MOLECULAR COOPERATIVITY AND COMPATIBILITY 
VIA 
FULL ATOMISTIC SIMULATION 

A Dissertation Presented 

By 

Kenny Kwan Yang 

to 

The Department of Civil and Environmental Engineering 
in partial fulfillment of the requirements 
for the degree of 

Doctor of Philosophy 
in the field of 

Civil Engineering 

Northeastern University 
Boston, Massachusetts 

May 2015
ABSTRACT

Civil engineering has customarily focused on problems from a large scale perspective, encompassing structures such as bridges, dams and infrastructure. However, present day challenges in conjunction with advances in nanotechnology have forced a re-focusing of expertise. The use of atomistic and molecular approaches to study material systems opens the door to significantly improve material properties. The understanding that material systems themselves are structures, where their assemblies can dictate design capacities and failure modes makes this problem well suited for those who possess expertise in structural engineering. At the same time, a focus has been given to the performance metrics of materials at the nanoscale, including strength, toughness, and transport properties (e.g., electrical, thermal). Little effort has been made in the systematic characterization of system compatibility – e.g., how to make disparate material building blocks behave in unison.

This research attempts to develop bottom-up molecular scale understanding of material behavior, with the global objective being the application of this understanding into material design/characterization at an ultimate functional scale. In particular, it addresses the subject of cooperativity at the nano-scale. This research aims to define the conditions which dictate when discrete molecules may behave as a single, functional unit, thereby facilitating homogenization and up-scaling approaches, setting bounds for assembly, and providing a transferable assessment tool across molecular systems.

Following a macro-scale pattern where the compatibility of deformation plays a vital role in structural design, novel geometrical cooperativity metrics based on the gyration tensor are
derived with the intention to define nano-cooperativity in a generalized way. The metrics objectively describe the general size, shape and orientation of the structure. To validate the derived measures, a pair of ideal macromolecules, where the density of cross-linking dictates cooperativity, is used to gauge the effectiveness of the triumvirate of gyration metrics. The metrics are shown to identify the critical number of cross-links that allowed the pair to deform together. The next step involves looking at the cooperativity features on a real system. We investigate a representative collagen molecule (i.e., tropocollagen), where single point mutations are known to produce kinks that create local unfolding. The results indicate that the metrics are effective, serving as validation of the cooperativity metrics in a palpable material system. Finally a preliminary study on a carbon nanotube and collagen composite is proposed with a long term objective of understanding the interactions between them as a means to corroborate experimental efforts in reproducing a d-banded collagen fiber.

The emerging needs for more robust and resilient structures, as well as sustainable are serving as motivation to think beyond the traditional design methods. The characterization of cooperativity is thus key in materiomics, an emerging field that focuses on developing a "nano-to-macro" synergistic platform, which provides the necessary tools and procedures to validate future structural models and other critical behavior in a holistic manner, from atoms to application.
ACKNOWLEDGEMENTS

First and foremost, I will take this opportunity to thank my beloved wife Lissa. I am extremely grateful and indebted to her for her support and most importantly her patience in these past few years. While it has been an up and down ride, she has constantly been there for me. This accomplishment is equally hers for bringing me peace and sanity during my trials.

I would also like to take this opportunity to thank my son Jeremy, while he does not understand much of what is happening, he has definitely brought a smile on my life. Unconsciously pushing me further by my desire to be a better person and example for him. I would also like to apologize to him for using my molecular dynamics text as bedtime stories, while I was trying to kill two birds with one stone.

My achievements are not my own, but a reflection of the people who have surrounded me during my life. At a personal level, it is an honour for me to thank my parents for all their hard work, guidance and unconditional support. As a parent I now fully understand all the sacrifices that you have made so that I can be standing here today. In particular to my father, Ken, who has unconsciously shown me that what matters ultimately is the actions and not words. I am equally thankful to my mother, Cristina, for her motivation, sacrifices and vision on my education. I also thank my brother, Young, who definitely was the person to look up to and BEAT, and provided an example of what to do and definitely what not to do. I wish to thank all my uncles and aunts in the United States, especially in no particular order, Yul, Chen, Maggie, Nancy and Ashraf, as well as my cousin Sophia and husband John, all of you have helped me and my family while we were here and I do not have enough words to express my gratitude for all your support and advice on the American way.
I want to express my gratitude for the many who have supported me during my studies at INTEC, and Northeastern. I would like to give special thanks to my mentor, boss and professor Rafael Taveras, he served as my professional inspiration and was the one who gave me the confidence to apply for the Fulbright scholarship, his experience has been invaluable. Additionally I would like to thank my colleagues from undergrad Kailing Joa, Kurt Hansen and John Mejia, for their friendship and support.

At Northeastern I wish to particularly acknowledge the smartest person I personally know in this world, Prof. Bernal, who gave me the insight as to what it takes to be part of graduate school and showed me the value of what it means to know the fundamental theory. In the same sentence I would like to extend my thanks to Dr. Michael Döhler from INRIA, not only for his friendship and support, but also for serving as my statistics and math consultant. I wish to extend my thanks to my friends, the original 427 Richards Hall, you guys know who you are, Laura Pritchard, and Jiangsha Meng, for your friendship, the laughs and the good times. I would also like to acknowledge my collaborators within NICE at Northeastern, with special thanks to Ruth and Ashley. I am grateful for the generous support by Dr. Hajjar, the chairman of CEE in addition to his advice and funding considerations.

Lastly, but most importantly, I want to recognize my advisor Prof. Steven W. Cranford, whom I owe my deepest gratitude. He is solely responsible for my introduction to the world of atoms and nano-scale subjects, I have learned a great deal beyond my background, furthermore I am forever in debt for his influence on my development as a scientist and writer. I want to thank Steve for his passion, patience and understanding, I know it is definitely not an easy task to advise me. There are not enough words to express my gratitude, and I hope that at least the intentions are clear.

Molecular Cooperativity/Compatibility

K. Kwan, 2015
Regarding academic guidance, I would like to thank my thesis committee of Professor Sandra Shefelbine, Professor Moneesh Upmanyu, and Professor Andrew Myers for the fruitful suggestions and valuable time.

This research was funded by Steven Cranford and NEU CEE Department as well as the Fulbright Student Program.
For Lissa, and Jeremy, thank you for being here with me on this ride.
# TABLE OF CONTENTS

LIST OF TABLES .............................................................................................................. xi
LIST OF FIGURES ............................................................................................................. xii
ABBREVIATIONS .............................................................................................................. xix

1. INTRODUCTION ........................................................................................................... 1
   1.1 Background and Problem Statement ................................................................. 2
   1.1.1 The Challenge of Complex Materials ......................................................... 4
   1.1.2 Cooperativity at the Nanoscale ................................................................. 10
   1.2 Literature Review ......................................................................................... 15
   1.3 Objectives ....................................................................................................... 17

2. METHODOLOGY ......................................................................................................... 20
   2.1 Molecular Dynamics .................................................................................... 20
   2.2 Basic Statistical Mechanics – The Ergodic Hypothesis ......................... 24
      2.2.1 Thermodynamic Ensembles ................................................................. 26
   2.3 Force Fields ................................................................................................... 28
      2.3.1 Pair Potentials .................................................................................... 29
      2.3.2 Force fields for Biological Materials ................................................. 30
      2.3.3 Bond Order and Reactive Potentials .................................................. 32
   2.4 Atomistic Water, Explicit Solvent, and TIP3P Water Model ............ 34
   2.5 Long range interactions ............................................................................. 36
   2.6 Multi-scale Modeling (Coarse Graining) .................................................. 38
      2.6.1 Elastic Network Models ................................................................. 40
      2.6.2 Two Potential Freely Jointed Chain Polymer Models .................... 42
      2.6.3 The MARTINI Force Field ............................................................. 43

3. COOPERATIVITY METRICS ................................................................................... 46
   3.1 Polymer Physics: Ideal Chain Configurations ....................................... 47
      3.1.1 End-to-end Distance and Polymer Size ........................................... 48
      3.1.2 Radius of gyration ............................................................................ 49
   3.2 The Gyration Tensor ................................................................................... 51
      3.2.1 The Eigenvalues of the Gyration Tensor .............................................. 54
   3.3 Size Metrics ................................................................................................. 59

Molecular Cooperativity/Compatibility

K. Kwan, 2015
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4</td>
<td>Shape Metric</td>
<td>60</td>
</tr>
<tr>
<td>3.5</td>
<td>Orientation Metric</td>
<td>61</td>
</tr>
<tr>
<td>3.6</td>
<td>Summary</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>Molecular Model Formulation</td>
<td>65</td>
</tr>
<tr>
<td>4.2</td>
<td>Simulation Protocol</td>
<td>67</td>
</tr>
<tr>
<td>4.3</td>
<td>Results and Discussion</td>
<td>70</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Slip and Norm</td>
<td>70</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Shape Metrics</td>
<td>73</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Orientation Metrics</td>
<td>78</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Closer look at cross-link density</td>
<td>80</td>
</tr>
<tr>
<td>4.4</td>
<td>Summary</td>
<td>81</td>
</tr>
<tr>
<td>5.1</td>
<td>Tropocollagen Molecule</td>
<td>83</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Single Point Mutations: O1 Mutations</td>
<td>84</td>
</tr>
<tr>
<td>5.2</td>
<td>Molecular Model Formulation</td>
<td>86</td>
</tr>
<tr>
<td>5.3</td>
<td>Simulation Protocol</td>
<td>88</td>
</tr>
<tr>
<td>5.4</td>
<td>Results and Discussion</td>
<td>89</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Size Metric – Slip</td>
<td>90</td>
</tr>
<tr>
<td>5.4.2</td>
<td>Shape Metric – Anisotropy</td>
<td>95</td>
</tr>
<tr>
<td>5.4.3</td>
<td>Orientation Metric – Skew</td>
<td>96</td>
</tr>
<tr>
<td>5.5</td>
<td>Summary</td>
<td>99</td>
</tr>
<tr>
<td>5.5</td>
<td>Summary</td>
<td>101</td>
</tr>
<tr>
<td>6.1</td>
<td>Mechanical Properties Characterization of CNT-Collagen</td>
<td>103</td>
</tr>
<tr>
<td>6.1.1</td>
<td>Tensile Tests</td>
<td>106</td>
</tr>
<tr>
<td>6.1.2</td>
<td>Bending Test</td>
<td>107</td>
</tr>
<tr>
<td>6.2</td>
<td>Steered Molecular Dynamics simulations for CNT-Collagen Interactions</td>
<td>108</td>
</tr>
<tr>
<td>6.3</td>
<td>Summary</td>
<td>108</td>
</tr>
<tr>
<td>7.1</td>
<td>CONCLUSION AND EXPECTED SIGNIFICANCE</td>
<td>112</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>118</td>
</tr>
<tr>
<td>Appendix A</td>
<td>THE EIGENVALUE PROBLEM</td>
<td>139</td>
</tr>
<tr>
<td>Appendix B</td>
<td>THE COLLAGEN STRUCTURE MODEL</td>
<td>142</td>
</tr>
</tbody>
</table>
B.1 Creating The .PDB File ........................................................................................................ 142
B.2 Generating the PSF File .................................................................................................... 146
B.3 Inserting Mutations .......................................................................................................... 147
B.4 Converting to LAMMPS Format - Solvating .................................................................. 148

Appendix C  FASTA File Alpha1.fasta ............................................................................. 149
Appendix D  Topology File Hydroxyproline ................................................................. 150
Appendix E  LAMMPS Input file for Ideal Chains ...................................................... 151
Appendix F  MATLAB Gyrations Ideal Chains ............................................................ 153
Appendix G  MATLAB Gyrations Input .......................................................................... 155
Appendix H  MATLAB Anisotropy and Orientation Metrics ................................. 157
# LIST OF TABLES

Table 1. Brief examination of different thermodynamical ensembles ............................................. 27
Table 2. The parameters for the modified TIP3P model for CHARMM is shown ................................. 36
Table 3. The parameters for the modified TIP3P+long range interactions for CHARMM ............... 37
Table 4. Radius of gyration for common polymer architectures ......................................................... 50
Table 5.OI produced by mutations of Type I Collagen ....................................................................... 87
Table 6. Frequency of Lethality of OI by Substitution. Adapted from Bodian et al.[132] .................. 89
Figure 1: Carbon allotropes (a) Diamond (b) graphite. Notice how the carbon atom structure is arranged. The resulting material diamond, from invincible in Greek, is the hardest natural material with a hardness of 10 in the Mohs scale, whereas graphite’s, softness (due to the non-bonded layered structure) is what gives its use as a writing tool, which not coincidentally is the meaning of graphite in Greek, to write. Used with permission © Wikipedia. ............................................................... 5

Figure 2: Analogy of structural engineering and complex material design. From a structural engineering perspective one can predict the behavior of the structure across all scales. This is what one hopes to achieve in complex material design. Adapted from Cranford [9], used with permission © 2012 Massachusetts Institute of Technology. .................................................................................................................... 6

Figure 3: Length vs time scale. The dilemma is how to balance these two variables to observe the behavior of interest. Clearly the relevant functional scales are orders of magnitude away from solutions presented by current molecular simulations. Adapted from Buehler & Yung [11], used with permission © 2009 Nature Publishing Group. .............................................................................................................................. 7

Figure 4: Building block problem. The assembly of proteins can be decomposed into hierarchical levels ranging from simple compounds, to amino acids, to alpha and beta sheets and finally the protein. Nature’s hierarchical order gives us an example of how hierarchical systems can be exploited to enhance material properties at the relevant scales. Figure adapted from Cranford & Buehler[13], used with permission © 2012, Springer Science+Business Media Dordrecht.............................................. 9

Figure 5: Cooperativity Analogy. On the left a macroscale composite beam (a) consider a simple composite cantilever beam system, 2 materials A and B (b) subject to load the components may not deform together, thus failing compatibility and cannot be considered as single element (c) if compatibility conditions are met they deform together (d) thus they can be considered a single unit. On the right compatibility at the molecular level (a) consider system of molecule A and B (b) if cross-linking is insufficient A and B do not deform together and are uncooperative (c) if cross-linking is sufficient the components deform together and can be considered cooperative (d) like the composite beam if cooperative one can consider it a single unit, with effective properties C, thus A and B are no longer necessary. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE............................... 12

Figure 6: Atomistic examples that invalidate compatibility laws and illustrate the need for cooperativity metrics at the nanoscale. (a) Graphdiyne system is ruptured under tension, creating a void in the structure. (b) Polymer complex system showing the crossing over of one chain to another. Both behaviors are incompatible under the continuum interpretation but might not necessarily be so atomistically. ................. 14

Figure 7: Full atomistic MD simulation of a pair of solvated polyelectrolytes complex (PAH and PAA). (a) Coupled polyelectrolytes under a pH of 7.0 have a high density of ionic cross linking that translates to cooperative mechanical behavior. (b) Coupled polyelectrolyte with a pH of 4.0 which results in a lack of cross-link thus creating
openings. Adapted from Cranford[30], used with permission © 2012, Royal Society of Chemistry.................................................................17

Figure 8: Molecular cooperativity summary. A) Theoretical formulation where gyration based metrics are developed. B) Testing the cooperativity metrics on an idealized macromolecule study and given positive results to ultimately carry out C) validation on an existing material system, the collagen molecule...................18

Figure 9: Molecular Dynamics Interactions. (a) The formal atomistic structure is replaced by a point mass representation with position ri(t), velocity vi(t), and acceleration ai(t). (b) Potential energy well, it is a basic energy decomposition based on geometrical constraints of a simple pair potential function. The depth of the well ties directly into the physical properties of the materials. Adapted from Cranford [9], used with permission © 2012 Massachusetts Institute of Technology..........................22

Figure 10: Depiction of the contribution of different atomic behavior to the potential energy defined in the CHARMM potential. Adapted from Buehler [51]...........................32

Figure 11: Composite graphene structured modeled using the AIREBO potential. In the system above the red spheres represent pseudo-nanoparticles, and the objective was to observe the saturation of the system by looking at the interaction energy, thus showing insight towards the structure's composite action. Adapted from Kwan and Cranford [43], used with permission © 2014 Elsevier B.V.…………………33

Figure 12: Unfolding of disulfide stitched graphene. The ReaxFF potential was used for this full atomistic study that shows the critical number of bonds that is necessary such that the bending energy of the capsule itself overcomes the total disulfide bond energy, which leads to the unraveling of the graphene sheet. Adapted from Kwan and Cranford [57], used with permission © 2014 AIP Publishing LLC…………………34

Figure 13: A 3-site rigid water molecule. For the case of the modified TIP3P molecule, each atom of the molecule is an interaction site as well as the usage of the Lennard-Jones potential for interatomic interactions. The TIP3P water model is employed in the CHARMM force field and parameters are shown in Table 2…………………35

Figure 14: Structural analysis of frames. The transition from (a) truss elements to (b) beam elements to (c) structural frame analysis, made possible by obtaining the constitutive models (material properties) and the structural organization. Adapted from Cranford[62], used with author’s permission…………………………..39

Figure 15: ENM for polyelectrolyte complex. (a) Full atomistic representation of coupled PAA and PAH chains. (b) General elastic network connecting the atoms with harmonic springs, in this case the model is unaware of the cross-linking scheme specific for this structure which in turn explains the absurd amount of cross-links. Adapted from Cranford [9], used with permission © 2012 Massachusetts Institute of Technology……………………………………..41

Figure 16: The coarse-grained representation of all the naturally occurring amino acids for the MARTINI force field. Reproduced with permission from Monticelli [79], used with permission © 2008, American Chemical Society…………………………..44

Figure 17: Simple polyethylene chain. A sample polymer, notice the one dimensional configuration characteristic of polymers in general, the backbone is composed of carbon atoms (grey) and hydrogen atoms (white)……………………………………..48

Figure 18: Ideal Chain. An ideal chain representation of a polymer with n monomers and l bond lengths. R represents the end-end distance, where the coiled nature of the polymer is not captured in this measurement.………………………………………………………..49

Molecular Cooperativity/Compatibility

K. Kwan, 2015
Figure 19: End-end distance vs. radius of gyration. The interpretation of the radius of gyration \( R_g \) can be thought of as a sphere, taking into account the coiled behavior that is unnoticed if one only looks at the end-end distance \( R \). 51

Figure 20: Gyration schematic for single molecule of any random configuration. A gyration tensor can be determined that describes the second moments of position of a collection of particles composing a macromolecule through geometry alone. Here the difference between particles \( i \) and \( j \) in x-direction are indicated. The resulting gyration tensor can be diagonalized, principal axes of gyrations can be defined which intersect at the center of mass for the molecule. The magnitude of the axes reflect the eigenvalues of the gyration tensor which can be used to define several parameters that can be exploited to define molecular cooperativity. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE. 52

Figure 21: A schematic of a collection of particles, in 1-D the relationship between the Moment of Inertia Eq.(39) and Radius of Gyration Eq.(40) is easier to visualize. 53

Figure 22: Asphericity of polymer. (a) the blue bead polymer shown has a linear configuration, meaning very possible that two of the three eigenvalues had short extension, the asphericity is close to 1. (b) the red bead polymer show a conformation that is most closely projected to a sphere thus its asphericity must be close to zero. 57

Figure 23: The deviation of gyration, for structures that have identical geometrical characteristics and are cooperating it is evident that the deviation is zero. In the case of \( S_1 \) and \( S_2 \) the uncooperativity is represented as a misalignment, thus the deviation is greater than the one for \( S_3 \) and \( S_4 \). 58

Figure 24: Skew, the orientation metric that illustrates the gyration alignment between molecule A(red) and molecule B (blue). 62

Figure 25: Cooperativity triangle. The derivation of this trifecta that describes size, shape and orientation is used as a definition of nano-cooperativity. The combined effort of the metrics are necessary to assess cooperative behavior in the structure. 64

Figure 26: Schematic of Coarse-grain model (a) Close-up of bead-spring model for molecular chains A and B with simple harmonic stretching (r), bending (\( \theta \)) and intermolecular repulsive terms (L-J), described by Eq.(55)-(56) and Eq.(21). Also shown are the discrete cross-links between molecules. (b) 2000 bead-spring system (MC1) with 251 cross links. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE. 67

Figure 27: Example of time evolution. Shows the dynamic evolution of the system as the simulation runs, notice that the configuration changes from a rather spherical one where the atoms are packed to a configuration that favors one direction, Section 3.4. 68

Figure 28: Slip results: (a) Magnitude of slip v. time for MC1. The computed size difference diminishes as the number of cross-links increase, one can note that past 21 cross-links the size difference is approximately zero, (b) Magnitude of slip v. time for MC2. Just like the MC1 counterpart the molecules size are relatively zero once they get past 42 cross-link scenario. (c) Euclidean norm of slip MC1 (blue) and MC2 (red) with respect to cross-links and unlinked length (inset). Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE. 72

Figure 29: Norm of Gyration Tensor results. (a) Gyration norm MC1. The size difference between each cross-link case reduces from a high of 1.2 to approximately zero, the variation between the first cases (3-10 cross-links) is evident. It is shown that past the 21 cross-link cases the gyration norm approaches zero. (b) Gyration norm MC2. The trend from MC1 is repeated here to a lesser extent, the range of 10-20 cross-
link shows more stability but one can definitely point that past 42 cross-link the size difference is approximately zero, thus suggesting cooperativity. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

Figure 30: Snapshot of molecular chains at 4 ns. a) The 3-cross-link scenario clearly shows the inharmonious motions between the chains, the independence between the chain motions affect the shape metric as well as the size and orientation metrics. b) The 100-cross-link case shows a fully integrated motion, where each pair of beads between the chains are displacing in the same direction, thus the particle configuration is similar between the chains.

Figure 31: Example of relative shape anisotropy evolution for MC1. Here, $\kappa A_2(t)$ and $\kappa B_2(t)$ for MC1 are plotted for both 3-cross-link (red) and 100-cross-link (blue) cases, to exemplify the differences. The wider the “gap” or separation between the two relative shape anisotropies indicates their geometric configuration differ at the same instances. The 100-cross-link case is interpreted as cooperative because its beads are distributing similarly on both chains. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

Figure 32: Shape and differential anisotropy results. (a) Relative shape anisotropy for MC1 as a function of time and number of cross-links. The gap for the 3-cross-link case is around 0.2 and it decreases steadily until the 21 cross-link case where the gap approaches zero. (b) Relative shape anisotropy MC2 as a function of time and number of cross-links. (c) Molecular shape for both MC1 (blue) and MC2 (red), as a function of number of cross-links and unlinked length. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

Figure 33: Orientation results. (a) Skew for MC1. The maximum angle of misorientation is 30 degrees, and similar to results from the previous metrics the chains gradually align themselves as the number of cross-links increases, reaching the same critical point at 21 cross-links. (b) Skew for MC2. The trend is similar to the one obtained from MC1, but like previous results the orientation for MC2 has less variability. (c) Norm of skewness, the molecular orientation, MC1 (blue) and MC2 (red) in terms of number of cross-links and unlinked length. The results confirm, like the previous metrics, that if this particular model has at least 21 cross-links then the misalignment of both chains approaches zero. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

Figure 34: Examples of non-cooperative single metrics. These examples stress the importance of evaluating cooperativity by using all three metrics. (a) A pair of identical carbon nanotubes, each carbon nanotube has the same configuration and radius of gyration, thus if one only looks at the slip and the differential anisotropy it would show cooperativity, however they are clearly oriented in the wrong direction. Thus the skewness metric would show incongruent behavior. (b) A system where a pair of axially deformable bars undergo opposite deformations, for this case the skewness metric would clearly show no misorientation, however since the loading conditions are opposite the deformation of both bars will clearly show a difference in the radius of gyration, thus making this another example illustrating the lack of validity of each individual metric as a single criteria for cooperativity.

Figure 35: Hierarchical depiction of collagenous tissues. Collagenous tissues are hierarchical structures formed through self-assembly, as is standard in Nature, one can consider this as a smart material since these tissues structure adapt to the mechanical forces. Adapted from Gautieri et. Al. [116], used with permission © 2011, American Chemical Society.

