2005).
15. J. P. Orgel, T. C. Irving, A. Miller and T. J. Wess, *Proc Natl Acad Sci U S A* **103**, 9001 (2006).
16. R. J. Radmer and T. E. Klein, *Biochemistry* **43**, 5314 (2004).
17. R. J. Radmer and T. E. Klein, *Biophys J* **90**, 578 (2006).
18. E. Lindahl, B. Hess and D. van der Spoel, *J Mol Model* **7**, 306 (2001).
19. <http://folding.stanford.edu/>
20. J. Wang, P. Cieplak and P. A. Kollman, *J Comput Chem* **21**, 1049 (2000).
21. S. Park, R. J. Radmer, T. E. Klein and V. S. Pande, *J Comput Chem* **26**, 1612 (2005).
22. C. C. Huang, G. S. Couch, E. F. Pettersen, T. E. Ferrin, A. E. Howard and T. E. Klein, *Pac Symp Biocomput* 349 (1998).
23. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J Chem Phys* **79**, 926 (1983).
24. D. J. Evans and B. L. Holian, *J Chem Phys* **83**, 4069 (1985).
25. H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J Chem Phys* **81**, 3684 (1984).
26. B. Hess, H. B. Herman, J. C. B. Johannes and G. E. M. Fraaije, *J Comput Chem* **18**, 1463 (1997).
27. R. J. Radmer and T. E. Klein, unpublished results.
28. S. Park, T. E. Klein and V. S. Pande, *Biophys J* **93**, 4108 (2007).
29. D. L. Bodian and T. E. Klein, *Hum Mutat* **30**, 946 (2009).
30. J. Emsley, C. G. Knight, R. W. Farndale and M. J. Barnes, *J Mol Biol* **335**, 1019 (2004).
31. R. Z. Kramer, J. Bella, P. Mayville, B. Brodsky and H. M. Berman, *Nat Struct Biol* **6**, 454 (1999).
32. J. K. Rainey and M. C. Goh, *Protein Sci* **11**, 2748 (2002).
33. A. V. Persikov, J. A. Ramshaw, A. Kirkpatrick and B. Brodsky, *Biochemistry* **39**, 14960 (2000).