Atomistic Simulations of Collagen Fibrils

I. Streeter and N.H. de Leeuw

Institute of Orthopaedics and Musculoskeletal Science, University College London, UK
Department of Chemistry, University College London, UK

Abstract— **Computational molecular dynamics simulations are used to investigate the supramolecular structure of a collagen fibril at an atomistic scale of resolution. The simulations use a newly developed protocol that allows the overall fibrillar structure of collagen to be studied, rather than only providing data on individual, fully solvated proteins. The data generated provide new information regarding the arrangement of collagen proteins and water molecules within a fibril, and the nature of the inter-protein attractions that are responsible for the process of fibrillogenesis. It is anticipated that this new approach to modelling a fibril will lead to a better understanding of how collagen's material properties depend on the dynamic behaviour of the underlying proteins and water molecules.**

Keywords— **collagen, molecular dynamics, simulation, fibril, structure.**

I. Introduction

Collagen is used in the field of tissue engineering as a material for building scaffolds that have the potential to direct cells into forming new tissue. It has an advantage over any synthetic material as it is naturally the principal protein component of the extracellular matrix, and so it is biocompatible and has low immunogenicity. Collagen is a highly complex material: the underlying proteins have a very long, non-repeating amino acid sequence, and they aggregate in combination with water to form fibrils with a specific supramolecular arrangement. This complexity in structure makes collagen correspondingly complex to handle as a material. A typical protocol for making a collagen scaffold might involve stimulating a solution of the proteins to gel into fibrils, compressing the gel to remove excess water, and then using cells to slowly resorb and remodel the collagen into the micro-architecture of a natural tissue [1].

Generally, when manipulating a material for a given function, it is advantageous to understand how its underlying structure contributes to its material properties. Molecular dynamics (MD) simulations provide an effective way of bridging between these atomistic-scale interactions and macroscopic observable properties [2]. In a typical MD simulation, a molecular system is first built on an atom by atom basis, and the system dynamics are then calculated by applying classical equations of motion to each atom. The atoms are allowed to interact with each other as they would in the real system, for example via covalent bonds, electrostatic interactions, or Van der Waals forces. The calculated trajectory represents a statistical ensemble of different atomic configurations accessible to the system, and so the material's bulk properties can be predicted by taking appropriate averages from this ensemble.

Although MD simulations have long been used in both the fields of protein science and materials science, simulations of collagen have so far remained fairly limited in terms of what they can tell us about collagen as a material, because most MD simulations to date have only modelled short segments of a collagen protein under conditions where they are fully solvated by water molecules [3-5]. Although this has given us insight into the nature of collagen's triple helix conformation, it does not tell us about the molecular organisation within a fibril, or how the dynamic behaviour of the proteins and water molecules are affected when they are packed together into a fibrillar environment. Extending these molecular dynamics simulations from a single molecule to an entire collagen fibril would ordinarily be problematic because such an increase in the system size would require prohibitively large computational resources.

We have developed a new protocol for modelling the structure of the full collagen fibril, including its constituent proteins and intrafibrillar water molecules, whilst retaining the detail of the system at the atomistic scale. This new approach works by modelling the system as a densely packed unit cell and then applying periodic boundary conditions in such a way that it generates the supramolecular structure inferred from x-ray diffraction experiments. The periodic unit cell is small compared to the size of an entire fibril, and so the computational requirements for this approach are relatively modest. This is an approach that would be more commonly associated with modelling crystalline solid materials, but we find that it can work for collagen even though there are many differences between a fibril and a genuine crystal, in that collagen proteins are large, flexible, disordered, and partially solvated.

A. El Haj and D. Bader (Eds.): 8th International Conference on Cell & Stem Cell Engineering, IFMBE Proceedings 30, pp. 39–42, 2011.
www.springerlink.com