Figure 36: Tropocollagen molecule. (a) Depiction of the three chains and characteristic triple helical structure. (b) stagger of the three strands highlighted by the alpha carbons
of a sequence of GLY-PRO-HYP (c) the orthographic view from the top of the triple helix (d) close up of the triple helix, each chain is plotted with one color and the spheres represent the main alpha carbons of the glycine amino acid, the helical structure is stabilized by hydrogen bonds. The molecule is surrounded by explicit water but not shown to improve visual clarity.

Figure 37: Effects of missense mutation, shows that the replacement of Glycine by other residues affect the mechanical properties. (a) Stiffness properties can reduce up to 15%. (b) Intermolecular adhesion is further reduced. (c) The fibril maximum stress lost is over 100%, and fracture occurs at a lower strain. Adapted from Gautieri et. al. [124].

Figure 38: Mutation Site. (a) Tropocollagen molecule, the orange spheres indicate the location where residue substitution takes place. (b) Glycine amino acid present in the healthy collagen substituted by Valine making it a mutated collagen. Notice the difference in size between the amino acids, which is one of the reasons for the presence of a kink. All chains are identically sequenced GLY-PRO-HYP triplets.

Figure 39: RMSD of the simulated collagen molecules. The RMSD of distances between atoms is used as a guide to establish whether systems have reached dynamical equilibrium. In red the RMSD of the healthy strand with a mean of 5.32 and a variance of 1.65, lower in comparison to the mutated collagen, in blue, that has a mean of 7.75 and a variance of 2.47, the higher average distance and larger fluctuations are expected as a result of the swelling of the molecule due to the presence of a mutation.

Figure 40: Visual confirmation of the kink in mutated collagen. (a) The healthy collagen molecule is shown, notice the tight packing of the helix at all times. (b) the mutated collagen shows a clear disruption in the triple helical structure, and although subtle one can perceive the looseness at all locations, demonstrating that the model is capturing the intended behavior.

Figure 41: Pre-established conditions of collagen. Because this is a natural existing material system, the metrics are used here to differentiate healthy and mutated collagen based on the previous knowledge that points to the existence of an unfolding as the result of glycine being replaced.

Figure 42: Spliced cooperativity approach. An \( \approx 8 \text{nm} \) long, 30 amino acids per chain, mutated collagen is presented, in this case the collagen is spliced into three uniform sections. The chains are identically sequenced as \((\text{GPO})^5\text{-VPO-}(\text{GPO})^4\), where V corresponds to the substitution of Glycine, orange beads with Valine, purple beads, which corresponds to the second section. The molecule is shown surrounded by TIP3P water molecules that are represented as a transparent blue surface.

Figure 43: Per chain analysis – Radius of Gyration. (a) Radius of gyration of each individual chain located at the section of mutation for the healthy collagen, the second chain presents a minor lack of cooperativity with the rest of the chains. (b) Corresponding section of the mutated collagen, the metrics reflect a relative harmonious behavior at all chains, a conflicting display of what the mutation is physically producing in the molecular behavior.

Figure 44: Section analysis of the slip in the collagen molecule. The slip metric is taken as the relative size difference between the healthy and mutated collagen for each section. The blue plot corresponds to section 2, which contains the mutation, has a relative size difference that varies with a mean of \( \approx 6.97 \), almost a 200% difference with respect to the other relative size differences. The gap in relative size presented is an indication of the swelling of the molecules due to the disruption caused by a single point mutation on the helical behavior of the molecule.
Figure 45: Shape configuration between sections. (a) The shape difference presented for section 2, where the mutation lies, the difference between shape configurations in time shows a drop-off of roughly 15%, indicating a clear gap in the way the atoms are configured within the mutated section. (b) The shape difference in section 3, where the difference between the mutated and heathy collagen section 3 are unnoticeable, the mutated collagen shadows the healthy collagen section, both have a mean of 0.47, and is for all intents and purposes a mirror image of the shape configurations for section 1.

Figure 46: Section analysis of differential shape in collagen molecule. The differential anisotropy metric shows a clear-cut margin between the mutation containing section 2 and the other sections. Noting that the anisotropy is contained between 0 and 1, the margin between the sections is ≈12%, the maximum standard deviation also belongs to section 2 and amounts to roughly 0.03, which is within one order of magnitude smaller than the mean, the relative minor variability only reinforces the confidence of the results.

Figure 47: Section analysis of orientation in the collagen molecule. Skew results, due to the tight packing of the triple helical structure, seem to indicate that collagen has a natural misorientation of ≈1 degree. In this case the mutated section shows some minor fluctuations, the mean =0.995 a reduction of approximately 0.5%. All three metrics seem to indicate local uncooperative behavior in section 3.

Figure 48: Moving average of orientation metric. The results are filtered through a moving average, each point is the average of 100 data points, and is used to convey a cleaner image where the swing in orientation is shown to be chaotic for the section that contains the mutation.

Figure 49: Moving standard deviation of orientation metric. The standard deviation for each 100 points are measured and show above, one can observe the distinct gap between the “oscillation” of the variability for the blue plot, belonging to the mutation affected section. The variation here exceeds over 3 times the variation of the other sections where the mutation is not present.

Figure 50: Slip of collagen molecule. For this case the slip metric is used on the whole collagen (8nm) molecule. The results are encouraging showing a distinct difference, a direct implication is the affordability of using less and larger sections to identify mutations on a complete 300nm strand of collagen molecule.

Figure 51: Schematic of a single-wall carbon nanotube (SWCNT). (a) top view (b) trimetric view. Notice the characteristic honeycomb pattern. A carbon nanotube is essentially a rolled up graphene sheet.

Figure 52: CNT-Collagen composite sketch. (a) Disordered bundles of collagen and CNTs (b) using CNTs as templates the intention is recreate a D-banded collagen fiber.

Figure 53: Schematic of a single-wall CNT-collagen composite. (a) orthographic view (b) side view. For the collagen molecule the main alpha carbons are represented by beads.

Figure 54: Example of .FASTA extension file, presents 90 amino acids in the customary GLY-X-Y sequence found in collagen.

Figure 55: Main window of The BuScr Script to generate the collagen structure.

Figure 56: Chain file selection for collagen structure, THe BuScr.

Figure 57: N- and C- terminal selection for THe BuScr, for the collagen study these are usually added as dummy amino acids for the FASTA file as they are irrelevant.
N-terminal refers to the end amino acid of the N-sequence and C-terminal to the start of the C-sequence. The termini sequence are shown in blue.

Figure 58: Helical propensity and helix settings window, here some information concerning helicity and melting temperature is presented. Leave everything as is and click Done.

Figure 59: Generate triple-helical coordinates window, this windows allows for the selection of atoms, the coordinate system and other options. It is the last step for The BuScr to generate the PDB file.

Figure 60: Auto PSF generator module in VMD, to operate just add the topology file, and manually assign the chain.

Figure 61: Mutator module in VMD, to use just browse the original PSF and PDB, select the residue to be replaced and input the three letter code of the replacement amino acid.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D</td>
<td>1 Dimension</td>
</tr>
<tr>
<td>Å</td>
<td>Angstrom</td>
</tr>
<tr>
<td>AFM</td>
<td>Atomic Force Microscope</td>
</tr>
<tr>
<td>AIREBO</td>
<td>Adaptive Intermolecular Reactive Empirical Bond Order</td>
</tr>
<tr>
<td>AMBER</td>
<td>Assisted Model Building with Energy Refinement</td>
</tr>
<tr>
<td>atm</td>
<td>atmosphere</td>
</tr>
<tr>
<td>BO</td>
<td>Bond Order</td>
</tr>
<tr>
<td>CG</td>
<td>Coarse-Grain</td>
</tr>
<tr>
<td>CHARMM</td>
<td>Chemistry at HARvard Macromolecular Mechanics</td>
</tr>
<tr>
<td>CNT</td>
<td>Carbon Nanotubes</td>
</tr>
<tr>
<td>DNA</td>
<td>DeoxyriboNucleic Acid</td>
</tr>
<tr>
<td>DOF</td>
<td>Degrees Of Freedom</td>
</tr>
<tr>
<td>ENM</td>
<td>Elastic Network Model</td>
</tr>
<tr>
<td>FJC</td>
<td>Freely Jointed Chain</td>
</tr>
<tr>
<td>GLY</td>
<td>GLYcine</td>
</tr>
<tr>
<td>GLY-PRO-HYP</td>
<td>GLYcine-PROline-HydroxyProline</td>
</tr>
<tr>
<td>GROMOS</td>
<td>GROningen MOlecular Simulation</td>
</tr>
<tr>
<td>H</td>
<td>Hamiltonian</td>
</tr>
<tr>
<td>JE</td>
<td>Jarzynski Equality</td>
</tr>
<tr>
<td>K</td>
<td>Kelvin</td>
</tr>
<tr>
<td>Kcal</td>
<td>Kilocalories</td>
</tr>
<tr>
<td>KE</td>
<td>Kinetic Energy</td>
</tr>
<tr>
<td>LAMMPS</td>
<td>Large-scale Atomistic/Molecular Massively Parallel Simulator</td>
</tr>
<tr>
<td>L-J</td>
<td>Lennard-Jones</td>
</tr>
<tr>
<td>MC1</td>
<td>Molecular Complex 1</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>NPT</td>
<td>Number of particles, constant Pressure, Temperature</td>
</tr>
<tr>
<td>ns</td>
<td>nanoseconds</td>
</tr>
<tr>
<td>nm</td>
<td>nanometer</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear Magnetic Resonance</td>
</tr>
<tr>
<td>NVE</td>
<td>Number of particles, system Volume and total Energy</td>
</tr>
<tr>
<td>NVT</td>
<td>Number of particles, system Volume, Temperature</td>
</tr>
<tr>
<td>OI</td>
<td>Osteogenesis Imperfecta</td>
</tr>
<tr>
<td>OPLS</td>
<td>Optimized Potentials for Liquid Simulatio</td>
</tr>
<tr>
<td>PAA</td>
<td>Poly(Acrylic Acid)</td>
</tr>
<tr>
<td>PAH</td>
<td>Poly(Allylamine Hydrochloride)</td>
</tr>
<tr>
<td>PPPM</td>
<td>Particle-Particle and Particle-Mesh</td>
</tr>
<tr>
<td>REAXFF</td>
<td>REActive Force Field</td>
</tr>
<tr>
<td>RMSD</td>
<td>Root Mean Square Deviation</td>
</tr>
<tr>
<td>SMD</td>
<td>Steered Molecular Dynamic</td>
</tr>
<tr>
<td>SPC</td>
<td>Simple Point Charge</td>
</tr>
<tr>
<td>SWCNT</td>
<td>Single-Walled Carbon Nanotube</td>
</tr>
<tr>
<td>THz</td>
<td>TeraHertz</td>
</tr>
<tr>
<td>T</td>
<td>Temperature</td>
</tr>
<tr>
<td>TIP3P</td>
<td>Transferable Intermolecular Potential 3 Point</td>
</tr>
<tr>
<td>THeBuScr</td>
<td>Triple Helical collagen Building Script</td>
</tr>
<tr>
<td>Tr</td>
<td>Trace</td>
</tr>
<tr>
<td>VAL</td>
<td>VALine</td>
</tr>
<tr>
<td>vdW</td>
<td>van der Waals</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Traditionally Civil Engineers have been trained to look at problems from a “big picture” perspective, encompassing large-scale structures such as bridges and dams, urban infrastructure, and the built environment. At such scales, engineers have pushed materials to the extreme limits of performance. Concurrently, the advancements, challenges and potential benefits of nanotechnology are pushing for a refocusing of expertise. The difficulties presented in multiple fields (medical, environmental and energy) cannot be simply dismissed by evolutionary approaches and incremental gains, but by overcoming them through transformative methods. Many of these approaches rely on exploiting the limits of material behavior – including strength, toughness and resilience - and there is an increasing trend in focusing on the fundamental behavior of materials, to redefine the process of materials selection for particular applications.

While the importance of macro-scale structural characterization cannot be undermined, the inclusion of material behavior on the modeling of structural response is typically idealized/homogenized (e.g., even high-fidelity finite elements presume the continuum hypothesis holds), or incorporates some empirical law to reflect small-scale effects (e.g., so-called micromechanical models). Such approaches can reflect the effect of materials, but do not apply to the systematic design of materials. The proposed research will attempt to develop bottom-up molecular-scale understanding of material behavior, with the long term objective being the application of this understanding to design functional material systems with applications at the ultimate structural or functional scale. As such, a “nano to macro” synergistic platform and subsequent applications to structural systems will provide the necessary tools and procedures to validate future structural models and other critical behavior across-scales in a holistic manner.
1.1 Background and Problem Statement

Current innovations in structural design are commonly based on proven material systems such as steel and concrete\(^1\). As a consequence, these structures are reaching the limits in terms of the design parameters and performance. The need for more robust and resilient infrastructure for protection against seismic risks, flooding, hurricane and other hazards are motivating to think beyond the traditional design methods. This requires another perspective where the material itself is considered a structure, albeit at a lower scale. From a mechanics viewpoint the materials can be viewed as simply smaller structures, from which their assemblies can dictate design capacities and failure modes. This is precisely where the expertise of structural engineering can be integrated into materials science.

Taking an atomistic approach to study materials systems opens the door to significantly improve material properties that concern to a resulting system strength, reliability, and sustainability. One method to explore systems at the atomistic or nanoscale is through computational modeling or simulation efforts (so-called \textit{in silico} approaches). It is believed that computational methods can be exploited to better understand the use, selection, development, and discovery of materials, with a goal to achieve high-speed and robust acquisition, management, analysis, and dissemination of diverse materials data. This is the basis of the recently proposed \textbf{Materials Genome Initiative [1]} and \textbf{Materials Project [2, 3]}, removing guesswork from materials design and “accelerating materials discovery through advanced scientific computing and innovative design tools.” The ultimate goal is to use computational models and data mining to achieve deeper understanding by applying lessons learned from data gathered on one type of materials.

\(^1\) A testament to both economic factors and traditional design comfort, rather than any material benefit.

\textit{Molecular Cooperativity/Compatibility} \hspace{1cm} K. Kwan, 2015
material to others. A National Academies report [4] describes the need for using multiscale materials modeling to capture the process-structures-properties-performance of a material. Here, such concepts are applied to the ultimate, macro-scale. The field is highly multi-disciplinary, requiring students with a range of backgrounds making use of cutting edge techniques from materials science, engineering, chemistry, physics, computer science, and mathematics. The multidisciplinary research integrates joint computational, theoretical, and experimental approaches, across scales and encompassing a range of materials.

The basis of such computational methods are full atomistic models – representing each atom of a material as the specific element (e.g., C, H, O, N, etc.) with associated chemical details (e.g., bonding, ionic interactions, etc.) Such models are typically parameterized through robust quantum mechanical approaches, and are thus rooted in fundamental physics rather than any presumed constitutive laws. Indeed, such laws are not necessary \textit{a priori} to define the system – the models are free to evolve and respond according to the simulation algorithms. Moreover, material systems can be varied without the need for presumed structural changes or assumptions of stability. Modeling systems atom by atom also enables precise control of boundary conditions, development of complex geometries, and extraction of system information inaccessible by experimental methods. One main advantage of a full atomistic approach is that material systems can be varied without the need for prescribed synthesis pathways – e.g., chemically stable systems can be characterized and quantified, even if they are not yet attainable experimentally. As a direct result, the material “solution space” can be greatly extended, potentially exploiting novel material combinations, providing an efficient means to screen, assess, and explore potential complex material systems. The computational effort can be thought of as a “virtual experiment”
or “computational microscope”, where one is able to explore the system without the burden of a physical experiment, which in many occasions one is limited to the resolution of the details.

One important but grossly overlooked aspect of the study of any material system is the cooperativity of its components. In simplest terms, we use the term cooperativity to infer two (or more) distinct material components behaving as a cohesive structural unit. Such cooperativity is necessary for the bottom-up assembly of complex materials systems, relying in hierarchical architectures of components upon components upon components - and so on down the rabbit hole - to the fundamental basis atoms. Materials can be improved by knowing how to manipulate the interaction between the parts, and significant advances can be expected by exploiting the intersection and integration between material science and structural engineering. While the objective of the research proposed herein can be broadly defined as molecular compatibility, the encompassing goal is to structure the physics, chemistry, mechanics and materials science community to develop a synergistic approach for the prediction of performance across temporal and spatial scales.

1.1.1 The Challenge of Complex Materials

It is easy to recognize that the advent of nanotechnology has been responsible for the recent advance in technology, (e.g. nanocomposites are being used in the automobile industry to reduce the weight of the chassis, organic light-emitting diodes (OLED) which are manufactured from nanostructures, have brought new possibilities to the display market, among thousands of other current technology.[5]) The use of a bottom-up approach, starting at the atomistic scale, has had an impact through new materials and design strategies. If this is so, why has there been little innovation in the availability of material considered for civil infrastructure? Throughout history,
construction materials have been based on wood, steel and concrete. With little modification these materials have been and are still the main choices in day to day building and design. Moreover, the nanoscale atomistic structures of these systems are only now being fundamentally explored [6-8], uncovering previously unknown atomistic details and mechanisms. We do not fully understand our most common building materials! Clearly, a new approach is warranted.

To overcome the complexities that have been affecting the industry now and tomorrow it is necessary to depart from the comforts of traditional methods and question how to involve the use of nanotechnology in our design approaches. Due to the ultimate limitations of combinable elements (in the periodic table), it is the current thought that structuring atoms is the key to new materials design. This is clearly evident in Nature. One of the classical examples is how the difference in atom structure of carbon atoms (i.e., carbon allotropes), see Figure 1, reflect substantial variation in material properties at the macroscale, e.g., a tetrahedral formation of carbon atoms results in diamond, the hardest known material (a 10 in Mohs hardness scale), whereas a layer of honeycomb arranged carbon atoms creates graphite one of the softest materials (2 in Mohs scale).

![Figure 1: Carbon allotropes (a) Diamond (b) graphite. Notice how the carbon atom structure is arranged. The resulting material diamond, from invincible in Greek, is the hardest natural material with a hardness of 10 in the Mohs scale, whereas graphite’s, softness (due to the non-bonded layered structure) is what gives its use as a writing tool, which not coincidentally is the meaning of graphite in Greek, to write. Used with permission © Wikipedia.](image)

*Molecular Cooperativity/Compatibility*  
K. Kwan, 2015
Just as in traditional structural engineering where the approach is to determine the arrangement of structural elements (e.g., trusses, beams, columns) to maximize the capacity of the global system (e.g., The Eiffel Tower), the current global challenge of material design is to come up with key atomic structures at the nanoscale that enhance the macro properties of the complex material, see Figure 2. Simply put, what is the equivalent “truss” at the nanoscale and can it be designed? This leads to the philosophy of using materials not by a top-down approach, where material is wasted, but to structurally build up from an atomistic level. The aforementioned is known as the bottom-up approach.

Figure 2: Analogy of structural engineering and complex material design. From a structural engineering perspective one can predict the behavior of the structure across all scales. This is what one hopes to achieve in complex material design. Adapted from Cranford [9], used with permission © 2012 Massachusetts Institute of Technology.

One of the crucial challenges in complex materials design arises from the need to obtain a complete theoretical foundation of the system beyond simple properties (e.g., not just “strength” but load paths through the system, not just “adhesion” but the chemical interactions that constitute it). Full understanding (and potential control) of such pathways could facilitate the
understanding (and potential design) of emergent properties at the largest scales [10]. The concept is linked to that of materiomics – a holistic knowledge of the material system across scale – where, simply put, “the whole is greater than the sum of the parts”. This implies knowledge of the system behavior across a wide range of time and spatial scales. Molecular simulations are generally taking place in the minimum time scales, of the order of $10^{-15}$ seconds (femtoseconds), whereas one can imagine, as the system grows the macro effects are taking place in the order of milliseconds, seconds, hours, or even years (such as the case for the mechanisms of concrete creep, for example). This is commonly known as the time versus length dilemma, depicted in Figure 3.

Figure 3: Length vs time scale. The dilemma is how to balance these two variables to observe the behavior of interest. Clearly the relevant functional scales are orders of
magnitude away from solutions presented by current molecular simulations. Adapted from Buehler & Yung [11], used with permission © 2009 Nature Publishing Group.

It demonstrates the colossal gap between the accessible time and space scales of the solutions of molecular simulations and the ultimate functional time scales. In terms of length scale, to accurately reflect full atomistic systems, bonds are typically on the order of a few Ångstrom (0.1 nanometers, or $10^{-10}$ meters), and associated vibrational displacements mere fractions of Ångstroms. In terms of time scales, the integration timestep must be small enough to capture the vibrational frequencies of the atoms. Hydrogen, for example, vibrations on the order of 100 THz ($10^{14}$ Hz or cycles per second), requiring time steps on the order of femtoseconds. This creates an impasse from a computational perspective, to be able to model macro-systems as fully atomistic would require technological capabilities that are currently unfeasible (i.e., time scale > 1 ns, requiring millions of integration steps per nanosecond, and length scale > 20nm resulting in a system size > 1 million atoms with multiple degrees of freedom per atom). Simulating a microsecond of “real time” is currently pushing the boundaries of high-performance super computers (for example, one of the largest MD runs involved a system of over 4 trillion atoms - a little over 25 cubic micrometers – and required over 140,000 computational cores using relatively simple molecular formulations [12]). Ultimately this gap represents the frontier that must be overcome for successful complex material system design.

Nature shows one possible way in which the length-scale dilemma can be overcome – via hierarchies. By assembling components into common structures, and then assembling higher-order structures, the same materials can be used to construct larger systems (circumventing length limitations) with varying functions (using the same base materials). Indeed, protein-based biological materials are composed of hierarchical structures which provide exceptional structural, mechanical, chemical and optical properties across all scales, yet composed of a limited number
of base amino acids (only twenty). How such hierarchies can be established in the assembly of such proteins (from chemical compounds to amino acids, to beta sheets to full protein structures) is illustrated in **Figure 4**. Depending on the scale of interest, the “fundamental component” of a tissue may be considered the proteins, the secondary structures, the peptide sequence, the amino acids, or the atoms (or even the electron orbitals!). The ability to construct higher hierarchical structures plays a direct role in the time scale, since larger structures can be used to observe large scale effects that take place in longer time scales. In the case of biological molecules different processes can be observed at different time scales. Smaller units compose larger units, across scales like Legos built from Legos. As such, this is popularly known as the **building block problem**.

![Figure 4: Building block problem. The assembly of proteins can be decomposed into hierarchical levels ranging from simple compounds, to amino acids, to alpha and beta sheets and finally the protein. Nature’s hierarchical order gives us an example of how hierarchical systems can be exploited to enhance material properties at the relevant scales. Figure adapted from Cranford & Buehler[13], used with permission © 2012, Springer Science+Business Media Dordrecht.](image-url)
From a modeling perspective, one can establish spatial hierarchies by using multi-scale or coarse-grain models, in other words by condensing the degrees of freedom (DOFs), to bridge the gap between full atomistic representations and continuum theory (see Figure 3). Condensation is predicated on assuming that some form of cooperativity exists at the lower scale. The understanding of the synergistic multi-scale transition from atomic, to meso-scale to macro-scale of the system is undoubtedly critical for structure-relationship predictions. Such scale-bridging coarse-graining methods allow the designer to enter these inaccessible time scales that would provide fundamental insight to material design.

Assembly of function blocks can be likened to the piece-by-piece design and construction of a building or bridge. As such, the term structural engineer in today’s application is ambiguous and can no longer be categorized as a profession were the expectations are a design of a building or similar scale. A simple observation leads one to question the definition of “structure”, and from a materials design perspective the structure itself is a material system, whether synthetic or biological. Indeed, design of a reinforced concrete beam necessitates understanding of the joint concrete/steel performance, yet it is typically not considered within the scope of materials science. As will be shown further, while the details change due to the fact that scale of operation is different, the fundamental concepts behind structural analysis can be applied.

1.1.2 Cooperativity at the Nanoscale

Nanotechnology has advanced to the point where almost any molecular functional group can be introduced into a material system. Using functional units – hypothetical material “building blocks” – it is thought that complex material systems can be engineered from the bottom-up [14-17]. However, the ensuing properties attained via the combination of arbitrary components are not yet predictable, and thus cannot be logically engineered [18-21]. Many studies have
demonstrated the high strength of synthetic materials such as carbon nanotubes [22] graphene [23], or polymer systems [24], as well as biological materials such as collagen [25] and spider silk [26], as examples. Such systems have greatly extended our understanding of nano- and molecular mechanics. Yet, one critical aspect has been grossly overlooked, a concept introduced to every undergraduate student in mechanics – compatibility. In the simplest mechanical terms, compatibility is the condition that materials deform together; in effect, when materials cooperate. Undoubtedly, integration of multiple components within advanced functional materials requires intimate knowledge of mechanical cooperativity between material interfaces, optimizing contact, adhesion, and deformation as well as strength.

At the nanoscale, the most basic form of molecular cooperativity can be captured by the cross-linking between two molecules [27-29]. If two (or more) molecules are adequately coupled, they can be considered a single functional unit – e.g., a fundamental hierarchical system. From a broader perspective, one no longer needs to differentiate between two molecules [30]. This is exactly why emerging techniques such as DNA-origami [31-33] is successful – even though DNA is two individual macromolecular chains, the helical and cross-linked structure allows consideration of a single DNA “building block”. Even at the macroscale, a successful composite is defined by the ability of the different components to act together as a single entity, and thus considered a single structural “building block” [34].

In other words a compatible behavior is experienced throughout the structure. As an illustrative example, we can compare a simple composite structural beam to cross-linked molecules (see Figure 5). For both cases, while clearly composed of two distinct material
constituents, if the components are sufficiently coupled, meeting compatibility conditions and thus “cooperative”, they can be considered a single unit (with transformed beam properties, for example, for the case of the beam, and macromolecular persistence length, for example, for the case of the molecular chains). A competent structural engineer need only consider the effect properties of the building block, without consideration of “smaller-scale” details and structure. We can therefore presume that a sufficiently cooperative system: (a) deforms together and (b) can be described as a single (homogenized) component. This is similar to enforcing compatibility conditions to macroscale deformations. However the issue is that the compatibility laws presented in continuum mechanics do not hold at the nanoscale - the most flagrant reason is that nanoscale elements are naturally discrete elements.

Figure 5: Cooperativity Analogy. On the left a macroscale composite beam (a) consider a simple composite cantilever beam system, 2 materials A and B (b) subject to load the components may not deform together, thus failing compatibility and cannot be considered as single element (c) if compatibility conditions are met they deform together (d) thus they can be considered a single unit. On the right compatibility at the molecular level (a) consider system of molecule A and B (b) if cross-linking is insufficient A and B do not deform together and are uncooperative (c) if cross-linking is sufficient the components deform together and can be considered cooperative (d) like the composite beam if cooperative one can consider it a single unit, with effective properties C, thus A and B are no longer necessary. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.
At the macroscale (continuum mechanics) the compatibility equations are based on the strain tensor. The strain tensor is normally defined as:

$$\varepsilon = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{21} & \varepsilon_{31} \\ \varepsilon_{12} & \varepsilon_{22} & \varepsilon_{32} \\ \varepsilon_{13} & \varepsilon_{23} & \varepsilon_{33} \end{bmatrix}$$ (1)

Where the diagonal terms of the strain tensor are defined as

$$\varepsilon_{ii} = \frac{dU_i}{dx_i}$$ (2)

Where $U_i$ is the component of the displacement field in the direction $x_i$

And the off-diagonal terms as

$$\varepsilon_{ij} = \frac{1}{2} \left( \frac{dU_i}{dx_j} + \frac{dU_j}{dx_i} \right)$$ (3)

The basic compatibility conditions are then six expressions (the so-called Saint-Venant’s compatibility conditions), three of the form:

$$\frac{\partial^2 \varepsilon_{11}}{\partial x_1^2} + \frac{\partial^2 \varepsilon_{22}}{\partial x_2^2} = 2 \frac{\partial^2 \varepsilon_{12}}{\partial x_1 \partial x_2}$$ (4)

and three more of the form:

$$\frac{\partial^2 \varepsilon_{33}}{\partial x_3 \partial x_2} = \frac{\partial}{\partial x_3} \left[ \frac{\partial^2 \varepsilon_{32}}{\partial x_1} + \frac{\partial^2 \varepsilon_{31}}{\partial x_2} - \frac{\partial^2 \varepsilon_{12}}{\partial x_3} \right]$$ (5)

A compatible strain tensor field in a body is a unique field that is obtained when the body is subjected to a continuous, single-valued, displacement field.

Molecular Cooperativity/Compatibility  K. Kwan, 2015
In the continuum description of a solid body, the body is composed of a set of infinitesimal material points. Each point is assumed to be connected to its neighbors without any gaps or overlaps. One can better understand the compatibility laws (Eqs. (4) and (5)), by picturing a large body divided into small differential cubes, straining each cube and putting them together once again, if the strains are compatible then it would be possible to put the distorted cubes back into a single continuous solid. These conditions do not hold at the atomistic scale – gaps and overlaps are permissible. Figure 6 shows a) the breaking of carbon bonds that lead to rupture in the graphdiyne sheet [36], which infringes the compatibility laws due to separation, and b) the crossover of atoms of one polymer to another. So the question arises at the nanoscale: What are the conditions for compatibility and cooperativity?

Figure 6: Atomistic examples that invalidate compatibility laws and illustrate the need for cooperativity metrics at the nanoscale. (a) Graphdiyne system is ruptured under tension, creating a void in the structure. (b) Polymer complex system showing the crossing over of one chain to another. Both behaviors are incompatible under the continuum interpretation but might not necessarily be so atomistically.
With this motivation, the concept of mechanical coupling is herein defined as “equivalent mechanical behavior between components and complex”, where the objective is not to determine the resulting mechanistic properties (whether enhanced or diminished), but the harmony of the parts, or, colloquially speaking, when “1 + 1 = 1”. This is evident in the macro-scale where the behavior of composite materials such as concrete, are not determined by their individual components but rather their combined behavior [34]. In Nature this cooperativity phenomena is highly evident at the molecular level [11], such as the aforementioned cross-linked double-helix of DNA [37, 38], the triple-helix of tropocollagen [39, 40] and the unified deformation of the β-crystalline structure of spider silks [41]. Previously, the characterization of cooperativity has only been carried out for specific material systems from an energy perspective.

1.2 Literature Review

The subject of cooperativity in nanoscale studies has typically not been directly addressed, however some works [30, 42, 43] in particular have been carried out where some aspect of cooperativity has been explored. Among them, a study based on the cross-linking of coupled polyelectrolytes, served as a precursor for the general cooperativity metrics derived herein.

The aforementioned study by Cranford and Buehler[30] looked at the effect of molecular rigidity (e.g., bending stiffness) and cross-linking density on the cooperative behavior of a coupled macromolecule – the ionized complexation of poly(acrylic acid), or PAA, with poly(allylamine hydrochloride), or PAH. Hinged on the motivation of the building block problem, see Section 1.1.1, the authors wished to determine cooperativity as a means to establish hierarchical structures that allows for the dismissal of lower scale effects. Polyelectrolyte complexes were selected because the molecular rigidity and cross-link density can be tuned by adjusting the pH of the system (e.g., an indirect control of ionization). The study determined the bending rigidity through the use of a
homogenized Elastic Network Model (ENM) and further validated their result through full atomistic MD simulation, where cross-linking of polyelectrolytes is a function of the pH of the system, see Figure 7.

The authors derived expressions that relate cooperativity through the relationship of bending rigidity and cross-linking density, where a coupled macromolecule was presented with lower bending rigidity as a result of a reduced number of cross-linking. The main challenge presented in these types of studies however is the fact that the properties needed are characteristic of each material system, in this case the bond stiffness values of the polyelectrolytes were a vital part of the study. I.e., cooperativity was based on strength and stiffness measures, rather than geometry alone. Continuum compatibility conditions, in contrast, are only dependent on the deformation field, and independent of material characteristics. The metrics that are being proposed here were derived in a way that avoids the use of characteristic properties of the system and are solely geometrical in nature, thus aiming for the concept of cooperativity as a general term. The goal is to define cooperativity/compatibility of nanoscale components independent of material type, and only via geometric descriptors (e.g., nanokinematics).
1.3 Objectives

In view of this, analogous to the composite beam example (see Figure 5), the idea of compatibility at the molecular level is explored. As discussed, the motivation lies in the more established macrocompatibility laws, typically focused on deformation and strain relationships, rather than strength or equilibrium conditions. As such, the metrics target a general procedure that relies solely on geometry. The fundamental question that we wish to answer is – at what point do coupled molecules behave as a single mechanical unit? The key idea is that by using general geometrical metrics, one can objectively assess the cooperation. In turn this would allow us to effectively define new building blocks that neglect lower-scale effects [34]. This study examines the use of the gyration tensor and its invariants to evaluate cooperativity at the molecular level. The derived shape descriptors are extensions of formulations widely used in polymer science as a means to characterize the geometrical properties of a specific polymer [44-46]. Figure 8 illustrates the key points of the research.

The objectives are summarized as the following:

I. Establish means of assessment and quantification of molecular cooperativity and compatibility through the novel application of geometric measures used in polymer physics.

II. Carry out a preliminary assessment of the performance of cooperative metrics on an idealized coupled molecules model.
III. Apply metrics to known material systems to validate. Here, a mutated strand of collagen molecule is used, to observe the potential relationship between lack of cooperativity and mutation severity.

IV. Propose the exploration of the interaction between the collagen molecule and carbon nanotubes as a means to corroborate experimental work that aims at reproducing D-banded collagen fiber.

Figure 8: Molecular cooperativity summary. A) Theoretical formulation where gyration based metrics are developed. B) Testing the cooperativity metrics on an idealized macromolecule study and given positive results to ultimately carry out C) validation on an existing material system, the collagen molecule.

The organization of the dissertation is as follows: Chapter 2 provides details on the methodologies used. Chapter 3 defines the gyration tensor for a single molecule, the radius of gyration and the shape anisotropy. It is followed by a presentation of the extension of these concepts to describe relative differences between molecules, considering size, shape and orientation – a triumvirate of cooperativity metrics. In Chapter 4 a pair of ideal macromolecule chains are used as proof of
concept showing promising results. **Chapter 5** consists of measuring cooperativity in human Collagen Type I that is affected by missense mutations. **Chapter 6** presents an outline of proposed work on the interaction between of a carbon nanotube-collagen molecule composite, where understanding cooperativity could be a key factor in the reproduction of synthetic d-banded collagen fiber. **Chapter 7** offers some discussion and conclusions into the significance of the research.
2. METHODOLOGY

It is evident that solutions require methods, the need to confirm and further explore materials have resulted in the emergence of the use of computational modeling and simulation (so-called \textit{in silico} approaches) whereby material systems can be probed efficiently without cost of synthesis. This chapter briefly reviews some of the methods that are commonly used for full atomistic and coarse-grained studies, applied to the current cooperativity investigation. The chapter is focused on the general theories of classical Molecular Dynamics (MD) and the use of atomistic potentials and force fields to describe material behavior. From a structural engineer's point of view one is able to see the parallels, the use of structural dynamics to MD, as well as the similarity of force fields to constitutive laws.

2.1 Molecular Dynamics

Molecular Dynamics (MD) is a computational modeling method to drive a system of atoms and molecules at the nanoscale. The method is capable of describing the atomistic mechanisms that control the physical phenomena, in particular the mechanics involved. In general, MD involves several key steps, namely: (1) construction of the atomistic geometry; (2) definition of atomic interaction; (3) governing equations for the system; (4) initialization and energy minimization; (5) prescribed conditions of simulation; (6) integration scheme and (7) calculation of concerned properties.

\footnote{Throughout this research, Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS; http://lammps.sandia.gov/)\textsuperscript{47} Plimpton, S.J., \textit{Fast parallel algorithms for short-range molecular dynamics}. Journal of Computational Physics, 1995. 117: p. 1-19., an open-source molecular dynamics software package, was used to perform all simulations}
The first step in MD simulation is an accurate construction of the atomistic geometry which refers to a clear definition of the atomic location, the element type (i.e., the atomic mass and associated chemical properties), the partial charge of each atom, the bond connectivity among all atoms and the boundary conditions. This information is required for evaluating the inter-atomic interactions that describe the chemical properties. Subsequently, determination of the system energy is critical throughout the simulation process, and force fields (potentials) play an important role in the accuracy of the computational modeling studies by proper description of the interaction among atoms. There are a number of force fields available in the literature for different types of interactions among atoms.

Once initiated, MD can accurately simulates the motion of a group of atoms, to observe an interested property with the ultimate objective of gathering a global system property. MD results in microscopic properties which aided by statistical mechanics theory can then be related to macroscopic bulk properties. The field of statistical mechanics has extensively covered the ways in which these microscopic properties can be related to the interested global properties through thermodynamic metrics. A basic explanation on statistical mechanics and how it is central to the application of MD is given.

The basic description of MD can be summarized as the solution to the dynamical trajectory of each atom by employing interaction potentials that describe the attractive or repulsive forces between atoms. The atom is simplified into a point mass description, and the interaction potentials are based on experimental and first principles calculations based on quantum mechanics (see Figure 9). The equation of motion for each atom is solved in a Newtonian fashion, \( \mathbf{F} = ma \), and just as structural dynamics, the solution presents itself in the form of
positions $r_i(t)$, velocities $v_i(t)$ and accelerations $a_i(t)$ at each time step, thus revealing the overall dynamics of the $N$ particle system.

Figure 9: Molecular Dynamics Interactions. (a) The formal atomistic structure is replaced by a point mass representation with position $r_i(t)$, velocity $v_i(t)$ and acceleration $a_i(t)$. (b) Potential energy well, it is a basic energy decomposition based on geometrical constraints of a simple pair potential function. The depth of the well ties directly into the physical properties of the materials. Adapted from Cranford [9], used with permission © 2012 Massachusetts Institute of Technology.

The total energy of a closed system can be expressed as the Hamiltonian $(H)$, the sum of the kinetic $(K)$ and potential energy $(U)$:
\[ H = K + U \] (6)

Where the kinetic energy is defined as

\[ K = \frac{1}{2} \sum_{i=1}^{N} m_i v_i^2 \] (7)

And the potential energy as

\[ U = \sum_{i=1}^{N} U_i(r_i) \] (8)

where \( U_i(r_i) \) is a defined potential energy surface. The use of positions makes it an effective potential energy expression, other more complicated approaches take into consideration the environmental effects. The basic formulation of the potential is based on geometric constraints for each atom, that represent different behavior of the DOF of interest, i.e. bond stretching \( U(r_{ij}) \), bending \( U(\theta_{ijk}) \), and torsion \( U(\theta_{ijkl}) \), Figure 9.

The forces are obtained from the potential energy forces (sometimes called potential or force field) as \( F = -\nabla_i U(r_i) \). Thus the equation of motion can now be expressed as a function of the potential energy

\[ f_i = m_i \frac{d^2 r_i}{dt^2} = -\frac{\partial U}{\partial r_i} \] (9)

For \( i = 1, 2, \ldots, N \). This is a system of coupled second-order nonlinear differential equations that have to be solved numerically for systems where \( N > 2 \). MD uses numerical integration of the equation of motion with a carefully selected force field to obtain the forces and dynamical properties of the system. As is typically the case for equation of motion the scheme consists of Molecular Cooperativity/Compatibility K. Kwan, 2015
updating new positions of atoms from the old positions, velocities and current acceleration of particles, as well as the inclusion of boundary conditions, forces and other constraints. This is commonly carried out through the Verlet scheme which can be formulated as:

\[ r_i(t_0 + \Delta t) = -r_i(t_0 - \Delta t) + 2r_i(t_0)\Delta t + a_i(t_0)(\Delta t)^2 + O(\Delta t^4) \]  \hspace{1cm} (10)

Other schemes include the leap-frog, velocity-Verlet, and depending on the thermodynamic ensemble there are modifications to be made to the algorithms. Needless to say Eq. (10) covers particles in general, so one can also condense the system and consider a molecule itself as a particle. A downside to MD is that due to the nature of the problem, where high frequency vibrations are considered, it requires a time step in the order of femtoseconds thus the size of the systems are limited to nanometers, in essence a tradeoff between a system size and simulation length. Given a well-defined force field MD is a powerful tool that can simulate different behaviors.

### 2.2 Basic Statistical Mechanics – The Ergodic Hypothesis

The field of statistical mechanics provides the tools that permits the interpretation and analysis of molecular dynamics simulations. In fact, the results by MD are not valid instantaneously and thus are required to be averaged into the multiple possibilities of the microscopic system, it is then through statistical mechanics that one can connect the microscopic states to macroscopic variables.

Perhaps the most fundamental contribution of statistical mechanics to MD is the use of the Ergodic hypothesis. This hypothesis merely states that given sufficient time the ensemble average of “A” equals the time average:

\[ \langle A \rangle_{\text{ensemble}} = \langle A \rangle_{\text{time}} \]  \hspace{1cm} (11)
To better understand this let us superficially discuss the appropriate factors from statistical mechanics that affect the simulation.

An ensemble is the collection of possible systems that have different microscopic states but have an identical macroscopic state. In statistical mechanics averages are defined as ensemble averages, which is defined as

$$\langle A \rangle_{\text{ensemble}} = \int_p \int_r A(p,r) \rho(p,r) dp dr$$  \hspace{1cm} (12)

a function of $p, r$ which correspond to momenta and position respectively. Where

$$A(p,r)$$  \hspace{1cm} (13)

is the property that is observed and the function $\rho(p,r)$ is the probability density distribution and can be chosen depending on the macroscopic variables (NVT, NPT, see Section 2.2.1). For the NVT ensemble, it is defined as:

$$\rho(p,r) = \frac{1}{Q} \exp \left( \frac{-H(p,r)}{k_B T} \right)$$  \hspace{1cm} (14)

with $H =$ Hamiltonian, and $k_B$ the Boltzmann constant and $Q$ the partition function.

$$Q = \int_p \int_r \exp \left( \frac{-H(p,r)}{k_B T} \right) dp dr$$  \hspace{1cm} (15)

Looking at the expression one can immediately notice the solution of the integral requires that all the possible states of the system be known, an intricate task. Sampling of such can be achieved through brute-force Monte Carlo methods, or more directed/bias umbrella-sampling approaches, for example. On the other hand the time average from a MD simulation is
\[ \langle A \rangle_{\text{time}} = \frac{1}{M} \sum_{t=1}^{M} A(p, r) \]

this immediately shows the value of ergodicity. The ergodic law states that if one allows the system enough time, it will eventually explore all possible states \((t \to \infty)\), thus the average of state space and time are equivalent. For MD simulations to be considered valid, the statistics must hold, which requires enough time so that the solution space is sufficiently sampled, and as consequence validating Eq.(11).

An important consequence of the ergodic hypothesis is that in theory long simulations can be split into several “systems” and each split can be treated as an independent simulation sampling various accessible states. Though a brief explanation, this section covers the necessary relationship of MD and statistical mechanics.4

2.2.1 Thermodynamic Ensembles

Initial positions of all the atoms are defined by the coordinates of the corresponding atoms in the simulation. The total potential energy of the system \((U \text{ via Eq. (8)})\) can be calculated as a function of the current atomic positions. Prior to dynamic simulation, the entire system is subjected to a minimization of energy by varying the atomic position to decrease the total potential. This ensures a stable initial structure. After minimization, the constraints of the simulation may then be defined. The purpose of applying these constraints is to control the simulation process such that the computational experiment can be accurately performed; this is similar to the concept of applying boundary conditions in a finite element analysis.

---

Part of the boundary conditions is selecting the appropriate thermodynamic ensemble – e.g., the energetic conditions in which the system is subject. The solution of the equation of motion in Eq. (10) represents the NVE ensemble, this means that the particle Number, system Volume and the total Energy of the system remain constant through the simulation. Modifications to the equations of motions can be made such that results are realized in other thermodynamic ensembles. Table 1 presents an overview of the most common ensembles.

Table 1. Brief examination of different thermodynamical ensembles.

<table>
<thead>
<tr>
<th>Ensemble</th>
<th>Ensemble Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVE</td>
<td>Microcanonical - Constant Number, Volume and Energy</td>
</tr>
<tr>
<td>NPT</td>
<td>Isobaric-Isothermal -Constant Number, Pressure, Temperature</td>
</tr>
<tr>
<td>NVT</td>
<td>Canonical – Constant Number, Volume, Temperature</td>
</tr>
<tr>
<td>µVT</td>
<td>Grand Canonical – Constant chemical potential, Volume, Temperature</td>
</tr>
</tbody>
</table>

It should be mentioned that macroscopic quantities are always related to their associated microscopic quantities. For instance, temperature ($T$) is a macroscopic quantity and its associated microscopic quantity is the atomic velocity ($v_i$). Their relationship can be established by considering the kinetic energy (KE) stored in the system macroscopically and microscopically, and given by the following equation:

$$
\frac{3}{2} k_B T = \frac{1}{N} \sum_i \frac{1}{2} m_i v_i^2
$$  \hspace{1cm} (17)

where $k_B$ is the Boltzmann constant and $m_i$ is the mass of the $i^{th}$ atom. The left hand side of Eq. (17) describes the KE of the system from a macroscopic point of view using the quantity $T$, while the right hand side of Eq. (17) describes the KE of the system from a microscopic point of view using the quantity $v_i$. 

Molecular Cooperativity/Compatibility

K. Kwan, 2015
As an example one can modify the equations of motions by rescaling velocities as a means to control temperature, the Berendsen thermostat is one of the approaches that enables an NVT ensemble. The approach couples the system with an external heat bath that is set at the target temperature. Due to the direct relationship of velocities to temperature, one can rescale velocities such that it hits the target temperature. The rescaling parameter $\lambda$ can be written as:

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau} \left( \frac{T}{T_{set}} - 1 \right)}$$

(18)

$\Delta t, \tau, T_{set}$ are the simulation time step, the rise time, also known as the coupling constant, the time interval in which heat is exchanged with the bath, and the target temperature respectively. Velocities can then be rescaled as:

$$v_{new} = \lambda v_i$$

(19)

There are other ways to rescale velocities, e.g., rescaling through Maxwell-Boltzmann distribution, Nose-Hoover thermostat, stochastic sampling, etc., which explicitly modify either the velocities or the kinetic energy of the system based on an isothermal reservoir approach.

### 2.3 Force Fields

The importance of force fields in MD cannot be stated enough. Atomistic techniques do not require any a priori information on the material properties, as long as the appropriate interatomic interactions are set, the complete material behavior is determined via time-integrated trajectories. Force fields are expressions, developed either experimentally, through first principle quantum calculations, or a combination thereof, to describe the effects of the electrons, bond order, ionization, etc. on the atomic interactions. They allow us to estimate the energy landscape of the total system and is as previously shown a fundamental piece of MD.

Molecular Cooperativity/Compatibility  
K. Kwan, 2015
There are a variety of force fields with different accuracy and complexity. Often the choice of force field is dependent on the material system (force fields for biological materials, carbon systems, etc.) and to the behavior one intends to capture (simple bond stretching, bond breaking and bond forming etc.). The simplest formulation of a force field is the interaction between two atoms and is referred to as pair potentials.

### 2.3.1 Pair Potentials

Pair potentials are the simplest atom-atom interactions where the potential energy only depends on the distance between two particles, \( r \), and it accounts for the repulsion and attraction of the atoms. The total energy is then expressed as the sum of all atomic bonds in the system and is given as:

\[
U_{\text{total}} = \frac{1}{2} \sum_{i \neq j=1}^{N} \sum_{j=i}^{N} \phi(r_{ij})
\]  

(20)

Where \( r_{ij} \) represents the distance between particles \( i \) and \( j \), the half term is included to account for the double counting of the bonds. As such, \( \phi(r_{ij}) \) describes the potential energy of a bond formed between atoms as a function of the distance. Pair potentials must describe the effect of atoms attracting at large distances, to form bonds, and atoms repulsing at short distances due to electron overlap.