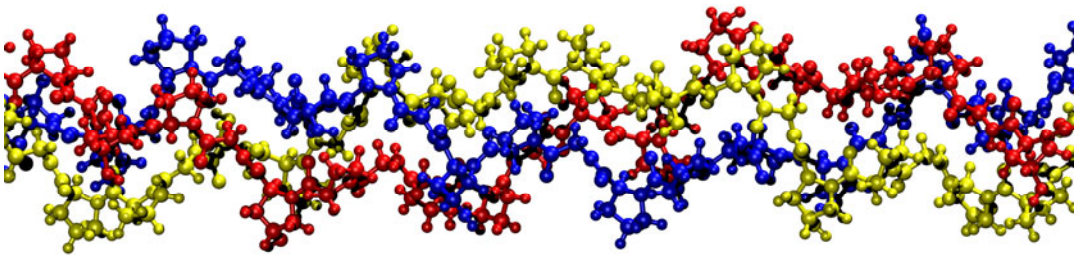

Fig. 1 Triple helical segment of a collagen protein showing atomistic resolution

II. Methods

The first stage in performing these simulations was to build the system with an appropriate starting conformation. The primary structure of the type I collagen protein was taken from the sequenced genes COL1A1 and COL1A2. The local conformation of collagen's triple helical structure is known from high resolution x-ray crystallography of collagen-mimetic peptides [6]. The supramolecular arrangement of collagen proteins is also known from x-ray diffraction of the fibril, but the spatial resolution of this latter experiment is a lot lower [7]. The initial system conformation was therefore inferred by combining these data: the local conformation was consistent with the high resolution data for collagen-mimetic peptides and the overall protein shape and topology was consistent with the low resolution data of the entire fibril. The programme THeBuScr was used to generate the local conformation of the helix [6], and then a python script written in-house was used to translate this into the correct supramolecular topology. Entry 1y0f in the online RCSB Protein Data Bank contains the crystallographic details of the supramolecular structure of collagen, including the orientation and position of the proteins in relation to the periodic unit cell [8].

Water molecules were placed in the unit cell to fill all the intrafibrillar spaces. For each collagen protein, 11980 waters were added, which equates to 0.75 g water / g collagen, and is therefore consistent with previous studies of fibrillar water content [9]. This quantity was selected using a trial and error approach: any more water caused the unit cell to expand beyond its crystallographic dimensions, but any less water led to a contraction.

This ratio of protein to water highlights the key difference between these simulations and standard simulations of proteins. Proteins are nearly always modelled by using a unit cell and by applying periodic boundary conditions, but normally an excess of water is used so that the unit cell is large to the extent that proteins from neighbouring cells do not directly interact with each other. Our simulations use relatively little water in such a way that the unit cells reflect the densely packed structure of the collagen fibril.

Molecular dynamics were performed using the software Amber 9, which is commercially available and is designed specifically for simulating large biomolecules [2]. A standard protocol was followed for the simulations: the system was first minimised to remove any bad contacts, and then heated and equilibrated at constant volume for 120 ps. Finally the trajectory was calculated under constant pressure conditions for 20 ns.

III. Results

The MD simulation of the collagen fibril ran successfully for 20 ns and the system was stable throughout. The tropocollagens flexed and gyrated within their positions due to thermal energy, but they remained in a triple helix conformation and the overall supramolecular arrangement of the proteins was preserved. Water molecules were more dynamic: they appeared to be fairly fluid as they flowed and rotated within the intrafibillar spaces left vacant by the proteins.

Figures 1-3 show various images of the system taken at the end of the 20 ns trajectory. Figure 1 shows a short segment of a tropocollagen with atomistic resolution. Each alpha chain is shown in a different colour, and it can be seen how they twist around each other in a triple helix conformation. Figures 2 and 3 show cross sectional slices of the fibril, taken perpendicular to and parallel to the fibril's long axis, respectively. For both figures, each protein is shown in a different colour, and it can be seen how they have a parallel allignment and are tightly packed to the extent that they interact directly with one another. Water molecules are also shown (red and white), and they fill all of the intrafibrillar spaces and are largely disordered in terms of their positions and orientation.

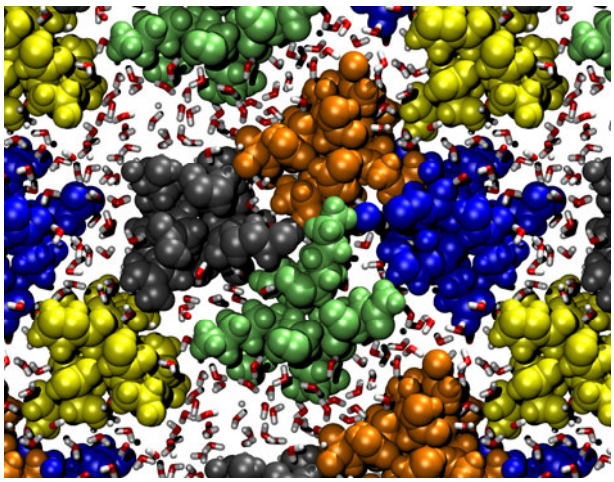

Fig. 2 Axial cross section of the fibril

The energy within the system decreased gradually for the first 8 ns, indicating that the molecules were slowly relaxing during this period to find more stable positions and to optimise intermolecular interactions. This is a fairly long time period for a relaxation, but it is no surprise given that the collagen molecules are so large, and that the starting conformation did not come directly from an experimental atomistic structure.