A common pair potential is the Lennard-Jones 12:6 potential and can be formulated as such:

\[
\phi_{LJ} = 4\epsilon_0 \left[ \left( \frac{\sigma}{r_{ij}} \right)^{12} - \left( \frac{\sigma}{r_{ij}} \right)^{6} \right]
\]  

(21)
where $\sigma$ represents the distance where potential is zero and $\epsilon$ is the depth of the potential well (Figure 9). The potential energy well is a concept that carries consequences not only for calculation of energy between atoms, but is also tied into the physical properties of materials. The depth of the well related to how tightly the atoms are constrained and has a direct impact on bulk properties such as the elastic modulus, thermal expansion coefficients and melting points. The usage of the simplified Lennard-Jones potential is usually controversial (primarily due to its overuse and the relatively simplicity of a two-parameter fitting), but it is generally accepted as a good fit to describe van der Waals (vdW) interactions.

One can immediately deduct an obvious limitation of pair potentials, it does not take into account the effects of the environment on the bond properties, and this oversimplification can lead to a bad fit for most applications. To remedy this multi-body potentials are employed and they provide more realistic behaviors.

### 2.3.2 Force fields for Biological Materials

The bases for organic material simulations requires that the description of various chemical interactions be taken care of. The set of characteristic chemical bonds involved in organic material behavior is more diverse, and requires the explicit consideration of covalent, ionic, and vdW interactions Figure 10.

Although there are several force fields that are used for organic modeling/simulation, a classical and widely used force field is the so-called CHARMM [52] potential. This force field is based on harmonic and anharmonic terms describing covalent interactions, in addition to long-range contributions that describe vdW, ionic interaction, Coulomb, as well as hydrogen bonds. The CHARMM force field is used for the collagen case study that is presented in Chapter 5. The mathematical formulation for empirical energy function is defined as:
\[ U_{\text{total}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{torsion}} + U_{\text{Coulomb}} + U_{\text{vdW}} \]  

(22)

Where the bond (stretching) energy is defined as:

\[ U_{\text{bond}} = \sum_{\text{pairs}} \left( a_0 + \frac{1}{2} k_{\text{bond}} (r_{ij} - r_0)^2 \right) \]  

(23)

The angle (bending) energy is

\[ U_{\text{angle}} = \sum_{\text{triples}} \left( b_0 + \frac{1}{2} k_{\text{bend}} (\theta_{ijk} - \theta_0)^2 \right) \]  

(24)

The torsion (rotational) energy is

\[ U_{\text{torsion}} = \sum_{\text{quadruples}} \left( c_0 + \frac{1}{2} k_{\text{torsion}} (1 - \cos(\theta_i)) \right) \]  

(25)

Where \(a_0, b_0, c_0\) are constant parameters, \(k\) is the respective stiffness parameter for the behavior, and \(r_0, \theta_0, \theta_i\) are the equilibrium distances, bond angles, and torsion angles, respectively. These terms are all harmonic, valid under a small deformation assumption. This implies that the systems cannot either (i) chemically react or (ii) mechanically break.

The energy due to ionic interactions takes the form of

\[ U_{\text{Coulomb}} = \sum_{\text{j} \neq \text{i}=1}^{N} \sum_{\text{j}=1}^{N} \frac{q_i q_j}{\epsilon r_{ij}} \]  

(26)

Where \(q_i, q_j\) are the partial charges of atoms \(i\) and \(j\) and \(\epsilon\) is the dielectric constant. The long-range contributions due to \(\text{vdW}\) forces are typically modeled using the Lennard-Jones potential [Eq. (21)].
Other force fields that are similar to CHARMM, are AMBER [53], GROMOS and the OPLS [54] force field.

### 2.3.3 Bond Order and Reactive Potentials

Reactive potentials are one type of potentials considered as multibody, they are typically based on the concept of bond order. Bond order (BO) can be thought of as the number of shared electron pairs between a pair of atoms, constituting a bond. BO based potentials basically describe the interactions between atoms by taking into account the atomic environment, because bond distances are mapped to bond orders and availability of electrons (so-called valence). It is especially suited to materials that have bond directionality dependence such as carbon materials. A classic and widely used BO potential is the adaptive intermolecular reactive bond order (AIREBO)[55] potential, this potential is empirically tested for carbon based materials, and its popularity can be attributed due to the computational efficiency for large atomic simulation, particularly nanotechnology critical all carbon-based systems such as carbon nanotubes and...
graphene. As an example this potential was successfully implemented in a previous study showing saturation in a graphene composite system [43], see Figure 11.

![Figure 11: Composite graphene structured modeled using the AiREBO potential. In the system above the red spheres represent pseudo-nanoparticles, and the objective was to observe the saturation of the system by looking at the interaction energy, thus showing insight towards the structure's composite action. Adapted from Kwan and Cranford [43], used with permission © 2014 Elsevier B.V.](image)

Reactive force fields are the results of efforts aimed at overcoming the limitations of classical potentials, which are unable to describe chemical reactions (e.g., bond forming/breaking). These potentials employ a bond length to bond order relationship to smooth the transition between different bond types. In other words, there is no small deformation assumption and no bond or connectivity information needs to be provided. Given a number of atoms and a reactive force field that has previously been trained for these elements, the environmental factors (e.g., temperature, pressure) will govern any bond formation of the system. In theory these force fields can handle any type of material systems based on the element type alone. A popular reactive force field is the ReaxFF formulation.

The ReaxFF [56] force field is derived from fitting against \textit{ab initio} quantum mechanical data, and is considered the state-of-the-art in terms of force fields. The complexities of the expressions for this force field is such that it is 50-100 times more expensive than the nonreactive
force fields, but it is still several orders of magnitude faster than doing complete quantum mechanical formulations. It is extensively used in case studies where one intends to capture the formation and breaking of bonds. I have previously used ReaxFF as part of a study to show the unfolding of disulfide stitched graphene as a potential structure for time delayed release [57], Figure 12.

Figure 12: Unfolding of disulfide stitched graphene. The ReaxFF potential was used for this full atomistic study that shows the critical number of bonds that is necessary such that the bending energy of the capsule itself overcomes the total disulfide bond energy, which leads to the unraveling of the graphene sheet. Adapted from Kwan and Cranford [57], used with permission © 2014 AIP Publishing LLC.

2.4 Atomistic Water, Explicit Solvent, and TIP3P Water Model

The necessity of approximating the modeling of proteins to reality requires the careful consideration of local environment – e.g., water solvation. In simplest terms, proteins do not behave like proteins in a vacuum. Solvation is fundamental for the accurate description of the structure, dynamics and thermodynamics of biological molecules. The exactitude of the results would be affected if the atomistic simulation is run in a vacuum because it would not be able to capture some of the effects of water on the material itself, e.g., the friction between material and solvent, as well as the hydrophobic effects, ionic screening and dielectric properties.

In general solvent modeling has been treated in two forms, explicit or implicit. The main difference between explicit and implicit, is in the way the water effects are treated, for the explicit version each individual water molecule is simulated and all of the interactions are calculated.
atomistically just as the solvated structure. The implicit scenario models the effect of water on the protein. These models treat the water itself as a continuum (or solvation field) and forces and energies are computed accordingly.

In all-atom force fields, water molecules are generally treated explicitly, at high computational expense. Most approaches take the molecule as rigid and consider the non-bonded interaction (electrostatic through Coulomb’s Law), with dispersion and repulsion forces generally taken care by the Lennard-Jones potential, see Eq. (21). In fact the parameters of the force field are fitted in a way that accounts for a specific water model, e.g., a TIP3P water model for CHARMM. Typically, they are classified as 1-site, 2-site, 3-site, 4-site models, attributing to interaction points. While water naturally has three, lower sites model (1-site, 2-site) are usually composed of coarse grained water models, and upper sites models (4-site) add an artificial interaction point to better distribute the electrostatic interactions. Among these models the most popular ones are TIP3P [58], TIP4P [59], SPC [60], and SPC/E [61]. For the following collagen studies using the CHARMM potential, the modified TIP3P model was used.

![Figure 13: A 3-site rigid water molecule. For the case of the modified TIP3P molecule, each atom of the molecule is an interaction site as well as the usage of the Lennard-Jones potential for interatomic interactions. The TIP3P water model is employed in the CHARMM force field and parameters are shown in Table 2.](image-url)
Table 2. The parameters for the modified TIP3P model for CHARMM is shown.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>TIP3P</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_{mass}$ (grams/mole)</td>
<td>15.9994</td>
</tr>
<tr>
<td>$H_{mass}$ (grams/mole)</td>
<td>1.008</td>
</tr>
<tr>
<td>$O_{charge}$</td>
<td>-0.834</td>
</tr>
<tr>
<td>$H_{charge}$</td>
<td>0.417</td>
</tr>
<tr>
<td>$O_{LJ}$ $\varepsilon$ (Kcal/mole)</td>
<td>0.1521</td>
</tr>
<tr>
<td>$O_{LJ}$ $\sigma$ (Å)</td>
<td>3.1507</td>
</tr>
<tr>
<td>$H_{LJ}$ $\varepsilon$ (Kcal/mole)</td>
<td>0.0460</td>
</tr>
<tr>
<td>$H_{LJ}$ $\sigma$ (Å)</td>
<td>0.4000</td>
</tr>
<tr>
<td>$OH_{LJ}$ $\varepsilon$ (Kcal/mole)</td>
<td>0.0836</td>
</tr>
<tr>
<td>$OH_{LJ}$ $\sigma$ (Å)</td>
<td>1.7753</td>
</tr>
<tr>
<td>OH K bond (Kcal/mole-Å²)</td>
<td>450</td>
</tr>
<tr>
<td>OH $r_0$ bond (Å)</td>
<td>0.9572</td>
</tr>
<tr>
<td>HOH K angle (Kcal/mole-rad²)</td>
<td>55</td>
</tr>
<tr>
<td>HOH $\theta_0$ angle (degrees)</td>
<td>104.52</td>
</tr>
</tbody>
</table>

As the majority of 3-site models, the modified TIP3P [52] water model describes water molecules as rigid with three interaction sites that correspond to each atom of the water molecule (two hydrogens bonded to one oxygen). Three site models are the full atomistic water model that are the most efficient computational-wise which adds to their popularity. Each site has a point charge and associated Lennard-Jones parameters (different from the original TIP3P water model which only accounted for interatomic behavior of the oxygen atom). These models are referred to as rigid because the initial geometry of the water molecule remains, the TIP3P model uses the actual geometry of the water molecules. Figure 13 shows the typical configuration of a 3-site model in addition to the parameter values shown in Table 2.

2.5 Long range interactions

For purposes of completeness, the long range interactions are discussed briefly. In the previous sections, see Section 2.3, the interactions presented are what are normally considered short
However in the case of modeling of protein, the fact that the system needs to be solvated in water introduces the need to account for the long range interactions of water, due to periodic boundary conditions. Water itself is a polar molecule and thus has dipole-dipole interactions - these forces present a grave problem from a computational perspective because if the interactions are felt across more than half of the simulation space, they propagate into the adjacent periodic image and compound exponentially. As such, special considerations need to be made.

Two main methods can be used in LAMMPS, the Ewald sum, which is a technique for efficient summing the interaction between an ion and its periodic images, and the particle-particle and particle-mesh (PPPM) algorithm, which is similar to the Ewald method but is usually computationally more efficient. The parameters for the TIP3P water model, see Section 2.4, change slightly when considering long range interactions and are presented in Table 3.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>TIP3P</th>
</tr>
</thead>
<tbody>
<tr>
<td>O_mass (grams/mole)</td>
<td>15.9994</td>
</tr>
<tr>
<td>H_mass (grams/mole)</td>
<td>1.008</td>
</tr>
<tr>
<td>O_charge</td>
<td>-0.830</td>
</tr>
<tr>
<td>H_charge</td>
<td>0.415</td>
</tr>
<tr>
<td>O_L-J ε (Kcal/mole)</td>
<td>0.102</td>
</tr>
<tr>
<td>O_L-J σ (Å)</td>
<td>3.188</td>
</tr>
<tr>
<td>H_L-J ε (Kcal/mole)</td>
<td>0</td>
</tr>
<tr>
<td>H_L-J σ (Å)</td>
<td>0</td>
</tr>
<tr>
<td>O_H-L-J ε (Kcal/mole)</td>
<td>0</td>
</tr>
<tr>
<td>OH_L-J σ (Å)</td>
<td>0</td>
</tr>
<tr>
<td>OH K bond (Kcal/mole-Å²)</td>
<td>450</td>
</tr>
<tr>
<td>OH r₀ bond (Å)</td>
<td>0.9572</td>
</tr>
<tr>
<td>HOH K angle (Kcal/mole-rad²)</td>
<td>55</td>
</tr>
<tr>
<td>HOH θ₀ angle (degrees)</td>
<td>104.52</td>
</tr>
</tbody>
</table>

Extensive details of long range interactions can be found in Allen and Tildesley.48. Ibid.

Molecular Cooperativity/Compatibility  
K. Kwan, 2015
2.6 Multi-scale Modeling (Coarse Graining)

Coarse graining is a modeling technique that provides a way to simulate and investigate systems where the properties of interest are in the mesoscale (effectively between full atomistic and micro), which are currently inaccessible through a full atomistic representation and unobtainable in continuum theory. The idea of coarse graining ties directly to the building block problem and the time-scale dilemma, where homogenized units can represent reduced DOFs, see Section 1.1.1. In general the community aims towards successfully establishment of hierarchical systems as a means to design materials from a bottom up approach, something that Nature repeatedly shows, and on the other hand current computational costs limit the time-length scale accessibility. The significance of coarse graining is that it would open the possibilities of observing a material system at all scales, thus obtaining a complete theoretical formulation, and key information for material design. In structural engineering, coarse graining is extensively used as a means to simplify the model by “ignoring” the degrees of freedom corresponding to superfluous observations. This process is known as condensation and it is just one way to coarse grain a model, albeit the usage of such a method is not necessarily tied the time-scale limitation. In addition, the general study of structural engineering itself can be considered a coarse-graining of sorts. A specific example in the form of structural frames is shown in Figure 14.
In this case the idealization of the truss as a beam based on beam theory is made possible due to the understanding of the components, the truss elements and the material properties. The everyday analysis carried out does not need to worry about each minute characteristic at other scales, which in this case is possible due to the exhaustive amount of testing at the material level and a theoretical framework that ensures a sufficient representative model of the beam element.

The arduous process of coarse-graining can be summarized as the development of a model by choosing a potential, which allows one to describe the behavior that the proposed model intends to capture. Each potential can then individually characterized by single atomistic
investigations, and finally the accuracy of the model is tested by satisfying the universal law of conservation of energy, two well-known approaches include the equivalence of direct energy between the atomistic and the coarse grained system and the energy conservation through consistent mechanical behavior. In this “finer trans coarser” approach, there is a kind of theoretical “handshaking” between scales, where pertinent information and parameters are passed to higher order models.

The actual method of coarse graining varies, depending on the material system and behavior that one intends to capture. In the first study that is presented shortly (Chapter 4), a pair of macromolecules are represented in a coarse-grained bead-spring model, where each “bead” can represent multiple atoms or a chemical group, and in this case the differentiation between molecule and atom can be ignored. The harmonic terms that one can normally use for atomic representations serve well for this coarse grained scenario. This is the case because of the objective and simplicity of the model structure by design. A brief discussion of some popular approaches follows, additional complexities of coarse graining are detailed in Cranford and Buehler [62].

### 2.6.1 Elastic Network Models

The Elastic Network Models (ENM) are the simplest form of coarse graining, it consists on assigning a single potential to the condensed structure, these models are usually tied normal mode analysis (NMA), a method to study long-time dynamics and elasticity of biomolecules. It can be formulated as single pair potential between atoms.
\[ U_{\text{ENM}} = \sum \phi_{\text{ENM}}(r) \]  

(27)

Where the potential function for each pair of atom is presented as a harmonic spring bond with stiffness \( K \) and an equilibrium distance of \( r_0 \)

\[ \phi_{\text{ENM}}(r) = \frac{1}{2} K (r - r_0)^2 \]  

(28)

These types of model are akin to a truss model in structural engineering, see and several variants of the model tweak the linearity of the stiffness of the spring, the simplest one being constant stiffness \( K \) [63]. The different definitions of the stiffness parameter \( K \) allow the model to provide more complex descriptions of interactions. An example is given in Figure 15.

(a)

(b)

Figure 15: ENM for polyelectrolyte complex. (a) Full atomistic representation of coupled PAA and PAH chains. (b) General elastic network connecting the atoms with harmonic springs, in this case the model is unaware of the cross-linking scheme specific for this structure which in turn explains the absurd amount of cross-links. Adapted from Cranford [9], used with permission © 2012 Massachusetts Institute of Technology.
ENMs, through a single potential description, are a simple model that can provide insight into the complex molecular deformations, their ability to describe unique phenomena of self-assembled systems have made them a popular choice in biological applications [63-71].

2.6.2 Two Potential Freely Jointed Chain Polymer Models

Due to the simplicity of the ENM formulation, they are useful as a means to do structural or modal analysis on a single large protein, however they are unable to capture dynamic behavior between macromolecules. This limitation can be overcome by combining two simple potentials: one to represent the intermolecular interaction and the other for intramolecular interactions. Polymer systems are frequently represented in this freely jointed chain (FJC) representation, where:

\[
U_{\text{FJC}} = U_{\text{bonded}} + U_{\text{nonbonded}} = \sum \phi_{\text{FENE}}(r) + \sum \phi_{\text{WCA}}(r) \tag{29}
\]

The bonded interactions are represented by a finitely extensive nonlinear elastic (FENE) potential, and is implemented to prevent polymer chains from crossing each other and is defined as

\[
\phi_{\text{FENE}}(r) = -\frac{1}{2} kr_0^2 \ln \left[1 - \left(\frac{r}{r_0}\right)^2\right] \tag{30}
\]

for \( r < r_0 \), for \( r \geq r_0 \) \( \phi_{\text{FENE}} = \infty \). The nonbonded interactions is presented by a variation of the L-J potential, see Section 2.3.1, Eq. (21), the Weeks-Chandler-Andersen (WCA) potential is 12:6 L-J potential truncated at the depth of the potential well, thus only taking into account the repulsive side of the L-J potential

\[
\phi_{\text{WCA}} = 4\varepsilon_0 \left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \tag{31}
\]
For $\frac{R}{\sigma} < 2^{1/6}$ and zero for $\frac{R}{\sigma} \geq 2^{1/6}$. In combination with the attractive FENE potential, this model can successfully describe stretching, orientation and deformation of polymer chains and simple biomolecules [72-77].

2.6.3 The MARTINI Force Field

As presented in earlier sections 2.6.1 and 2.6.2, the natural progression shows addressing the single potential shortcoming by the addition of a second potential. However FJC's are not general enough, and thus the MARTINI [78, 79] force field was developed as an attempt to provide a generalization of interactions for biological systems. The force field was created in a manner that took advantage of the fact that the majority of the biological systems are composed of a group of 20 amino acids. Essentially by coarse-graining the naturally occurring amino acids into common groups, see Figure 16, one could easily assemble different molecular structures with very few parameters.

The intention of this force field was to define as little “building blocks” as possible without sacrificing the chemical nature of the components. It achieved this by formulating a general parameterization that is remains independent of the model. So the dilemma becomes a trade-off between accuracy of the model and wide range applicability to different biological systems. The interactions considered are polar, nonpolar, apolar and charged. The force field can be represented as
Figure 16: The coarse-grained representation of all the naturally occurring amino acids for the MARTINI force field. Reproduced with permission from Monticelli [79], used with permission © 2008, American Chemical Society.

\[
U_{MARTINI} = U_{\text{bonded}} + U_{\text{angle}} + U_{\text{nontbonded}} = \sum \phi_T(r) + \sum \phi_\theta(\theta) + \sum \phi_{LJ}(r)
\]  

(32)

where the bonded and angle energies are implemented through harmonic potentials

\[
\phi_T(r) = \frac{1}{2} k_T (r - r_0)^2
\]  

(33)

\[
\phi_\theta = \frac{1}{2} k_\theta (\theta - \theta_0)^2
\]  

(34)

Where the harmonic potential parameters, \( k_T, r_0, k_\theta \) and \( \theta_0 \) are common to all bead types. The nonbonded interactions are taken care by using a 12:6 L-J potential, Section 2.3.1, where the parameters \( \epsilon_{ij} \) and \( \sigma_{ij} \) are grouped into distinct subsets of interaction groups.

As implied earlier the advantage of the MARTINI force field is the applicability to a wide range of biological systems, however a major disadvantage is that the coarse graining scalability...
of the force field is fixed by the relative small size of the functional groups, even so, it is much more efficient than full atomistic simulations. For the particular case of the collagen molecule, the MARTINI force field had to be extended[80] for two reasons, the collagen molecule contains the amino acid hydroxyproline which is not part of the 20 naturally occurring amino acids, so the parameters for this amino acid had to be fitted and secondly, the original formulation cannot describe the characteristic triple helical structure of collagen. To address the inability to describe the triple helicity, the stiffness parameters as well as bond, angle and dihedral information were adjusted through a statistical endeavor.

Here, we use a MARTINI-like formulation in Chapter 4, to represent a semi-flexible macromolecule with finite axial stiffness, bending rigidity, and pairwise interactions. The same model interactions are taken into account, Eq. (32), however the parameters of the potential themselves are not specific to any amino acid. As a result the structure can be considered a generic homogeneous polymer chain or the backbone of a polypeptide.
3. COOPERATIVITY METRICS

The main issue of introducing the concept of “cooperativity” between molecules is the current lack of any quantitative measures at the molecular scale (or at any scale, for that matter). Thus, new metrics must be developed for comparison between systems for adequate assessment of “cooperative” behavior. As a starting point the metrics will focus on large molecules, most of which can be ideally represented in a general ideal chain system. We also wish the metrics to be general, transferable across systems and scales, and geometric based (e.g., independent of specific chemical properties or molecular interactions).

As an inspiration, the field of polymer science has rigorously developed metrics to describe the size and shape of single polymers [44-46, 81-85], which may be useful in characterizing geometric cooperativity. These metrics are based on the concept of gyration, or, specifically, the components of a gyration tensor. A key advantage of such a geometric measure is that mechanical properties (such as molecular stiffness or cross-link strength) are not required, and thus the metrics should be transferable across multiple molecular systems. With the assumption that cooperative molecules would have strongly correlated shapes and orientations subject to deformation, and this data is reflected in the gyration components, the gyration tensor of the macromolecule is exploited as a basis to define cooperativity.

The chapter starts with a brief discussion on ideal chain configurations and the relevance to cooperativity.6

6 A deeper understanding of polymers physics in general, and the study of statistical conformations in ideal chains can be sought in 86. Flory, P.J., Principles of Polymer Chemistry. 1953: Cornell University Press.
3.1 Polymer Physics: Ideal Chain Configurations

Polymers, see

Figure 17, are large molecules composed of repeated sequences known as monomers, these monomers can be thought of as fundamental building blocks, consistent with the discussion in Section 1.1.1, it is a chain of atoms that essentially provide a backbone for which other groups of atoms can connect. Synthetic polymers are one of the most essential materials, they are mainly used for fibers [87], and when molded into required shapes are commonly known as plastics. Polymer physics is responsible for the characterization of the bulk properties of polymer related materials, in particular it observes the occurrence of chain entanglements which plays a vital role in the shape transitions, behavior under compression, hysteresis as well as flow, viscosity and swelling [86, 88, 89]. The characterization of coil behavior of the chain is a key factor in the correlations between structural and physical properties. Some conformational and statistical measures of polymers can be extracted by inspecting a simplified ideal chain model representation. These measures are for single chains only, and are not cast in a manner in which multiple chains or cooperative chains are to be observed or characterized.
Figure 17: Simple polyethylene chain. A sample polymer, notice the one dimensional configuration characteristic of polymers in general, the backbone is composed of carbon atoms (grey) and hydrogen atoms (white).

3.1.1 End-to-end Distance and Polymer Size

The end-to-end distance is generally used as a size measurement and represents the average distance between the first and last segment. This measurement has been used to determine pore size, a vital parameter in drug release applications [90], adsorption process [91], DNA mobility [92], among others. For an ideal chain of \( n \) monomers with bond length \( l \), see Figure 18, given the random variations of the monomer the mean end-to-end distance \( \langle R \rangle = 0 \), the mean squared end-end distance however is not trivial and defined as:

\[
\langle R^2 \rangle = l^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \langle \cos(\theta_{ij}) \rangle
\]

(35)

For the case of a freely jointed chain, where all bond lengths are the same, there are no correlations between angles and no treatment of torsional angles, \( i.e., \langle \cos(\theta_{ij}) \rangle = 0 \) for \( i \neq j \), then the expression of Eq.(35) is simplified into \( \langle R^2 \rangle = nl^2 \). Additionally in the case of worm-like chain (WLC) models, a theoretical relationship between the persistence length and end-to-end distance can be defined as:

\[
\langle R^2 \rangle = 2Pl \left[ 1 - \frac{P}{l} \left( 1 - e^{-l/P} \right) \right]
\]

(36)

Where \( P \) is the persistence length, a measure of the polymer stiffness, and \( l \) the length of polymer.

One can observe that for \( l >> P \) this relationship then yields \( R^2 = 2Pl \), thus for continuously flexible chains a larger end-to-end distance implicates a stiffer polymer.
Due to the natural inclinations of polymers to coil, the end-end distance is not a sufficient measure as it cannot represent the internal changes of conformation within the polymer, shown in Figure 20.

Figure 18: Ideal Chain. An ideal chain representation of a polymer with \( n \) monomers and \( l \) bond lengths. \( R \) represents the end-end distance, where the coiled nature of the polymer is not captured in this measurement.

### 3.1.2 Radius of gyration

A more intuitive way to measure the “size” of a polymer is through the radius of gyration because it gives a “3D” indication of the size of the polymer coil. The radius of gyration is used to evaluate solvent affinity, \( i.e. \), a larger radius of gyration would indicate a solvent in which the chains are fully solvated, whereas a poor solvent cannot prevent the coiling of the polymer [93]. The radius of gyration is also related to the molecular weight which is crucial for the determination of numerous physical properties such as transition temperatures, stiffness, strength, viscoelasticity, toughness and viscosity [94]. The radius of gyration is defined as:

\[
R_g^2 = \frac{1}{n} \sum_{j=1}^{n} \sum_{j=1}^{n} (\mathbf{R}_i - \mathbf{R}_j)^2
\]

\[ (37) \]

---

*Molecular Cooperativity/Compatibility*  
K. Kwan, 2015
For the case of equivalent freely jointed chain the radius of gyration is \( \langle R_g^2 \rangle = n l^2 / 6 = \left( \langle \overrightarrow{R}^2 \rangle \right) / 6 \).

The radius of gyration is smaller than the root mean squared (RMS) end-end distance by a factor of \( \sqrt{6} \), an illustration of the physical significance of the end-end distance and radius of gyration is shown in Figure 19. Table 4 shows the radius of gyration for some common polymers, provided that “N” is known some indication of shape can be inferred for these ideal cases.

Table 4. Radius of gyration for common polymer architectures.

<table>
<thead>
<tr>
<th>Polymer Architecture</th>
<th>Rg²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Chain</td>
<td>Nb²/6</td>
</tr>
<tr>
<td>Rings</td>
<td>Nb²/12</td>
</tr>
<tr>
<td>f-arm star</td>
<td>((N/f)b²/6)(3-2/f)</td>
</tr>
<tr>
<td>H-Polymer</td>
<td>(Nb²/6)89/125</td>
</tr>
</tbody>
</table>

In the case of polymer science, these measurements are used to qualitatively judge the coiled nature of a single polymer, and the implication of size in the physical and mechanical properties of the system. For the purposes of cooperativity assessment these metrics can be employed in a novel manner, taking advantage of the geometrical significance of each. The end-to-end distance however was not considered due to the one dimensional description that lacks information about the direction and size which are necessary for a complete illustration of cooperativity. With this let us formally introduce the gyration tensor and how it contains the properties that are of interest when assessing cooperativity.
Figure 19: End-end distance vs. radius of gyration. The interpretation of the radius of gyration \( R_g \) can be thought of as a sphere, taking into account the coiled behavior that is unnoticed if one only looks at the end-end distance \( R \).

### 3.2 The Gyration Tensor

The gyration tensor \( S \) is a second-order, real, and symmetric tensor that describes the second moments of position of a collection of particles composing a macromolecule, defined as:

\[
S = \frac{1}{N} \left[ \begin{array}{ccc} 
\sum_i (x_i - x_{cm})^2 & \sum_i (x_i - x_{cm})(y_i - y_{cm}) & \sum_i (x_i - x_{cm})(z_i - z_{cm}) \\
\sum_i (x_i - x_{cm})(y_i - y_{cm}) & \sum_i (y_i - y_{cm})^2 & \sum_i (y_i - y_{cm})(z_i - z_{cm}) \\
\sum_i (x_i - x_{cm})(z_i - z_{cm}) & \sum_i (y_i - y_{cm})(z_i - z_{cm}) & \sum_i (z_i - z_{cm})^2 
\end{array} \right] \tag{38}
\]

Where \( x_i \) denotes the component of position in the x direction of particle \( i \), and \( x_{cm} \) the center of mass in the x direction, and is valid for the other directions of the Cartesian coordinates (see Figure 20).
Figure 20: Gyration schematic for single molecule of any random configuration. A gyration tensor can be determined that describes the second moments of position of a collection of particles composing a macromolecule through geometry alone. Here the difference between particles $i$ and $j$ in $x$-direction are indicated. The resulting gyration tensor can be diagonalized, principal axes of gyration can be defined which intersect at the center of mass for the molecule. The magnitude of the axes reflect the eigenvalues of the gyration tensor which can be used to define several parameters that can be exploited to define molecular cooperativity. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

Similar to the mechanics stress or strain tensors, the gyration tensor can be manipulated in orthogonal space to reflect principal directions and magnitudes of gyration. This is accomplished via solution of the classic eigenvalue problem. The principal directions of the gyration tensor show the linear extent of the system and the intersection of the directions correspond to the center of mass of the system. Of note, the moment of inertia and radius of gyration are related and can be proportional. The difference between the radius of gyration and moment of inertia is that the moment of inertia is weighted by the mass of each particle, thus if all particles have equal mass the radius of gyration is directly proportional.

---

7 Please refer to Appendix A for the solution of the general eigenvalue problem.

Molecular Cooperativity/Compatibility

K. Kwan, 2015
To illustrate the physical meaning of radius of gyration and its relationship to the moment of inertia, let us consider the 1-D case, see Figure 21.

**Figure 21.** The moment of inertia for a set of particles can be expressed as:

\[ I = \sum_i m_i r_i \]  

where \( I \) is the inertia, \( m_i \) and \( r_i \) are the mass and distance from arbitrary position of each particle.

The radius of gyration is then expressed as

\[ R_g = \frac{I}{M} \]

where \( R_g \) is the radius of gyration and \( M \) is the total mass of the system. The physical understanding of the radius of gyration in the 1-dimensional case, is that it corresponds to the distance from the center where one can treat the collection of particles as a single point.

Figure 21: A schematic of a collection of particles, in 1-D the relationship between the Moment of Inertia Eq.(39) and Radius of Gyration Eq.(40) is easier to visualize.

One of the benefits of using the geometry-based gyration tensor to define cooperativity metrics - rather than other quantities such as potential energies - is the fact that the gyration tensor can
be measured experimentally (through microscopic image analysis [95] or indirectly through NMR spectra [96]). Due to the properties of the tensor, a set of shape descriptors used to show molecular cooperativity can then be derived from the invariants or eigenvalues of the gyration tensor.

3.2.1 The Eigenvalues of the Gyration Tensor

Let us conveniently start the discussion of eigenvalues\(^8\) of the gyration tensor through a diagonalized version of the matrix \(S\), Eq.(38), attained via transformation to the principal coordinate system, where:

\[
S = \begin{bmatrix}
\lambda_1 & 0 & 0 \\
0 & \lambda_2 & 0 \\
0 & 0 & \lambda_3 \\
\end{bmatrix} \equiv \begin{bmatrix}
R_1^2 & 0 & 0 \\
0 & R_2^2 & 0 \\
0 & 0 & R_3^2 \\
\end{bmatrix}
\]

\(\lambda_1 > \lambda_2 > \lambda_3\) are the three eigenvalues of the gyration tensor. The values \(R_1, R_2, R_3\) reflect orthogonal components of a vector in 3D space. As such, the ratio of the eigenvalues represent the relative spatial extension of system. For example, if all eigenvalues are equal, then the linear extension is the same for all directions, which one can picture as a spherical projection/shape. In contrast, if \(R_1 \gg R_2\) and \(R_1 \gg R_3\), then the extension is predominantly in the \(R_1\)-direction, and the molecule can be considered predominantly oriented along a common axis.

The eigenvalues of any system in general are invariants to the rotations of the system, or selection of basis coordinates, and for that reason become invaluable when looking for mathematical properties of the system. Any quantity that is a result of operations of eigenvalues is known as an invariant. Three dimensional matrices possess three independent invariants, \(I_1, I_2, I_3\), as a result of the characteristic eigenvalue equation. The invariants of the gyration tensor can

---

\(^8\) See Appendix A.
be formulated to give descriptors (e.g., parameters that describe the distribution of particles), including radius of gyration, shape anisotropy, asphericity, acylindricity, etc [97].

The first invariant of the gyration tensor, is given by the trace of the tensor, and is known as the mean square radius of gyration:

\[ I_1 = Tr(S) = \lambda_1 + \lambda_2 + \lambda_3 = R_g^2 + R_g^2 + R_g^2 = R_g^2 \]  \hspace{1cm} (42)

where \( R_g \) – the radius of gyration - defines the size of the chain and is indicative of molecular size (which can easily be shown using Pythagoras relation for the 3D vector). In protein analysis, for example, the radius of gyration is used as a means to indicate how folded a polypeptide chain is. The radius increases with an increase in molecular length, mass, and rigidity.

The second single molecule descriptor comes as a consequence of computational difficulties. As mentioned earlier the ratios of eigenvalues can be used to determine how close the system is to being spherical, however solving the eigenvalue problem has been, historically, an expensive process. As a turnaround a ratio of invariants coined as the asphericity\(^9\)/shape anisotropy [83] was created as a convenient way to characterize the shape, and is defined:

\[ \kappa^2 = \frac{3}{2} \frac{Tr(\hat{S}^2)}{(Tr(S))^2} \]  \hspace{1cm} (43)

where

\[ \hat{S} = \begin{bmatrix} \lambda_1 - \bar{\lambda} & 0 & 0 \\ 0 & \lambda_2 - \bar{\lambda} & 0 \\ 0 & 0 & \lambda_3 - \bar{\lambda} \end{bmatrix} \]  \hspace{1cm} (44)

\(^9\) The term asphericity and shape anisotropy are equivalent and used interchangeably.
with \( \bar{\lambda} \) denoted as the average eigenvalue, alternatively \( \hat{S} = S - \frac{1}{3}(Tr(S))I_{3 \times 3} \), and \( I_{3 \times 3} \) is the identity matrix. This represents the deviatoric matrix whose eigenvalues characterize the asymmetry of the system. As an analogy, in mechanics such term in the form of the stress and strain tensors represents the deviatoric stress/strain which are related to the shear stresses/strains of the object.

The trace of \( \hat{S}^2 \) is defined as:

\[
\text{Tr}(\hat{S}^2) = \sum_{i=1}^{3} (\lambda_i - \bar{\lambda})^2
\]  

(45)

Which is the variance of the eigenvalues of the gyration tensor Eq.(41), and hence is the measure of how much the polymer deviates from spherical symmetry.\(^{10}\) By definition, its value must range from 0 to 1, see Figure 22, one meaning a perfectly linear chain and zero for symmetrical conformations.

Note that Eq.(38) and Eqs. (42)-(43) are all calculated at a specific time instance - \( S(t) \), \( R_g(t) \), and \( \kappa^2(t) \) - and thus the result is a vector of each quantity. Moreover, depending on purpose, they can be expressed as a time-averaged quantity, \( \langle S \rangle \), \( \langle R_g \rangle \), and \( \langle \kappa^2 \rangle \).

\(^{10}\) A spherical conformation would require that all eigenvalues be the same, thus subtracting the average would be zero.
It is of importance to note that the beauty in the mathematics involved is born out of the necessity to avoid the eigenvalue computation. The previous expressions all have equivalent eigenvalue based expressions, however the computational efficiency lies on the ones mentioned here. It all traces back to the linear algebra property that states that for trace, the sum of the diagonal terms, is equal to the sum of the eigenvalues. Thus the computations are carried out by using the trace operator, who itself is an invariant Eq.(42).

With this framework set, let us consider the case of two macromolecules. Whether uncoupled or cross-linked, each molecule can be described individually by a gyration tensor (and its associated invariants). The primary assumption is that if molecules with the same characteristics ideally cooperate, – e.g., deform together in every manner – they will have equivalent gyration tensors. For example, based on the previous definition of cooperativity, if molecule “A” has a propensity to extend in space and is sufficiently coupled with molecule “B”, then molecule “B” would extend in the same manner. The metrics based on the gyration tensor can then be formulated to describe the three most relevant features of the molecules, or a body in space in general: (i) size, (ii) shape, and (iii) orientation. This concept is illustrated in Figure 23.
The deviation of gyration, for structures that have identical geometrical characteristics and are cooperating it is evident that the deviation is zero. In the case of S1 and S2 the uncooperativity is represented as a misalignment, thus the deviation is greater than the one for S3 and S4.

The first proposed metric for molecular cooperativity is the deviation of gyration. The idea is to calculate the gyration tensor for each individual chain and obtain a gyration compensator tensor (so-called because it compensates for the difference between “A” and “B”), which is defined as:

$$\Delta S = \left| S_A - S_B \right|$$  (46)

We note the absolute value is taken to ensure positive terms that describe the relative difference between $S_A$ and $S_B$. Moreover, from inspection of Eq.(46), it seems evident that if both chains were identical and cooperating perfectly the gyration tensor would be exactly zero ($\Delta S = 0$), however because the behavior is highly dependent on the number of potential cross linking and due to randomness in the system it would be limited to approach zero ($\Delta S \to 0$).

More advantageously, since Eq.(46) is a linear operation, similar to the gyration tensor, $\Delta S$ is a symmetric $3\times3$ matrix with associated eigenvalues and eigenvectors. As such, manipulation of the...
gyration variance tensor (e.g., invariants, principal directions, etc.) can be used to quantify molecular cooperativity in an objective manner. We first explore the differences in size.

3.3 Size Metrics

Since the form of $\Delta S$, from Eq.(46), is functionally the same as a gyration tensor, it can be thought of as the gyration terms describing a “virtual” molecule, which expresses the difference in gyration needed to compensate the behavior of the molecular composite. In simpler terms, $\Delta S$ is a small molecule that, if added to molecule “A”, the gyration would match that of molecule “B” (or vice versa). If we can quantify the relative size of the “virtual” molecule, with this in hand the deviation between the original molecules can be objectively compared.

Motivated by Eq.(42), the trace of such virtual gyration tensor, $\Delta S$, labeled hereon as slip ($\Omega$), can be expressed as:

$$\Omega = Tr(\Delta S) \quad (47)$$

Simply put, Eq. (47) represents the radius of gyration of the “virtual” molecule, and gives a measure of the relative size difference and projected mismatch between the chains that form an arbitrary molecular complex. The smaller this number is the more similar in size the components are along the same axis. It is clear that, with cooperativity of identical molecules, the slip tends to zero ($\Omega \rightarrow 0$), but that may not necessary be the case always, thus a more encompassing statement is that the slip must remain relatively constant to the initial slip of the system. Also clear is that the slip is dependent on size – that is, larger macromolecules will necessarily have larger absolute slip values, regardless of cooperativity. As such, the slip is only a relative measure, which can be applied to molecules of similar size, length, and mass. However, to alleviate the size dependence, one can normalize by the size of one of the molecules if beneficial.

Molecular Cooperativity/Compatibility  
K. Kwan, 2015
In addition to slip, because the gyration tensors for each chain would be the same under perfect cooperating circumstances, another metric that can be used to confirm cooperativity is the Euclidean norm of a square matrix (a common measure of size). This is defined as:

$$\||\Delta S\||_2 = \sqrt{\lambda_{\text{max}}(\Delta S^* \Delta S)}$$  \hspace{1cm} (48)

where $\lambda_{\text{max}}$ is the largest eigenvalue of the squared gyration compensator tensor. The resulting relative “distance” between gyrations, normalized by the norm of a single molecule, can be expressed as:

$$\||S_{\text{relative}}|| = \frac{\||\Delta S\||_2}{||S_A||}$$  \hspace{1cm} (49)

Again, it is clear that, with cooperativity, $\||S_{\text{relative}}|| \to 0$. It is also noted that both the slip and the relative distance are also time-varying, dependent on a particular molecular conformation.

We next proceed to assess the mismatch in shape of a general complex.

### 3.4 Shape Metric

The asphericity for each chain is calculated and the premise is that if both chains are again working in a synergistic fashion, they will have the same symmetric or asymmetric configuration at the same time instance (independent of the deviation). The differential anisotropy at a time instance $t$ is expressed as:

$$\delta \kappa^2(t) = \kappa_A^2(t) - \kappa_B^2(t)$$  \hspace{1cm} (50)

where each term is calculated via Eq.(43). Note that different from the slip, Eq.(47), the differential anisotropy is expressed as the difference of each chain - the reason being that the trace of a matrix representing slip is a linear operation, while the shape anisotropy definition is non-linear. Even though one can presume that a relative shape anisotropy of a “virtual” molecule...
can be obtained as $\kappa^2(\Delta S)$, this would lose its physical meaning in terms of cooperativity. For example, if the gyration deviation is more spherical (which indicates a spatially homogeneous distribution; $\kappa^2 \to 0$) or rod-like (an extended distribution; $\kappa^2 \to 1$), it may indicate the extent of coupling, but can fluctuate over time.

To compare across systems over time, the relative molecular shape is defined as the norm of the differential asphericity vector, or:

$$||\Delta \kappa^2|| = \sqrt{\delta \kappa^2(1)^2 + \ldots + \delta \kappa^2(n)^2}$$

(51)

with $n$ being the number of steps. Again, over time the molecules are free to evolve from spherical to extended configurations, but the differential asphericity – the difference between asphericity at an instant - should decrease as cooperativity increases. Note also that, as $\kappa^2 \in (0,1)$, then $\delta \kappa^2 \in (0,1)$. With metrics for size and shape established, we turn to a final measure of molecular orientation.

### 3.5 Orientation Metric

If we consider a single molecule suspended in space, it has extension along an arbitrary vector. Since the vector is dependent on the basis Cartesian (or similar) coordinate system, defining the orientation of a single molecule has little value. However, a pair of molecules A and B, shown in Figure 24, (coupled or otherwise) defines two vectors, and thus can be oriented with respect to each other. Again, it is presumed that, if cooperative the resulting orientation at all times would remain relatively constant, in the ideal case where both cooperative structures are identical this orientation would approximate zero.
To quantify the orientation, we consider the diagonal components of $S$ (which define the square of the radius of gyration along the predominant axes molecule). A vector of the molecules predominant gyration axis can be defined by these components:

$$\vec{r} = S^{(i)}_{xx} \hat{i} + S^{(i)}_{yy} \hat{j} + S^{(i)}_{zz} \hat{k}$$

(52)

We note that this is not the principal axis, but rather approximate axis of alignment within the basis coordinate system. We consider the angle between the axes of each molecule, such that the skew, $\alpha$, is defined as:

$$\alpha = \cos^{-1} \left[ \frac{S^{(1)}_{xx} S^{(2)}_{xx} + S^{(1)}_{yy} S^{(2)}_{yy} + S^{(1)}_{zz} S^{(2)}_{zz}}{\left( S^{(1)}_{xx} S^{(1)}_{yy} + S^{(1)}_{yy} S^{(1)}_{zz} + S^{(1)}_{zz} S^{(1)}_{xx} \right)^{1/2} \left( S^{(2)}_{xx} S^{(2)}_{yy} + S^{(2)}_{yy} S^{(2)}_{zz} + S^{(2)}_{zz} S^{(2)}_{xx} \right)^{1/2}} \right]$$

(53)

### 3.6 Summary

A triangle that illustrates cooperativity at the nanoscale is presented in Figure 25, each side corresponding to the size, shape and orientation which are represented by the slip, differential anisotropy and skewness. Moving forward, the intention is to subsequently validate and explore...
their application in dynamical systems that evolve in time and space. Before delving into the specifics of the validation, an important observation concerning the metrics needs to be specified, while it is tempting to think of the metrics as being limited to chains as per the discussion in the chapter, a more complete way is to think of the metrics as just differentials based on the gyration tensor, where the difference is between structures that one determines on a per case basis of the system of interest (e.g., gyration difference between bundles of molecules as opposed to the individual chains of the bundle).

The first system selected for the examination of the cooperativity metrics is an ideal case of a pair of chain-like molecules (Chapter 4), a model that is inspired from the aforementioned electrolyte study [30]. The usage of an ideal chain model allows for the explicit addition/removal of cross-links to achieve target densities and distributions and explore the resulting effect on coupling, compatibility, and cooperativity. The second system considers a tropocollagen molecule, particularly the effect of mutations (Chapter 5). The triple helical structure is presumed cooperative, and we explore if the derived metrics can indicate the presence of mutations by gyration measures alone (e.g., the inverse of the problem explored in Chapter 4). Both physical systems serve as validation for the cooperativity metrics.
Figure 25: Cooperativity triangle. The derivation of this trifecta that describes size, shape and orientation is used as a definition of nano-cooperativity. The combined effort of the metrics are necessary to assess cooperative behavior in the structure.
4. COOPERATIVITY IN IDEAL MACROMOLECULES

To first explore the validity of the proposed metrics, the simplest case is required to demonstrate proof-of-concept. A general model of two coupled molecules is developed to be implemented in equilibrium molecule dynamics (MD) simulations. As any new formulations, the confirmation of the usefulness is an arduous process. This basic model is used as a preliminary stepping stone to more complex material systems.

4.1 Molecular Model Formulation

A pair of molecular chains (molecule “A” and molecule “B”) is described by a minimalistic coarse-grained model for semi-flexible homopolymers (see Figure 26), with a corresponding MARTINI-like force field expressed as:

\[
U_{\text{Total}} = U_{\text{Bond}} + U_{\text{Angle}} + U_{\text{Nonbonded}} = \sum \phi_T(r) + \sum \phi_\theta(\theta) + \sum \phi_{\text{LJ}}(r)
\]

incorporating intramolecular stretching \(U_{\text{bond}}\) and bending \(U_{\text{angle}}\), and intermolecular nonbonded interactions \(U_{\text{nonbonded}}\). Harmonic potentials are implemented for the bonded and angle potentials, where:

\[
\phi_T(r) = \frac{1}{2} k_T (r - r_0)^2
\]

\[
\phi_\theta = \frac{1}{2} k_\theta (\theta - \theta_0)^2
\]

with \(k_T\) and \(k_\theta\) describing axial and bending stiffness respectively, \(r_0\) the equilibrium spacing, and \(\theta_0\) the equilibrium angle. Non-bonded pair interactions (intermolecular) are described as a Lennard-Jones 12:6 potential given by Eq.(21). We implement a cutoff equal to the equilibrium spacing such that there is no attraction between molecules (e.g., only the repulsive regime of the
LJ potential). As such, molecular coupling is only controlled \textit{via} finite cross-links, wherein a harmonic bond is added between molecule “A” and molecule “B”. The parameters for the molecular model are shown in Table 3, and described as follows: the corresponding harmonic potentials for bond and angles between the beads within the same chain are $k_T = 100, r_0 = 10 \text{Å}$ and $k_\theta = 200, \theta = 180°$, respectively. In the current study, the axial stretching parameter ($k_T$) is relatively high while the bending stiffness parameter ($k_\theta$) is chosen to be relatively weak, so that the polymer chains are both (approximately) inextensible and highly flexible under equilibrium conditions. To ensure sufficient cross-links, the harmonic potential for the cross-link bonds used $k_T = 200, r_0 = 20 \text{Å}$. For the LJ potential, $\sigma=8.9\text{Å}$, $\varepsilon=100 \text{Kcal/mol}$, and a cutoff of 10Å. It is important to note that in reality the distance of cross-links and monomers are not arbitrarily selected but rather a natural property of a specific material system that is modeled. However for this study the intent is to test these metrics as general as possible, thus the only criteria in the selection of $r_0$ was to avoid steric hindrance.