Although the system energy was stable after the initial 8 ns, the collagen proteins continued to gyrate and flex, and this was therefore interpreted as motion due to thermal energy rather than motion due to any relaxation process. The proteins were observed to be more flexible in the gap region of the fibril where the molecular packing is not as dense as in the overlap region. This can be quantified in terms of the root mean square fluctuation (rmsf) of the protein's alpha carbon atoms; that is the root mean square distance of each alpha carbon from its average position over the course of the simulation. In the densely packed overlap region, the average rmsf of alpha carbon atoms was 1.5 Å; the corresponding rmsf in the gap region was 2.7 Å. It is important to note that this increased flexibility in the gap region is manifested as the bulk movement of entire proteins, and not as any conformational change in the helix itself.

IV. Discussion

The simulations of the collagen fibril described here have a number of unusual features that are worth emphasising. Typically, MD simulations of proteins only ever consider a single protein in a fully solvated state, not in a densely packed environment as in a collagen fibril. Conversely, densely packed environments are often considered in computational studies of materials such as crystalline solids, but these studies rarely have to consider molecules as large as type I collagen, which is over 300 nm long. Our simulations have therefore combined techniques used by both of these branches of computational science to achieve a completely new approach to modelling the collagen fibril with atomistic resolution.

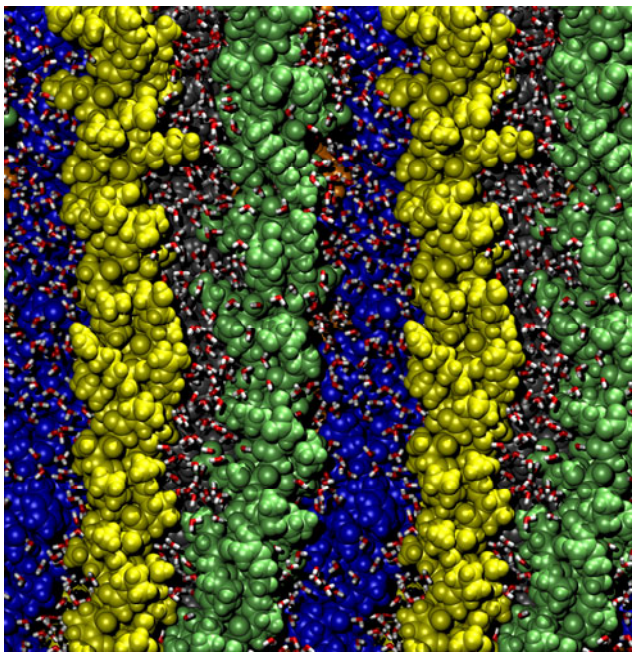

Fig. 3 Lateral slice through the fibril

The simulations reported here give a helpful insight into the nature of the collagen fibril. They give an impression of a fibril that is dynamic rather than static, with collagen proteins continually flexing and gyrating within their fibrillar positions, and with water molecules lubricating this motion by continually flowing and rotating in all of the intrafibrillar spaces. This flexibility of structure is particularly apparent in the gap region of the fibril, as indicated by the rmsf of the alpha carbon atoms in this region. The tropocollagens are less densely packed in this gap region, and there is a higher proportion of water molecules, which increases fluidity.

An important feature of this simulation method is that it allows a study of the manner in which the densely packed collagen proteins interact with each other. One type of important interaction is the inter-protein hydrogen bond, which could be an important factor for bringing strength to the collagen fibril and for driving the process of fibrillogenesis. Such inter-protein hydrogen bonds were observed to be a common feature in these simulations. They generally occurred either between two hydrophillic amino acid sidechains from two neighbouring tropocollagens, or between the protein backbone

of one protein and an amino acid side chain in its neighbour. Conversely, it was very rare to observe a hydrogen bond between two protein backbones in two neighbouring tropocollagens. All inter-protein hydrogen bonds were dynamic and transient; they were continually being broken and reformed due to the fluidity in the fibril structure. These observations provide new insight into the nature of protein-protein interactions, which would not have been possible had we only performed a more traditional MD simulation of a single fully solvated protein.