\textit{Table 3. Intermolecular parameters Values for Bead Spring System}

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Chains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MC1</td>
</tr>
<tr>
<td>Number of Beads</td>
<td>1000</td>
</tr>
<tr>
<td>Bond Stiffness $k_T$ (Kcal/mole-Å²)</td>
<td>100</td>
</tr>
<tr>
<td>Equilibrium Distance $r_0$ (Å)</td>
<td>10</td>
</tr>
<tr>
<td>Angle Stiffness $k_\theta$ (Kcal/mole-rad²)</td>
<td>200</td>
</tr>
<tr>
<td>Equilibrium Angle $\theta_0$ (degrees)</td>
<td>180</td>
</tr>
<tr>
<td>LJ $\varepsilon$ Chains A-B (Kcal/mole)</td>
<td>100</td>
</tr>
<tr>
<td>LJ $\sigma$ Chains A-B (Å)</td>
<td>8.9</td>
</tr>
<tr>
<td>LJ Cutoff Distance (Å)</td>
<td>10</td>
</tr>
</tbody>
</table>
Figure 26: Schematic of Coarse-grain model (a) Close-up of bead-spring model for molecular chains A and B with simple harmonic stretching $(r)$, bending $(\theta)$ and intermolecular repulsive terms (L-J), described by Eq.(55)-(56) and Eq.(21). Also shown are the discrete cross-links between molecules. (b) 2000 bead-spring system (MC1) with 251 cross links. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

4.2 Simulation Protocol

A series of coarse-grain simulations for two scenarios of the molecular chains were implemented by classical molecular dynamics (MD), as described in Chapter 2. The two cases investigated are:
1. a molecular complex whose pair of chains have a length of 1000 nm (i.e., 1000 beads), hereon designated “MC1”, and;

2. a complex whose chains have twice the length, 2000 nm (i.e., 2000 beads), designated “MC2”, allowing the comparison of cross-link density, rather than a finite number of cross-links.

The individual chains are then cross-linked into a single molecular complex through the addition of regularly spaced cross-link bonds. To assess the cooperativity as a function of cross-linking, for MC1 the number of cross-links was changed in the following sequence: 3, 5, 8, 10, 21, 30, 40, 50, 100, 143, 200 and 251 (and doubled for MC2), this odd numbering of cross-links was necessary to keep them uniformly spaced. The initial state of the composite was obtained randomly from a fully cross-linked scenario, and cross-links where then removed accordingly. This was done to remove fluctuations due to initial conditions. Example simulation snapshots depicting dynamic structural configurations are shown in Figure 27.

![Figure 27: Example of time evolution. Shows the dynamic evolution of the system as the simulation runs, notice that the configuration changes from a rather spherical one where the atoms are packed to a configuration that favors one direction, Section 3.4.](image)
All simulations are subject to a canonical (NVT) ensemble, carried out at a temperature of 300K to simulate ambient equilibration temperature, thereby approximating it to “standard” temperature conditions. Temperature fluctuation is controlled using a Nose-Hoover thermostat, with a damping parameter of 100 fs, resulting in a variation of approximately ±15K during simulation. The time step is chosen to be on the order of femtoseconds (2 x 10\(^{-15}\) s), a total simulation time of 10ns. The magnitude of the time step chosen ensures the stability of the simulations. All MD simulations are performed using the massively parallel modeling code LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) [47].

Due to the stochastic nature of the equilibrium processes, it is understood that there needs to be a statistical framework to ascertain the robustness of the metrics. This would include variation in chain length, bond stiffness, angle stiffness, cross-link stiffness, variation in temperature, initial conditions, etc. However the complexities arising from a rigorous statistical study would go beyond the scope of the current study and are left for future work. That being said, to explore the potential variation, a small sample of the space was surveyed by testing the metrics for the 5 and 40 cross-link scenarios of MC1. These two particular cases, which are representative cases for uncooperativity and cooperativity respectively, were simulated for a period of five times longer than the other cases. Taking advantage of the ergodic nature, see Section 2.2, the simulation runs were split into five and assumed to be distinct from one another for statistical sampling purposes.

\[11\] We note that temperature definition for coarse-grain models is not universally accepted, as internal degrees of freedom are condensed to larger-scale vibrations. However, the temperature here is applied to facilitate dynamic evolution of the system and achieve variation in the sampling space. Indeed, the nature of the harmonic bonds and cross-links prevent any thermal breakdown or transitions.
4.3 Results and Discussion

Due to the complexity involved with the possible configuration/sample space of a random polymer, it is important to clarify that the intent here is not explore the exact trends in the defined metrics as a function of cross-linking. The sheer number of sequence possibilities for proteins and polymers, molecular masses, environmental influences, and cross-linking behaviors in general make such trends model/system dependent. From this initial study, through a small number of simulations the objective is to observe any clear transitions from “uncooperative” to “cooperative”, using the metrics defined in Chapter 3 for size (slip), shape (differential anisotropy), and orientation (skew). The goal here is to explore whether such metrics are consistent in defining cooperativity objectively.

4.3.1 Slip and Norm

As expected by the definition and depicted in Figure 28, the results of using Eq. (47) show that slip between the two chains diminishes as the number of cross links increases – e.g., the differences between the gyration of the two molecules decreases. The critical point being 21 cross-links for MC1 and 42 (double) for MC2, representing a “free length” of approximately 50 nm for both models. However it is necessary to clarify that this critical point is a product of the selected scenarios, since the whole possible combinations of number of uniformly spaced cross-links was not evaluated, it is plausible that the actual bifurcation in cooperativity occur at a lower number of uniform spaced cross-links. Cases with to 3, 5, 8, and 10 cross-links are at least three orders of magnitudes larger in comparison with the others, indicating uncoupled behavior. Figure 28 shows how the initial slip value drops from a high of $2.84 \times 10^5$ and $4.68 \times 10^5$, respectively to lows of approximately $5 \times 10^2$ for both which represents a drop of at least 500% in the size difference. The max slip for MC1 corresponds to the 5-cross-link scenario and when normalized to the gyration value of molecule “A”, it represents a 37% of the magnitude. In turn, the relative
value of the slip for the 251 cross-link case represents only a 1.04% difference. The same analysis for MC2 yields a relative slip of 3 cross-link scenario of 42% difference, whereas the 251 cross-link results in a 0.0235% difference. This diminishing size difference is drastically reduced once the number of cross-links reaches 21 and 42 for MC1 and MC2 respectively, and the slip is relative constant in terms of magnitude, regardless of 100 or 200 cross-links.

As mentioned before, these quantities are time dependent, and by looking at the norm, we can summarize each quantity into a single number, shown in Figure 28. It is observed that despite the randomness in the system the trend shows that with respect to the number of cross-links the macromolecules can be considered to be cooperating once the molecular size of both MC1 and MC2 reach 21/42 cross-link cases. When looked at it from the perspective of unlinked “free” length, Figure 28, the trend remains fairly consistent – the critical free length is approximately 50 nm, or a cross-link density of 0.02 cross-links/nm.

The statistical sample from MC1 for the 5 cross-link scenario resulted in an average norm of $3.0 \times 10^6$ with a deviation of $\pm 1.4 \times 10^6$, indicating a consistency between measures. As shown previously after the 21 cross-link scenario the slip difference decreases orders of magnitude and the sample of the 40 cross-link scenario confirms that by showing that the average norm is $0.19 \times 10^6$ with a deviation of $\pm 0.065 \times 10^6$, indicating less variance when “cooperativity” is attained.

To corroborate the slip, the results obtained by Eq. (49) also confirm that, for these particular cases, 21 and 42 cross-links are already sufficient to ensure small size differences which strongly suggest cooperative behavior. Figure 29 illustrates this agreement, where the distance between gyrations (through the relative norm) reaches a maxima of 1.2 and 0.66 (MC1 and MC2 respectively) and descends steadily until it reaches 21/42 cross-links case, at which point the sharp
drop approaches zero, making these cases the critical number of cross-links to reduce the size difference between the molecules.

Figure 28: Slip results: (a) Magnitude of slip v. time for MC1. The computed size difference diminishes as the number of cross-links increase, one can note that past 21 cross-links the size difference is approximately zero, (b) Magnitude of slip v. time for MC2. Just like the MC1 counterpart the molecules size are relatively zero once they get past 42 cross-link scenario. (c) Euclidean norm of slip MC1 (blue) and MC2 (red)
with respect to cross-links and unlinked length (inset). Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

![Graph](image)

Figure 29: Norm of Gyration Tensor results. (a) Gyration norm MC1. The size difference between each cross-link case reduces from a high of 1.2 to approximately zero, the variation between the first cases (3-10 cross-links) is evident. It is shown that past the 21 cross-link cases the gyration norm approaches zero. (b) Gyration norm MC2. The trend from MC1 is repeated here to a lesser extent, the range of 10-20 cross-link shows more stability but one can definitely point that past 42 cross-link the size difference is approximately zero, thus suggesting cooperativity. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

### 4.3.2 Shape Metrics

An example of what represents uncooperative shape is shown in Figure 31, snapshot of the MC1 case at 4ns is presented. The 3-cross-link case presents a visible disparity in the arrangement of
the chains that is not only valid from a shape perspective but also from the size and orientation point of view, where the fluctuation between chains are independent from each other. On the other hand the 100-cross-link case shows a closely and firmly integrated motion between the pair of chains.

Figure 30: Snapshot of molecular chains at 4 ns. a) The 3-cross-link scenario clearly shows the inharmonious motions between the chains, the independence between the chain motions affect the shape metric as well as the size and orientation metrics. b) The 100-cross-link case shows a fully integrated motion, where each pair of beads between the chains are displacing in the same direction, thus the particle configuration is similar between the chains.

This behavior is translated in the observation of asphericity, shown in Figure 31, the 3-cross-link case, the red chains, represents a molecular complex with chains of drastically different particle distributions, looking at the snapshot presented in Figure 30, the asphericity difference between the chains amounts to 80%, a significant gap. The 100-cross-link case, represented as blue chains, indicates the chains share a very similar configuration at each time step which is physically presented in the previous figure.
Figure 31: Example of relative shape anisotropy evolution for MC1. Here, \( \kappa_A^2(t) \) and \( \kappa_B^2(t) \) for MC1 are plotted for both 3-cross-link (red) and 100-cross-link (blue) cases, to exemplify the differences. The wider the “gap” or separation between the two relative shape anisotropies indicates their geometric configuration differ at the same instances. The 100-cross-link case is interpreted as cooperative because its beads are distributing similarly on both chains. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.

The relative shape anisotropies (\( \kappa_A^2(t) \) and \( \kappa_B^2(t) \)) for all cross-link cases of MC1 are shown in Figure 32. The “gap” between each plotted pair represents the differential anisotropy (\( \delta\kappa(t) \)) Eq.(50), representing the differences in particle configuration. For the 3 cross-link case is around 0.2 and, again, it decreases steadily until the 21 cross-link case where the gap is marginal, a similar trend is observed for MC2.
Figure 32: Shape and differential anisotropy results. (a) Relative shape anisotropy for MC1 as a function of time and number of cross-links. The gap for the 3 cross-link case is around 0.2 and it decreases steadily until the 21 cross-link case where the gap approaches zero. (b) Relative shape anisotropy MC2 as a function of time and number of cross-links. (c) Molecular shape for both MC1 (blue) and MC2 (red), as a function of number of cross-links and unlinked length. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.
Figure 32, illustrates the difference between the two chain systems, the initial relative shape anisotropy has a value of 0.2 for MC1, indicating a symmetrical, possibly spherical initial shape, whereas the initial shape anisotropy value for MC2 is 0.8 which resembles a more extended/linear configuration. While MC1 displays relatively oscillatory behavior (relative extension and compaction), for MC2, it is observed that the overall shape fluctuations remain within a band that seems to increase with time – the longer molecules are relative extended initially, and begin to collapse to a more compact configuration. However the initial value for each composite is a non-factor and arbitrary, because as stated previously, a random but constant initialization point is chosen for each simulation. The concern here is not a particular molecular structure or behavior, but only the coupled behavior of the molecules over time. It is also important to clarify that the idea behind the shape anisotropy does not imply cooperativity if they are symmetrical, but rather emphasize that the molecules components must be similarly distributed. In simpler terms, it does not matter if the shape anisotropy is 0 or 1, what matters is that at the same instance both have similar values.

Again, statistically, a decrease in variation with coupling is observed. The sample from MC1 for the 5 cross-link scenario resulted in an average norm of 3.14 with a deviation of ±1.0923. As shown previously after the 21 cross-link scenario the molecular shape decreases several orders of magnitude and the sample of the 40 cross-link scenario confirms that by showing that the average norm is 0.11 with a deviation from the mean of ±0.05.

Finally Figure 32 shows the molecular shape Eq. (51), where similar to the previously discussed molecular size, and how the measurement presents a distinct transition at the fourth case which once more belongs to the 21/42 cross-link simulations. This clearly indicates that as the number of cross-links is increased the particle configuration changes are congruent between
the components of the macromolecule – further indicating cooperativity and a consistency between size and shape metrics.

4.3.3 Orientation Metrics

Completing the trifecta of the geometrical metrics, is the molecular orientation. The skew Eq. (53), of the macromolecule is shown in Figure 33. For the particular case of the MC1, the maximum misorientation as 30 degrees which belongs to the lowest 3-cross-link case. As the number of cross-links increases, the chains gradually align themselves, once more reaching a critical point at 21 cross-links where the misorientation angle suddenly drops to near zero values. As with the previous metrics, the skew trend remains in the evaluation of MC2, but the total variability between maximum and minimum orientation is only 6 degrees. This can be associated with the relatively extended configurations of the molecules, indicated by the relative shape anisotropies see Figure 32, and thus a restriction of change in molecular orientation. Regardless, in spite of such a relatively small mismatch in orientation, there is a clear decrease in the skew value with increasing cross-links. These results are encouraging because it suggests limited restriction in scale – the skew indicates relative cooperativity, regardless of a large mismatched orientation (30 degrees for MC1) or small mismatch (6 degrees for MC2).
Figure 33: Orientation results. (a) Skew for MC1. The maximum angle of misorientation is 30 degrees, and similar to results from the previous metrics the chains gradually align themselves as the number of cross-links increases, reaching the same critical point at 21 cross-links. (b) Skew for MC2. The trend is similar to the one obtained from MC1, but like previous results the orientation for MC2 has less variability. (c) Norm of skewness, the molecular orientation, MC1 (blue) and MC2 (red) in terms of number of cross-links and unlinked length. The results confirm, like the previous metrics, that if this particular model has at least 21 cross-links then the misalignment of both chains approaches zero. Adapted from Kwan and Cranford [35]. Used with permission © 2014 ASCE.
The molecular orientation, Figure 33, gives a direct comparison of how the angle difference between both chains of the same composite decreases as the number of cross-link increases. We see that skew is consistent with the previous metrics, again indicating that after 21/42 cross-links, there is not much of a difference in terms of the orientation. Once again, the statistical sample from MC1 for the 5 cross-link scenario was consistent, resulting in an average norm of 340.20 with a deviation of ±88.32. As shown previously after the 21 cross-link scenario the skewness decreases several orders of magnitude and the sample of the 40 cross-link scenario confirms that by showing that the average norm is 12.82 with a deviation of ±12.68.

4.3.4 Closer look at cross-link density

The results from the gyration metrics indicate that the tipping point between cooperative and uncooperative behavior corresponds to a cross-link density of 50nm. The peculiarity of this number certainly requires some insight based on factors such as temperature, cross-link strength and other specific properties of the system. Based on the studies by Cranford [30], cooperativity can be defined as a limit, where the relationship between molecular rigidity (D), cross-link density (p) and cross-link stiffness (k) are taken into account and is defined as

$$\frac{2k_BT}{p} < D < \frac{k}{6p^3}$$

(57)

where $k_BT = 0.593$ kcal/mol for $T=300K$. Based on the system properties shown in Section 4.1, the corresponding cross-link density for the lower limit is 168 nm, which is one order of magnitude greater than the 50nm result. However a closer inspection notes that the definition of temperature for coarse-grain models is controversial, thus a one to one comparison is not valid. A back calculation suggests that for the molecular rigidity the actual temperature of each particle
that is condensed as a bead is in the region of 1000 K – e.g., the motion of the coarse-grain chains set at 300K is similar to a full atomistic system at 1000K, due to the lumped degrees of freedom.

### 4.4 Summary

The cooperativity metrics have shown the ability to identify when the system is deforming together, from a cross-link perspective that number was 21/42 for MC1 and MC2, respectively. However it is important to specify that these metrics represent geometric/mechanical cooperativity and must be used together. The perspective of cooperativity from a geometric standpoint is essential because systems can be of different size and shapes but still be cooperative from a chemical perspective which is the case of ligands and receptors. For this case where the cooperativity is deemed from a mechanical point of view, a simple thought experiment can be evaluated to justify the usage of the three metrics. A simple mathematics examination reveals that gyration tensors for different cooperative structures might have the same trace Eq.(42), but the components of the sum might be different, thus leading one to believe that the system is cooperating because the slip is the same. However, this is not the same for the shape metric, while the slip might be the same, since the asphericity depends on the deviatoric component, where the components of the sum are taken into account. From a physical point of view, one can certainly envision cases in which cooperativity may be suggested by a single metric, but in fact, the molecules are in discord. In Figure 34(a), for example, presents a pair of identical carbon nanotubes in a vacuum. The radius of gyration and anisotropy would be the same for both, and consequently the slip will be zero indicating no size change. However, it is clear that the carbon nanotubes could have different orientations.
Figure 34: Examples of non-cooperative single metrics. These examples stress the importance of evaluating cooperativity by using all three metrics. (a) A pair of identical carbon nanotubes, each carbon nanotube has the same configuration and radius of gyration, thus if one only looks at the slip and the differential anisotropy it would show cooperativity, however they are clearly oriented in the wrong direction. Thus the skewness metric would show incongruent behavior. (b) A system where a pair of axially deformable bars undergo opposite deformations, for this case the skewness metric would clearly show no misorientation, however since the loading conditions are opposite the deformation of both bars will clearly show a difference in the radius of gyration, thus making this another example illustrating the lack of validity of each individual metric as a single criteria for cooperativity.

A second example is shown as a pair of axially deformable bars (Figure 32(b)), where one is under tension and the other under compression. While the slip is certainly different and variable, the orientation remains the same, which, independently, may suggested nonexistent cooperativity. These examples, among others, show the inability of using an individual metric to gauge cooperativity. Encouraged by the promising results on the coarse-grained ideal molecular chains, we proceed to test the cooperativity metrics on the tropocollagen molecule as a continuing effort of validating the metrics on different physical systems.
5. COOPERATIVITY IN COLLAGEN MOLECULE

Collagen is the most abundant protein in found in mammals, they are the basic components of tissues such as tendon, and bone, and thus they are the foundation that provides organisms with mechanical stability, elasticity and strength [40, 98-103]. Similar to other biological materials, collagen has captured much of the focus of bio-inspired material science for its complex hierarchical structure serving as a perfect example of the building block problem (see Section 1.1.1).

At the lower hierarchical level there are tropocollagen molecules, these are arranged in a staggered configuration with an observable periodicity known as the D-band, where D= 67 nm. The molecules have a gap between them which is known as the gap region. Five molecules bundled together form a micro fibrils which are bundled together to form collagen fibrils, where multiple fibers make up the collagen fiber [104-107]. It is also labeled as a smart material [108] due to its ability to change properties under mechanical forces, resulting in a rearrangement of structure, changes made at the molecular level permeate into the higher scales resulting in different mechanical and biological functions. In Nature some changes are inadvertent and result in loss of function, mutations in the tropocollagen molecule sequence have been shown to be responsible for diseases with detrimental effects, one of these diseases is osteogenesis imperfecta (OI), or brittle bone disease. The repercussions of OI includes the loss of bone ductility, stunted growth, weak tendons, among other symptoms [109].

The collagen molecule has been the subject of numerous studies [106, 110-115]and it is known that point mutations result in local unfolding and potentially a kink in the molecular structure. The objective of this study is to discern the local loss of triple helicity through the use of the cooperativity metrics, to ultimately reflect the severity and understand how the lack of
cooperativity can be related to the severity of the mutation that depends on the replacing amino acid. The idea is to do a comparative analysis of the geometric features of a healthy collagen molecules versus those of the mutated version, where the mutated version gyration based metrics show a deviation from the healthy molecule. The selection of this ideal case is *solely based on the previous understanding of the physical effect that is displayed as a result of the mutation process in the collagen molecule.*

### 5.1 Tropocollagen Molecule

Currently 28 types of collagen have been identified. Type I collagen is the most abundant in the human body and it is present in bones, tendon, skin, teeth and cornea. Deficiencies in Type I collagen are the source of most of the collagen related diseases. The protein itself is structured from a hierarchical assembly of tropocollagen molecules, **Figure 35,** that are arranged into fibrils which are the building blocks for larger-scale fibers. Collagen molecules are the fundamental building blocks, the changes in the composition of these molecules result in changes at the macro scale where the mechanical properties and biological processes are affected.
The collagen polypeptide consist of three branched polymer-like chains that form a triple helical structure, see Figure 36. Each chain of collagen is composed of amino acids that are arranged in a repeating sequence \((\text{GLY-X-Y})_n\), where GLY refers to the residue Glycine, and the X,Y positions are usually occupied by proline and hydroxyproline [106, 107]. Glycine is essential to the structure because its side chain, a hydrogen bond, is the only thing that can fit in the closely packed molecule. Hydrogen bonds play a vital role of holding the three chains together. The specific form of \(\text{GLY-PRO-HYP}\) is the most stable form of collagen[117], as well as the most observed sequence in fibrillary type collagen. A normal type I collagen molecule consists of two alpha-1 chains and one alpha-2 chain. It has an approximate length of 300 nm with a radius of 1.6nm. The OI mutation is primarily caused by the replacement of Glycine in the sequence.
Figure 36: Tropocollagen molecule. (a) Depiction of the three chains and characteristic triple helical structure. (b) stagger of the three strands highlighted by the alpha carbons of a sequence of GLY-PRO-HYP (c) the orthographic view from the top of the triple helix (d) close up of the triple helix, each chain is plotted with one color and the spheres represent the main alpha carbons of the glycine amino acid, the helical structure is stabilized by hydrogen bonds. The molecule is surrounded by explicit water but not shown to improve visual clarity.

5.1.1 Single Point Mutations: OI Mutations

There are over 250 different mutations for OI [118, 119], however in humans a single point mutation (1 amino acid in the entire chain) is typically responsible for brittle bone disease. Currently six mutation types of Type I collagen have been identified and were originally classified by Sillence [114, 120], as severe, moderate and mild conditions, see Table 5. The OI mutation is primarily caused by the replacement (also known as a missense mutation) of the Glycine residue (amino acid) in the \((\text{GLY-X-Y})_n\) sequence. Since Glycine is the smallest amino acid and is located in the interior axis, any replacement results in a kink due to the excess bulk of the larger amino acids. In addition because of the role it plays in maintaining the triple helical behavior, Glycine mutations are more damaging. It is known that mutations/changes in the collagen molecule lead to changes in the collagenous tissue, however the how remains unclear.

Recent studies [121-127] have revealed the negative effects, in particular on the mechanical properties, due to the mutations that resound across the molecular scales. Figure 37
shows results from Gautieri, the single molecules lose around 15% of their stiffness (depending on the replacement amino acid), in addition the intermolecular adhesion is considerably degraded and at the fiber scale a stress-strain test shows a loss over 100% in ultimate stress and a 0.2 difference in ultimate strain. The OI mutations disrupt the triple helical structure of the collagen in the vicinity of the mutation, an unfolding occurs due to the disruption of the hydrogen bonds which can be interpreted as a lack of cooperativity at the mutation site. The severity of the disruption is dependent on the type of residue that replaces the Glycine in the triplet.

**Table 5. OI produced by mutations of Type I Collagen**

<table>
<thead>
<tr>
<th>Type</th>
<th>Deformity Level</th>
<th>Lethality</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Moderate</td>
<td>Nonlethal</td>
</tr>
<tr>
<td>II</td>
<td>Severe</td>
<td>Majority are stillborn, 90% die before 4 weeks</td>
</tr>
<tr>
<td>III</td>
<td>Progressively Deforming</td>
<td>Nonlethal affects children</td>
</tr>
<tr>
<td>IV</td>
<td>Mild to Moderate</td>
<td>Nonlethal</td>
</tr>
<tr>
<td>V</td>
<td>Mild to Moderate</td>
<td>Nonlethal</td>
</tr>
<tr>
<td>VI</td>
<td>Mild to Moderate</td>
<td>Nonlethal</td>
</tr>
</tbody>
</table>
Figure 37: Effects of missense mutation, shows that the replacement of Glycine by other residues affect the mechanical properties. (a) Stiffness properties can reduce up to 15%. (b) Intermolecular adhesion is further reduced. (c) The fibril maximum stress lost is over 100%, and fracture occurs at a lower strain. Adapted from Gautieri et. al. [124].

5.2 Molecular Model Formulation

The molecular model of collagen presented in Gautieri [123] is used. The tropocollagen molecule is built with various sequences, to reflect mutations, using the software THeBuScr (Triple Helical collagen Building Script)[128]. The simplest model of collagen where all three identical chains are composed of the repeating sequence of Glycine-Proline-Hydroxproline also denoted by their three and one letter abbreviations as GLY-PRO-HYP (GPO) respectively. The central triplet is used to introduce the GLY replacement, see Figure 38. The peptide structure is described as \((\text{GPO})_5\text{XPO}(\text{GPO})_4\)_3, where X is the position occupied by Valine one of the possible replacement amino
acids for which an 83% of mutations are classified as severe. Table 6 shows the frequency of severe OI percentages per amino acid replacement. This peptide is a truncated version limited to 30 residues per chain to reduce computational costs, similar studies have been carried out using comparable length models. [129-131]

![Figure 38: Mutation Site. (a) Tropocollagen molecule, the orange spheres indicate the location where residue substitution takes place. (b) Glycine amino acid present in the healthy collagen substituted by Valine making it a mutated collagen. Notice the difference in size between the amino acids, which is one of the reasons for the presence of a kink. All chains are identically sequenced GLY-PRO-HYP triplets.](image)

<table>
<thead>
<tr>
<th>Substitution</th>
<th>total</th>
<th>lethal</th>
<th>nonlethal</th>
<th>%lethal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala</td>
<td>21</td>
<td>4</td>
<td>17</td>
<td>19</td>
</tr>
<tr>
<td>Cys</td>
<td>49</td>
<td>19</td>
<td>30</td>
<td>39</td>
</tr>
<tr>
<td>Asp</td>
<td>22</td>
<td>15</td>
<td>7</td>
<td>68</td>
</tr>
<tr>
<td>Glu</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>50</td>
</tr>
<tr>
<td>Arg</td>
<td>31</td>
<td>17</td>
<td>14</td>
<td>55</td>
</tr>
<tr>
<td>Ser</td>
<td>61</td>
<td>18</td>
<td>43</td>
<td>30</td>
</tr>
<tr>
<td>Val</td>
<td>18</td>
<td>15</td>
<td>3</td>
<td>83</td>
</tr>
</tbody>
</table>

**Table 6. Frequency of Lethality of OI by Substitution. Adapted from Bodian et al.[132]**

**5.3 Simulation Protocol**

A series of equilibrium simulations are implemented by classical molecular dynamics. The protein molecules are solvated in a 20nm x 5 nm x 5nm periodic explicit water box. A three site rigid water
molecule model is used (TIP3P), charges and L-J parameters are assigned to each of the three atoms. The PPPM (Particle-Particle-Particle-Mesh) method is used to account for the long range coulombic interactions. The total system size is 24765 for the healthy collagen and 24807 atoms for the mutated collagen. In the case of charged molecules, counter ions (Cl- or NA+) were added to keep the system neutral. System is initially minimized using the steepest descent algorithm, and the water is slowly equilibrated with an amped up approach to stabilize water molecules, while the atoms of the collagen molecule are restricted. The NVE ensemble is used using a steady increase of temperature (e.g. 1-50-100-200-300K) to avoid potential instabilities in the system. Simulations are then subject to a canonical (NVT) ensemble, carried out at 400K, and is meant to accelerate the process of exploring the solution space, so that one can reduce the simulation time, see Section 2.2.1. Temperature fluctuation is controlled using a Nose-Hoover thermostat, with a damping parameter of 100 fs, resulting in a plus/minus of approximately 15 K during simulation. Afterwards the system temperature is lowered to 300K using the NPT ensemble to control the pressure of the surrounding solvent at an average of 1 atm. The time step is chosen to be on the order of femtoseconds (0.5 x 10^{-15} s) for a total of 4.5 ns. The magnitude of the time step chosen ensures the stability of the simulations All MD simulations are performed using the massively parallel modeling code LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator; http://lammps.sandia.gov/ ) [47], using the CHARMM potential, see Section2.3.2, implemented in the LAMMPS code.

5.4 Results and Discussion

Preceding the discussion of cooperativity, a quick confirmation of the models is presented, the root mean square deviation (RMSD) of atomic positions is used to confirm that the simulated systems are fluctuating about some average position and thus have reached dynamical
equilibrium, see Figure 39. As one can suspect the average distance between atoms of the mutated scenario, in blue, present both a larger mean (7.75 vs. 5.32) and variance (2.47 vs. 1.65) compared to the healthy molecule, presented in red, in accordance to the loss of helicity, see Section 5.1.1. The mutated collagen shows a larger fluctuation in the local average, and could potentially reflect a transient effect of contraction and expansion due to local unfolding of the triple helix.

Figure 39: RMSD of the simulated collagen molecules. The RMSD of distances between atoms is used as a guide to establish whether systems have reached dynamical equilibrium. In red the RMSD of the healthy strand with a mean of 5.32 and a variance of 1.65, lower in comparison to the mutated collagen, in blue, that has a mean of 7.75 and a variance of 2.47, the higher average distance and larger fluctuations are expected as a result of the swelling of the molecule due to the presence of a mutation.

In addition a visual inspection confirms the presence of the kink due to the mutation, Figure 40, the tightly packed triple helix seen in the healthy case shows a maximum expression of disruption.
in the location of the glycine substitution, this disturbance fluctuated at around 3 Å from alpha carbon to alpha carbon atoms, efforts to find a value from literature as an added confirmation were unsuccessful. Encouraged by the visual presence of the kink, let us proceed to the cooperativity discussion.

(a)

(b)

Figure 40: Visual confirmation of the kink in mutated collagen. (a) The healthy collagen molecule is shown, notice the tight packing of the helix at all times. (b) the mutated collagen shows a clear disruption in the triple helical structure, and although subtle one can perceive the looseness at all locations, demonstrating that the model is capturing the intended behavior.

From a fundamental perspective the geometric similarity of the collagen system with the ideal macromolecule model, in the sense that both are essentially pseudo-linear systems, makes the scenario of transferring the cooperativity metrics a favorable one. In this case the cooperative metrics can only be used to discern the differences between the systems, Figure 41, because systems in Nature already have pre-established conditions, in other words the inverse problem of the ideal molecule case presented in Chapter 4.
Figure 41: Pre-established conditions of collagen. Because this is a natural existing material system, the metrics are used here to differentiate healthy and mutated collagen based on the previous knowledge that points to the existence of an unfolding as the result of glycine being replaced.

The implementation of the metrics are slightly modified, due to the obvious difference in the system, two chains vs. three chains, a comparative analysis between repeated pairs of chains is considered. Another concern is the effect of the collagen’s natural movements under a solvated environment, which can affect the resolution of the results. To counter this, two measures are taken into account, first the gyration tensor is computed by only using the backbone atoms of the residue, discarding the dangling behavior of the “branched” atoms, and second the molecule is spliced into different sections and the metrics are computed for each section as if they were individual molecules. For this case the collagen has been spliced into three uniform sections, the middle section containing the mutation, where the first triplet of each end for all three chains is cast-off to avoid any potential boundary effects, Figure 42.
Figure 42: Spliced cooperativity approach. An ≈8nm long, 30 amino acids per chain, mutated collagen is presented, in this case the collagen is spliced into three uniform sections. The chains are identically sequenced as \((GPO)_5-VPO-(GPO)_4\), where V corresponds to the substitution of Glycine, orange beads with Valine, purple beads, which corresponds to the second section. The molecule is shown surrounded by TIP3P water molecules that are represented as a transparent blue surface.

Conversely the implementation of these changes were unsatisfactory, the preliminary results of the comparison based on the per chain section analysis was unable to differentiate the mutated section from the healthy section. Figure 43 shows no clear distinction in the slip between the chains, and similar results can be obtained from the other metrics. However if one observes the expansion of the molecule in the mutated case you can pose the question, are the individual chains really not cooperating? If one deconstructs each chain of the mutated collagen shown in Figure 40, and folds the three chains together then you can see that, in fact, from the perspective of the metrics, the slip would be minimal, the shape anisotropy wouldn’t show visible differences since the “hump” presented by the unfolding is relatively similar for all three chains and the misorientation is minimum, while the direction of gyrations might be different the magnitudes are all relatively the same. Thus, it seems that the metrics are indeed suggesting that the mutated collagen molecule is cooperative. Given this, it is clear that the definition of structure for the metrics requires modification, see Section 3.6, as a consequence the cooperativity “structure” is defined as the bundle of the three chains, and the gyration tensor is extracted from each section as a whole, the comparison is then made by looking at the similarities between the same section of the healthy and mutated collagen. With these changes taken care of, the focus here is then to
employ the gyration based metrics to attempt to identify the lack of cooperativity produced by the single point mutation in a collagen molecule.

Figure 43: Per chain analysis – Radius of Gyration. (a) Radius of gyration of each individual chain located at the section of mutation for the healthy collagen, the second chain presents a minor lack of cooperativity with the rest of the chains. (b) Corresponding section of the mutated collagen, the metrics reflect a relative harmonious behavior at all chains, a conflicting display of what the mutation is physically producing in the molecular behavior.

5.4.1 Size Metric – Slip

Utilizing the slip metric, Eq.(47), in a sectional fashion, where the size difference is measured for each of the three sections between the healthy and mutated collagen, the results show the discrepancy in the extension of the system in the section that contains the mutation. As depicted in Figure 44, the blue plot, corresponding to section 2, the size difference fluctuates around 6.97 Å, while sections 1 and 3 have a mean of approximately 2.20 Å and 2.29 Å correspondingly. This relative gap in size represents an eccentricity of approximately 4.67 Å, which exceeds the size difference of the other sections by roughly 200%. The size gap conveniently ties into the previous knowledge indicating the bulge of the collagen molecules as a product of the loss of helicity due to the mutation in OI. From a statistical perspective the data presented by the slip metric has at most a standard deviation of 1.4, meaning that the variance presented is still considerably low if
one looks at the absolute difference between the means, thus further serving as an added confidence in the results.

Figure 44: Section analysis of the slip in the collagen molecule. The slip metric is taken as the relative size difference between the healthy and mutated collagen for each section. The blue plot corresponds to section 2, which contains the mutation, has a relative size difference that varies with a mean of \( \approx 6.97 \), almost a 200% difference with respect to the other relative size differences. The gap in relative size presented is an indication of the swelling of the molecules due to the disruption caused by a single point mutation on the helical behavior of the molecule.

5.4.2 Shape Metric – Anisotropy

For the shape configuration, the relative shape anisotropy Eq.(50), is inspected. Figure 45 shows the individual anisotropy configurations for both section 2 and section 3. For section 2 the separation between the particle arrangement at the same time instance is clear, where the decline in anisotropy amounts to nearly 15%, recall that a anisotropy value that tends to zero indicates a spherical symmetrical conformation, thus the drop reflects the trend of the mutation.
in furthering the structure from a linear configuration, i.e. a swelling behavior in all directions, see Figure 40. In contrast, section 3 shows no discernible difference, in fact both (healthy and mutated) contain similar statistical parameters (mean ≈0.47 and standard deviation ≈0.026), the same outcome is observed for section 1.

Figure 45: Shape configuration between sections. (a) The shape difference presented for section 2, where the mutation lies, the difference between shape configurations in time shows a drop-off of roughly 15%, indicating a clear gap in the way the atoms are configured within the mutated section. (b) the shape difference in section 3, where the difference between the mutated and healthy collagen section 3 are unnoticeable, the mutated collagen shadows the healthy collagen section, both have a mean of 0.47, and is for all intents and purposes a mirror image of the shape configurations for section 1.
In agreement to the individual anisotropy, the relative shape anisotropy reflects the relative difference in particle conformations, see Figure 46 the blue plot has a mean of 0.147 with a standard deviation of 0.03, compared to the mean of 0.03 and standard deviation of 0.02 for sections 1 and 3. The margin of difference is near 12%, a considerable amount if one takes into account that the shape anisotropy value ranges from 0 to 1. In addition the small variance indicates the quality of the data generated by the metric, where there are no overlap between standard deviations and means, showing a clear gap between the comparisons. As was the case for the slip metric shown in the previous section, the differential anisotropy is also able to correctly identify the section where the mutation is located.

Figure 46: Section analysis of differential shape in collagen molecule. The differential anisotropy metric shows a clear-cut margin between the mutation containing section 2 and the other sections. Noting that the anisotropy is contained between 0 and 1, the margin between the sections is ≈12%, the maximum standard deviation also belongs
to section 2 and amounts to roughly 0.03, which is within one order of magnitude smaller than the mean, the relative minor variability only reinforces the confidence of the results.

5.4.3 Orientation Metric – Skew

The final leg of the cooperativity triangle metric is evaluated, the sectional analysis of the misorientation in the mutated collagen further agrees with the results from the previously discussed size and shape metrics. Figure 47, shows an interesting trend, it would seem that the natural orientation of gyration of the collagen is approximately 1 degree. In the case of the blue plot corresponding to the section of interest one can see an insignificant, in magnitude, descent of the angle, however the relative statistics show the undeniable fluctuations of the angle in the presence of the mutation.

Figure 47: Section analysis of orientation in the collagen molecule. Skew results, due to the tight packing of the triple helical structure, seem to indicate that collagen has a
natural misorientation of ≈1 degree. In this case the mutated section shows some minor fluctuations, the mean ≈0.995 a reduction of approximately 0.5%. All three metrics seem to indicate local uncooperative behavior in section 3.

For the relative statistics the moving average and moving standard deviations are observed, Figure 48 and Figure 49, while the mean and standard deviation reflect the statistic of the system of a whole, in this case because the magnitude is insignificant the focus is on the transient of the statistics. Using a window size of 100 data points and the central method, one can cleanly observe the effects of the mutation on the rate of fluctuations of orientation in the system, the moving standard deviation in particular shows that the flux in angle change exceeds by 300%, the variability exhibited in the sections that do not contain the mutation. The usage of the moving average basically acts as a filter and results in a cleaner look of Figure 47.

![Figure 48: Moving average of orientation metric.](image)

**Figure 48**: Moving average of orientation metric. The results are filtered through a moving average, each point is the average of 100 data points, and is used to convey a cleaner image where the swing in orientation is shown to be chaotic for the section that contains the mutation.
5.5 Summary

The cooperativity metrics have successfully identified the local unfolding of the triple helix in the mutated collagen molecule. While changes needed to be made to the definition of the cooperativity structure, (e.g. taking the bundle of collagen as the structure of interest), the results are recognizing section 2 as the section containing the mutation consistently across the metrics. In this case cooperativity was assessed based on the relative aspect of the geometrical conformations, where the comparison takes place on the sections, of each mutated and healthy collagen, and looking for a disruption in the trends. So while sections 1 and 3 were consistently behaving within similar statistical descriptors, section 2 presented a perceptible variability. Concerning the windowing approach, a quick inspection of taking the whole 8nm as a section...
proved fruitful, Figure 50 in particular shows the relative difference in slip between the collagen molecules and similarly distinctions are present in the other two metrics. Yet a detailed description of comparing larger molecules was currently out of the possible time frame. Indeed this result is encouraging because it would seem to imply that larger sections of molecule gyrations can be taken into account, thus increasing the usability of the metrics. With the metrics success at hand, ultimately one would like to use it to correlate the results based on the severity of the mutation as presented in Gautieri[124]. Furthermore, while the focus was the imperative observation on whether the mutation could be identified or not, one wonders whether one can exploit the results presented to explain how mutations affect the hierarchical performance, e.g. the orientation results show a significant variation of the variability in the skewness of the gyration. With the cooperativity metrics have been validated for an existing physical system, the next step is to attempt to use cooperativity metrics as a tool for design aid. In the following chapter a proposed work is briefly detailed with the objective of looking at the cooperativity between carbon nanotube bundles and collagen molecules.

Figure 50: Slip of collagen molecule. For this case the slip metric is used on the whole collagen (8nm) molecule. The results are encouraging showing a distinct difference, a direct implication is the affordability of using less and larger sections to identify mutations on a complete 300nm strand of collagen molecule.
6. PROPOSED COLLAGEN-CARBON NANOTUBE INTERACTION

While the ideal molecule (Chapter 4) and collagen (Chapter 5) studies demonstrated that cooperativity metrics can offer insight into coupled molecular systems, the next step would be to apply the metrics as a parameter to help facilitate materials system design and characterization. An interesting emerging system where cooperativity efforts can play a role are in potential carbon nanotube (CNT) – collagen composites.12 As part of immediate future work, the interaction between CNT and collagen will be investigated. Here, we outline some key future modeling work.

The advancement of nanotechnology has without a doubt made carbon allotropes one of the prominent areas in the materials science field. Their inherent molecular stability, in combination with promising mechanical, electrical and thermal properties [133-144] have certainly merited the continuing focus. Carbon nanotubes (or CNTs; see Figure 51) in particular are an appealing material given its superior mechanical properties, with many proposed applications [22, 145].

Figure 51: Schematic of a single-wall carbon nanotube (SWCNT). (a) top view (b) trimetric view. Notice the characteristic honeycomb pattern. A carbon nanotube is essentially a rolled up graphene sheet.

12 A system currently being explored by Prof. M. Minus, Department of Mechanical and Industrial Engineering, at Northeastern University
The field of bio-nanotechnology has explored the possibility of incorporating carbon nanotubes substrates for cell culture[146], drug delivery systems [147-150], bio-sensing [151-156] and as a component to biomaterials [157, 158]. In the area of tissue engineering, considerable effort is being made into incorporating CNTs to reinforce scaffolding materials. These scaffolding materials are being designed to mimic the natural extracellular matrix (ECM) which is responsible of signaling the tissue cells so that tissue regeneration can take place. Natural ECM is composed of collagen, adhesion proteins, signaling molecules and proteoglycans, thus an artificial ECM should be as similar as possible to the natural ECM to activate the tissue regeneration mechanism. In addition to the potential on increased mechanical strength as reinforcements for artificial ECMs, the composite shows other benefits such as being minimally invasive, biocompatible, and biodegradable. As an added bonus, the CNTs electrical properties are also being studied in specific tissue regenerating cases where electrical signaling is important, e.g. nerve and heart.

While designed ECM as a platform for tissue growth is an ongoing challenge, a different perspective on tissue engineering is also being considered: instead of recreating the ECM to heal the tissue, why not directly create the tissue itself? In other words, rather than designing ECM to grow osteogenic, myogenic or neurogenic tissues, can the tissue itself be produced directly? Currently, the ability to synthetically reproduce an actual collagen fiber (as a precursor to tendon, for example) has eluded the field [21]. Current efforts are being made to use carbon nanotubes as a template to reproduce the characteristic D-band of collagen fibers (structure described in Chapter 5). As such, the cooperativity between collagen and carbon nanotube interaction is fundamental concern if one wishes to understand the makings of the fiber. (see Figure 52).
Figure 52: CNT-Collagen composite sketch. (a) Disordered bundles of collagen and CNTs (b) using CNTs as templates the intention is recreate a D-banded collagen fiber.

Surprisingly, in spite of the numerous studies focused on CNT-collagen [159] composite in tissue engineering, research involving the atomistic description of the material system, Figure 53, has not been exhaustive, and promising research can be done [160, 161]. In contrast to the collagen system in Chapter 5, where it is widely known that a mutation produces local uncooperativity, the CNT-collagen composite is poorly understood. Thus fundamental studies that describe the interaction properties between CNT and collagen need to be carried out before cooperativity can be assessed, in conjunction with ongoing experimental efforts.

Figure 53: Schematic of a single-wall CNT-collagen composite. (a) orthographic view (b) side view. For the collagen molecule the main alpha carbons are represented by beads.

The contribution of this work would be to integrate cooperativity as a potential metric to gauge successful integration of a CNT-collagen system. The objectives of the future work are outlined as:

Molecular Cooperativity/Compatibility

K. Kwan, 2015
I. Employ suite of *in-silico* tests to study the mechanical properties (i.e. Young’s Modulus, tensile strength, bending strength) of composite systems, to serve as a basis.

II. An assessment of atomistic interactions between carbon nanotube and collagen in different geometric orientations.

III. Assess cooperativity of carbon nanotube bundles with collagen fibrils using the developed metrics and correlate with the results from (I) and (II).

IV. Validate with ongoing experimental efforts.

The objective of this proposed work is to analyze the properties of the CNT-collagen composite, as well as characterize the interactions that take place between the components. This information will allow for the interpretation of the physical meaning of compatibility, thus facilitating results and interpretation of the cooperativity metrics. The scope of the proposed work is beyond the current initial development of cooperativity metrics, and should proceed in tandem with synthetic/experimental efforts. However, we proceed to outline the potential modeling approaches that may be implemented.

### 6.1 Mechanical Properties Characterization of CNT-Collagen

A literary review search resulted in no fundamental studies on the mechanical characterization of this composite. Note that while the cooperativity metrics (*e.g.*, Chapter 3) are intentionally formulated in a manner that is *independent* of specific system properties, the lack of information on how the system interacts requires that some effort be taken to gain insight. Basic characterization is necessary. To start a suite of two *in silico* tests may be applied: first, a tensile test to approximate a properties such as stiffness (*e.g.*, Young’s modulus), tensile strength, and shear transfer, and; second, characterizing bending by imposing the curvature, to determine
potential unified deformation of CNT-collagen. We note that these characterization techniques have been applied in a recent investigation of diamond nanothreads [162].