It is interesting to consider the role of intrafibrillar water molecules in the collagen fibril, as it is known from experiments that water is critical to collagen's material properties [10]. It can be seen from Figure 2 that water molecules surround all of the proteins, and in many places find themselves sandwiched between two neighbouring proteins. We find that on average 44.0% of all intrafibrillar water molecules are found within 3.2 Å of a protein, measured as the distance from the water's oxygen atom to the nearest protein non-hydrogen atom. This distance corresponds to the first hydration shell, and it demonstrates the high proportion of intrafibrillar water that is in direct contact with a protein. Water molecules in close association with proteins generally behave differently compared to the bulk liquid phase, and they have slower rates of translation and rotation. Previous computational studies have shown that collagen distorts the structure of water up to a distance of 6 Å away from its surface. We find that only 5.8% of intrafibrillar water molecules lie further than 6 Å away from their nearest protein in our simulations of a collagen fibril. This highlights a feature of collagen that is common to many biological systems in that the vast majority of the constituent water is very different in character to bulk water in the liquid phase. In the case of collagen, this is likely to be a pertinent point when considering mineralisation and the formation of apatite crystals, which are nucleated in the intrafibrillar spaces, and are likely to be sensitive to the nature and dynamics of their local environment.

V. Conclusions

The most important conclusion from this work is that it is now possible to simulate an entire collagen fibril whilst retaining atomistic resolution, provided one takes advantage of the periodic unit cell as revealed by x-ray diffraction. This is an important step forward that takes us away from modelling single collagen molecules, and brings us towards models that give us a better understanding of larger scale collagen structures. We acknowledge that the model described still cannot tell us everything about a collagen fibril. For example, because the simulations use an infinitely repeating unit cell, they tell us nothing about the outer surface of a collagen fibril. However, it is hoped that this new simulation protocol can be just the first in a series of steps towards achieving simulations that reveal the underlying interactions responsible for collagen's material properties.

The initial results from this simulation method have given an interesting description of the fibril. In particular, it has highlighted the dynamic and disordered nature of the proteins and water molecules at small length scales, even though the fibril has a fairly regular packing arrangement at larger length scales. It has told us new information regarding the positions of inter-protein hydrogen bonds, and it has demonstrated the close association of the water and protein phases, which emphasises that the water molecules are in a very different environment to a pure liquid phase and the proteins are in a very different environment to a fully solvated state.

References

1. Cheema U, Nazhat S, Alp Burçak et al. (2007) Fabricating tissues: an analysis of farming versus engineering strategies. Biotechnol Bioprocess Eng 12:9–14
2. Case D, Cheatham T, Darden T et al. (2005) The Amber biomolecular simulations programs. J Comput Chem 26:1668–1688
3. Klein T, Huan C (1999) Computational investigations of structural changes resulting from point mutations in a collagen-like peptide. Biopolymers 49:167–183
4. Ravikumar K, Hwang W (2008) Region-specific role of water in collagen unwinding and assembly. Proteins 72:3120–1332
5. Raman S, Parthasarathi R, Subramanian V et al. (2008) Role of length-dependent stability of collagen-like peptides. J Phys Chem B 112:1533–1539
6. Rainey J, Goh M (2004) An interactive triple-helical collagen builder. Bioinformatics 20:2485–2459
7. Orgel J, Irving T, Miller A (2006) Microfibrillar structure of type I collagen *in situ*. Proc Natl Acad Sci USA 103:9001–9005
8. RCSB Protein Data Bank at www.rcsb.org/pdb
9. Sivan S, Merkher Y, Wachtel E et al. (2006) Correlation of swelling pressure and intrafibrillar water in young and aged human intervertebral discs. J Orthop Res 24:1292–1298
10. Nomura S, Hiltner A, Lando J et al. (1977) Interaction of water with native collagen. Biopolymers 16:231–246

Corresponding author:

Ian Streeter
Institute of Orthopaedics and Muscoluskeletal Science
University College London
Brockley Hill
Stanmore HA7 4LP
UK
Email: i.streeter@ucl.ac.uk