### 6.1.1 Tensile Tests

A *quasi*-static approach can be used to determine strength and stiffness. A suitable constant strain increment can be applied to approximate molecular mechanics and avoid any temperature effects or rate dependence on the determination of the limit states. This facilitated comparison to experimental results (which typically have slow loading rates). During simulation, quantities such as virial stress and potential energy can be recorded, which are difficult to measure empirically. The virial stress is commonly used to relate to the macroscopic (continuum) stress in molecular dynamics computations [163, 164]. The virial components, $\tau_{ij}$ in a representative volume $\Omega$, are:

\[
\tau_{ij} = \sum_{a \in \Omega} \sum_{b \in \Omega} \left\{ \frac{1}{2} \left( r_{i}^{(a)} - r_{i}^{(b)} \right) F_{j}^{(ab)} \right\}
\]

which generates the six components of the symmetric stress-volume tensor, where, $r_{i}^{(a)} - r_{i}^{(b)}$ denotes the distance between particle ‘a’ and atom ‘b’ along the ith vector component, while $F_{j}^{(ab)}$ is the force on particle ‘a’ exerted by particle ‘b’ along the jth vector component. Note that, if the system is minimized and temperature free, the velocity component associated with the virial can be neglected. The total stress can be calculated as $\sigma_{ij} = \tau_{ij} \Omega^{-1} W$ here $\Omega$ is taken a distance that is free from boundary effects. The model should be large enough to approach bulk behavior and enable stress/strain results to be taken from an interior section where the stress/strain field is relatively homogeneous, yet small enough for computational efficiency.
With this the Young’s modulus, tensile and shear strengths can be computed, and delineated between CNT and collagen components. The next setup looks at the bending properties of the composite.

### 6.1.2 Bending Test

To bend a composite, curvature can be imposed indirectly by adhering the system to an idealized surface, and then imposing curvature to the surface. Since the adhesion energy between composite and surface is constant, the only change in the system energy will be due to the bending energy of the composite. The surface can be potentially modeled by a Lennard-Jones 9:3 potential, where:

\[
\phi(r) = \varepsilon \left[ \frac{2}{15} \left( \frac{\sigma}{r} \right)^9 - \left( \frac{\sigma}{r} \right)^3 \right]
\]

(59)

The Lennard-Jones 9:3 potential is derived by integrating over a 3D half-lattice of Lennard-Jones 12:6 particles, effectively representing a semi-infinite LJ surface. The strain energy can then be minimized with respect to the curvature, yielding the change in bending energy. In particular, system configurations can be used to isolate CNT curvature from collagen curvature, interaction energies as a function of bending, etc.

### 6.2 Steered Molecular Dynamics simulations for CNT-Collagen Interactions

With basic mechanical knowledge, the interaction energies between CNT and collagen would be required to fully understand any cooperativity descriptors. Steered Molecular Dynamics, or SMD, is a powerful tool that is commonly used for interaction studies in biomolecular systems [165-167], as well as strength/strain characterizations. The SMD approach applies a moving spring force such that the molecule can behave in a manner not captured by either force or displacement...
loading alone, allowing induced conformational changes in a system along a prescribed reaction vector [28, 168]. In SMD, a driving force can be applied to selected atoms of magnitude

\[ F_{SMD} = K_{spring} (R - R_0) \]  

(60)

where \( K_{spring} \) is the spring constant, and \( R_0 \) is the distance from the end of spring from a target tether point. A constant velocity can be prescribed which monotonously increments or decrements the distance \( R \) towards the tether point, or target coordinate. Advantageously, the application of SMD simulations corroborates with atomic force microscopy (AFM) experiments at the atomic scale [168].

Interaction studies involve the pulling of one of the components of the composite, requiring the use of SMD. Presently, one main concern in computational experiments related to adhesion properties is the assumption that the “error” induced by the finite rate of conformational changes is neglected (e.g., the observed behavior is both a function of the load rate and path). Verifying this however requires extremely slow manipulation and trajectory space sampling which results in excessive computational expense. SMD is a non-equilibrium process, which accepts irreversibility, while it sacrifices accuracy in evaluating binding affinities and potential of mean force, it is able to gain access to relevant information at a longer time scale.

The idea that thermodynamical potentials cannot be obtained from irreversible processes has been proven unfounded by the *Jarzynski Equality* (JE) [169]. The JE is a statistical mechanical equation, similar to the ergodic hypothesis see Section 2.2, that relates the change in free energy, \( \Delta F \), between two equilibrium states via a non-equilibrium process. In a quasi-static process, the work \( W \), done on a system from \( A \to B \) can be defined as

\[ \Delta F = F_B - F_A \leq W \]  

(61)
when the system transition from A to B\textit{ infinitely} slowly. On the other hand the JE is independent of the process speed, stating
\[\exp\left(-\frac{\Delta F}{k_B T}\right) = \left\langle \exp\left(-\frac{W}{k_B T}\right) \right\rangle \]

where \(k_B\) is the Boltzmann constant, and T temperature. Right hand side brackets represent an ensemble average which is an average over all realizations of the external process that takes A\(\rightarrow\)B. This identity relates the free energy difference of the system, an equilibrium property to the ensemble average of the total work \(W\) performed on a system during a non-equilibrium shift from A\(\rightarrow\)B. Since its derivation the validity of the JE has been confirmed to be accurate (although non-exact [170]) approach in both experiments[171] and simulations [172] of biomolecules.

In theory, the left hand side of \textbf{Eq.}\textbf{(62)}, would require that SMD simulations probe the complete sample space of initial conditions and directions of the adhesion as it is necessary to capture the free energy of the system. However the right hand side of the JE implies an average of the work over all phase space trajectories. So probing a finite sample of approaches/trajectories of the binding process would allow for a sufficient approximation of \(W\), which allows for an approximation of the free energy.

There have been extensive research on CNT-polymer interactions [173-179] through pulling and introduction of CNTs into polymer matrices. Here, we propose the same type of studies for the Collagen-CNT composite, the interaction energy is obtained is expressed as
\[U_{\text{Int}} = U_{\text{total}} - (U_{\text{CNT}} + U_{\text{Collagen}})\]
\( U_{\text{int}}, U_{\text{CNT}}, U_{\text{Collagen}} \) represent the total energy of the system, energy of the CNT and energy of the collagen respectively.

### 6.3 Summary

A short proposal is presented on a poorly understood system where cooperativity has the potential to play a role in the multi-scale understanding of making a D-banded collagen fiber. The expectation is that this knowledge can offer insight into an actual experimental problem that is being addressed today. With the results of these studies one can better observe what physically what the native cooperativity of the system is, so that there can be a better understanding and interpretation of the gyration-based cooperativity metrics.
7. CONCLUSION AND EXPECTED SIGNIFICANCE

With the driving motif that the “whole” can be “greater than the sum of its parts”, we herein explored the “rules of cooperativity” for coupled molecules. Here, cooperativity is defined as coupled deformation – similar to the concept of compatibility at the continuum scale - where a molecular pair can be considered as a single unit or “building block” under equilibrium and dynamic conditions. The aim of this dissertation was to develop a mechanistic framework for the action of building blocks in assemblies based on geometry, beyond particular chemistry or cross-linking, extending earlier concepts [180]. If two molecules are sufficiently cooperative, then they can be defined as a “building block” without the need to consider lower-order structure (i.e., “A+B=C” as depicted in Figure 5). Such approaches are necessary if we wish to exploit the material assembly process and use achieve diverse material functions in distinct assemblies.

Motivated by the geometric descriptors inherent to the gyration tensor, a novel trifecta of quantities that show when a pair of macromolecules are cooperating is derived. These quantities are based on the invariants of both the gyration tensor, $S$, and the proposed gyration deviation tensor, $\Delta S$, and reflect differences in size (e.g., slip), shape (e.g., differential anisotropy), and orientation (e.g., skew), to objectively compare the transition from uncooperative to cooperative behavior – i.e., the critical cross-linking density when “1+1=1”. Chapter 3

COOPERATIVITY METRICS covered the derivation of these unique and original metrics, and are briefly shown again:

- The slip defined as

$$\Omega = Tr(\Delta S)$$

where the gyration deviation is presented as the difference of gyration between two molecules.
The differential shape configuration or anisotropy

\[ \delta \kappa^2(t) = \kappa_a^2(t) - \kappa_b^2(t) \]

Based on the anisotropy of each molecule Eq.(43).

And the misorientation of gyration, or skew

\[ \alpha = \cos^{-1} \left[ \frac{S^{(1)}_{xx} S^{(2)}_{xx} + S^{(1)}_{yy} S^{(2)}_{yy} + S^{(1)}_{zz} S^{(2)}_{zz}}{\left( S^{(1)}_{xx} S^{(1)}_{yy} + S^{(1)}_{yy} S^{(1)}_{zz} + S^{(1)}_{zz} S^{(1)}_{xx} \right)^{1/2} \left( S^{(2)}_{xx} S^{(2)}_{yy} + S^{(2)}_{yy} S^{(2)}_{zz} + S^{(2)}_{zz} S^{(2)}_{xx} \right)^{1/2}} \right] \]

Upon the derivation of these metrics, a first system was selected for the examination of the cooperativity metrics. The key question was if the metrics can indicate adequate coupling – e.g., how may cross-links are necessary to induce “cooperativity”. In this case an ideal pair of flexible pseudo linear chain molecules, a model motivated by the cooperativity of electrolyte complexes, served as one validation of a physical system. The usage of ideal chains allowed for the systematic tuning of cross-link density and distribution. In Chapter 4
COOPERATIVITY IN IDEAL MACROMOLECULES, it was determined:

• The slip and norm metrics, Eqs. (47)-(49), indicated that the size difference dropped after the 21/42 cross-link scenario, Figure 28, Figure 29. The initial slip value rises to a maximum of 2.84x10^5 Å, for MC1 and 4.68x10^5 for MC2 and hits a low of at least 500% once it reaches the critical cross-link point.

• The shape metrics Eq.(43) and Eq.(50), likewise, show an initial gap of 0.2 that decreases steadily until the critical cross-link case of 21/42 cross-links where the gap becomes essentially zero, Figure 32.

• The skewness metric Eq. (53), also corroborates the results from previous metrics. With a maximum misorientation of 30 degrees and 6 degrees for both MC1 and MC2 that drop to near zero values after reaching the critical cross-link number of 21 and 42 respectively Figure 33.

• The robustness of the trifecta of geometrical metrics is explored, by exploiting the concept of ergodicity, preliminary results show consistent statistical behavior from all three metrics.

• All three metrics indicated cooperativity after 21/42 cross-links which reflects a free length of 50nm between linkages.

Next, a tropocollagen molecule is selected as a secondary physical system in the continuing efforts to validate the cooperativity metrics. In the case of the collagenous material, there is a well-defined hierarchical system, and one issue is how to determine handshaking between the scales. Looking at the collagen molecule and how the mutations produce deficiencies that show at the macroscale one wonders if there is additional information that the cooperativity metrics can show. The importance of knowing when the parts cooperate can be a key factor and can be used in the future as a tool for material design. Ultimately, the ideal scenario would show
insight on how the lack of cooperativity of the collagen molecule affects the macroscopic properties of bone in OI mutations. Here, the key question was flipped – can the metrics sufficiently indicate a known case of decoupling – e.g., reflect the local unfolding of a mutated tropocollagen? In Chapter 5 COOPERATIVITY IN COLLAGEN MOLECULE it was shown:

- The metrics failed to assess cooperativity on a per chain basis, prompting for a redefinition of the structure, for this case cooperativity was assessed in a sectional fashion, where the comparison of the metrics is based on the relative differences.

- The slip metrics, Eqs. (47), indicated that the size difference increased on section 2, see Figure 44, where the mutation is contained, consistent with the a priori knowledge that mutations produce a protuberance. The average size difference between the healthy and mutated collagen for section 2 was around 6.97 Å, a size gap that exceeds the other sections by 200%. The variance of the data is relatively low compared to the mean values thus, ensuring a clear gap between the sections.

- The shape metrics Eq.(43) and Eq.(50), likewise, presented congruent behavior expected with the presence of the mutation. The individual anisotropy for the section 2 comparison shows a drop of nearly 15% in the anisotropy value for the mutated molecule, bringing the anisotropy value closer to zero, therefore suggesting a more spherical configuration which can be interpreted as a swelling of the collagen molecule in the replaced glycine. The other sections were statistically similar showing consistency between sections that do not contain the mutation, Figure 45 and Figure 46.

- The skewness metric Eq. (53), also agrees with the results from the previous metrics. In this case the natural orientation of the gyration collagen is ≈1 degree, Figure 47. While the change in magnitude of the orientation is negligible, local statistical parameters, Figure 48 and Figure 49, show that the fluctuations of section 2 are in fact distinct from the other 2 sections.
The local variability exceeds the others by 300%, and might provide insight as to the how mutations affect the hierarchical performance of the collagen fibers.

- Briefly showed the viability of using larger, thus less, sections to judge cooperativity, increasing the usability if one were to address the issue of using the whole 300nm collagen molecule, Figure 50.

Finishing the cooperativity study of tropocollagen molecule, an interesting topic is left as proposed work. With the objective of using cooperativity as an aid in experimental work, an outline of a study that characterizes the system and the interactions is shown in Chapter 6

**PROPOSED COLLAGEN-CARBON NANOTUBE INTERACTION.** The main tasks involved in the study are as follow:

- Employ suite of *in-silico* tests to study the mechanical properties (i.e. Young’s Modulus, tensile strength, bending strength).
- Use SMD to obtain the interaction energy from a full description between Carbon Nanotube and Collagen.
- Assess a higher hierarchical scale of cooperativity of carbon nanotube bundles with collagen fibrils and correlate with the results obtained from the tests.
- Validate with ongoing experimental work.

The expectation is that this study can be a key factor in the current endeavor of reproducing a D-banded collagen fiber.

Individually, the metrics described cannot differentiate between cooperative/coupled and uncooperative/uncoupled conditions, see Section 4.4. Certainly, two molecules of different sizes could be sufficiently cross-linked, while expressing “slip” of a relatively large, finite (by constant) magnitude. Likewise, two uncoupled molecules could both be relatively extended in...
conformation, and thus the associated differential anisotropy would approach zero. Cooperativity must be assessed through prudent combination of multiple geometric metrics, across time, and in consideration of system particulars (such as differences in molecular mass). The results indicate that the use of these three metrics together offer insight to whether the molecules are in a cooperative state. Additionally, the statistical descriptions of the metrics results can potentially be used to inquire about specific mechanical properties.

Reliability at a single scale requires the assumption that molecules behave as a “single unit” and thus a basis for assembly. This, in turn, begs the question under what conditions can coupled molecules be considered a single entity. In nano- and molecular mechanics the idea of compatibility is rarely mentioned, and most of the focus has been on reaching a required performance and/or limit states. However this missing information can become a powerful tool in the assessment and development of functional materials systems. The results presented here are promising and certainly open further discussion of how to both determine and quantify if macromolecular systems are adequately cooperative.

Indeed, the desire for material self-assembly to propagate across scales involves the formation of building blocks that seamlessly assemble higher-order building blocks [15, 16], from one scale to another in a hierarchical manner, and molecular cooperation/compatibility is a necessary requirement not yet fully understood.
REFERENCES


53. Pearlman, D.A., D.A. Case, and e. al., *AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free


Molecular Cooperativity/Compatibility

K. Kwan, 2015


*Molecular Cooperativity/Compatibility*  
K. Kwan, 2015


*Molecular Cooperativity/Compatibility*  
K. Kwan, 2015


Appendix A  THE EIGENVALUE PROBLEM

In general for any square n by n matrix, the eigenvalue problem can be stated as

$$As = \lambda s$$  \hspace{1cm} (64)

where $\lambda$ is a constant called the eigenvalue and $s$ the associated eigenvector. For the sake of keeping the discussion in a mechanics point of view let us take $n$, the dimension, to be 3 and $A$ to be symmetric and real. As mentioned previously the eigenvalues are an invariant quantity, meaning they remain unchanged under coordinate transformations. Other invariant quantities are the trace, determinant and any subsequent combination of invariants, (e.g. ratios of eigenvalues and traces, as shown in Section 3.2.1.

One can rearrange Eq.(64) into

$$\left( A - \lambda I_{3 \times 3} \right) s = 0$$  \hspace{1cm} (65)

From a mathematics perspective Eq.(65), is a homogenous equation and it has a nontrivial solution if and only if the determinant

$$\left| A - \lambda I_{3 \times 3} \right| = 0$$  \hspace{1cm} (66)

where

$$\left| A - \lambda I_{3 \times 3} \right| = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} = 0$$  \hspace{1cm} (67)

The characteristic equation Eq. (66) can then be expressed as

$$\lambda^3 - (a_{11} + a_{22} + a_{33}) \lambda^2 + (a_{11}a_{22} + a_{22}a_{33} + a_{33}a_{11} - a_{12}^2 - a_{13}^2 - a_{23}^2) \lambda -$$

$$\left( a_{11}a_{22}a_{33} - a_{11}a_{22}a_{33} - a_{11}a_{22}a_{33} - a_{11}a_{22}a_{33} + 2a_{12}a_{23}a_{13} \right) = 0$$  \hspace{1cm} (68)

The roots of the characteristic equations are then

$$\lambda_1, \lambda_2, \lambda_3$$  \hspace{1cm} $\lambda_1 < \lambda_2 < \lambda_3$  \hspace{1cm} (69)

the directions of these roots are given by substituting the value of each root into Eq. (65) and solving for each $s$.

\[\text{To the painful disgust of mathematicians, the word tensor and matrix are used interchangeably here, the term tensor is usually reserved for mathematical objects that are needed in physics to define certain quantities, whereas a matrix is just a structure of numbers. While a second order tensor is certainly a matrix, the same cannot be said from a matrix.}\]
From a mechanics context the matrix $A$ can be the stress tensor $E$, strain tensor $\varepsilon$, Eq.(1), or the gyration tensor $S$, Eq.(38) then the principal values have a physical significance and are then known as the principal stresses, principal strain and principal radius of gyration respectively, and their associated eigenvectors are the principal directions.

The characteristic equation can then be expressed as an equation

$$\lambda^3 - I_1\lambda^2 + I_2\lambda - I_3 = 0 \quad (70)$$

where the coefficients $I_i$, not to be confused with the identity matrix $I_{3x3}$, of $\lambda$ are known as the invariants and each is defined as

$$I_1 = (a_{11} + a_{22} + a_{33}) = Tr(A) \quad (71)$$

$$I_2 = (a_{11}a_{22} + a_{22}a_{33} - a_{12}^2 - a_{13}^2 - a_{23}^2) = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{13} \\ a_{13} & a_{33} \end{bmatrix} + \begin{bmatrix} a_{22} & a_{23} \\ a_{23} & a_{33} \end{bmatrix} \quad (72)$$

$$I_3 = (a_{11}a_{22}a_{33} - a_{11}a_{23}^2 - a_{22}a_{13}^2 - a_{33}a_{12}^2 + 2a_{12}a_{23}a_{31}) = \det(A) \quad (73)$$

The physical interpretation of the invariants will depend on the nature of the tensor, for stress and strain tensors the first invariant is directly related to the hydrostatic stress, (i.e. $\sigma_{\text{hydrostatic}} = \text{Trace}(E)/3$), the second invariant is related to the deviatoric components of the tensor and the third variant is applied when looking at deformation gradients. The use of invariants are of importance when looking at failure analysis. (e.g. von Mises and Tresca failure models).

For the case of the gyration tensor the first invariant corresponds to Eq.(42) and represents the radius of gyration. In polymer science the invariance of the trace is used as a way to circumvent the computational difficulties of shape analysis with the eigenvalues themselves.
Appendix B  THE COLLAGEN STRUCTURE MODEL

In general to generate a full atomistic model for biological materials three files are needed.

1. .PDB File – Protein Data Bank file, it contains the atomic coordinates and other information describing proteins and macromolecules.

2. Topology File – Contains all the information needed to convert a list of residue names into PSF structure file. It also contains the internal coordinates and allows for automatic assignment of coordinates to hydrogen and missing atoms.

3. Parameter File – Supplies specific numerical value for generic CHARMM potential functions. It maps the topology file.

From these three files you can generate the PSF, Protein Structure File, which contains all the molecule-specific information to apply a force field to a molecular system.

B.1 Creating The .PDB File

The Interactive Triple-Helical collagen Building Script (TheBuScr)\cite{128}, is used to generate the collagen structure. The script allows one to generate the collagen molecule with any chosen sequence. It requires Tcl/Tk programming language to run.

TheBuScr can be obtained through the website of The Rainey lab:

\url{http://structbio.biochem.dal.ca/jrainey/THeBuScr.html}

To utilize the script one must first generate a .FASTA extension file with the intended collagen sequence.

An example FASTA file is presented in Figure 54, with a random number of Gly-X-Y sequence.
Figure 54: Example of .FASTA extension file, presents 90 amino acids in the customary GLY-X-Y sequence found in collagen.

In Linux ensure that active directory is THeBuScr and type the following command to run the script

```
wish THeBuScr.1.07.tcl
```

The following window opens up, Figure 55

![Main window of The BuScr Script to generate the collagen structure.](image)

Figure 55: Main window of The BuScr Script to generate the collagen structure.

To start one must click on the **Select file** button and insert the FASTA file generated earlier for each chain, if all alpha chains are the same, which is the case for the current collagen study the next window Figure 56, has the option to select the same file for all chains.

![Chain file selection for collagen structure, THe BuScr.](image)

Figure 56: Chain file selection for collagen structure, THe BuScr.
After selecting all files for chain, click on Select Triple-Helix, here one must select the N-terminal ending amino acid and the C-terminal starting Amino Acid, see Figure 57. Polypeptides and proteins are usually referred to as the start and end of the structure, and for the purposes of the study are irrelevant. Remembering that the collagen structure sequence is GLY-X-Y, it is important that the selection of N-terminal acid be an amino acid at the Y position and that the selection of the C terminal amino acid be Glycine, if not the program would crash since it is a violation of the collagen sequence. Apply these termini for all chains.

Next step is to click on the Set/Show helical params button, after ensuring that the termini are set, the Helical propensity and helix settings window opens, this window shows some information concerning the melting temperature. Leave everything as is and click on Done, see Figure 58.
Figure 58: Helical propensity and helix settings window, here some information concerning helicity and melting temperature is presented. Leave everything as is and click Done.

The last step is the **Build Helix** button, see **Figure 59**, this gives you the option of selecting which atoms to output in your collagen structure. The default options satisfy the needs of the current study, Select File for output and click on Produce PDB file and Done. This creates the PDB file that is needed to create the PSF file.

Figure 59: Generate triple-helical coordinates window, this windows allows for the selection of atoms, the coordinate system and other options. It is the last step for THE BuScr to generate the PDB file.
B.2 Generating the PSF File

To generate the PSF file, the Visual Molecular Dynamics (VMD) software is used. You will need to load the generated PDB file from the previous section. VMD is a molecular visualization program for displaying, animating and analyzing molecular systems. VMD contains the AutoPSF generator, see Figure 60, it is located in Extensions > Modeling > Automatic PSF Builder.

The first task is to add the in Step 1 topology file, go to Step 3 assign the chains manually, and click on the create chains button, see Figure 60.

![AutoPSF generator module in VMD](image)

**Figure 60:** Auto PSF generator module in VMD, to operate just add the topology file, and manually assign the chain.
B.3 Inserting Mutations

Thankfully VMD also contains a module for mutations. Remember from Section 5.1.1, the point mutation in collagen can be represented by replacing the Glycine amino acid with a select variety of amino acids. The module in VMD can be accessed by

Extension>Modeling>Mutate Residue

This opens a window, see Figure 61, where the PSF and PDB files, generated earlier, need to be inputted. The Target Residue field corresponds to the amino acid and chain that is being replaced and the Mutation field has the three letter code of the replacement amino acid. After these fields are completed clicking Run Mutator generates a new set of PSF files with the mutated version of collagen.

Figure 61: Mutator module in VMD, to use just browse the original PSF and PDB, select the residue to be replaced and input the three letter code of the replacement amino acid.

Molecular Cooperativity/Compatibility  
K. Kwan, 2015
B.4 Converting to LAMMPS Format - Solvating

All of the above generated files are formatted as is for CHARMM, to use LAMMPS one must use a tool called ch2lmp located in:

lammps->tools>ch2lmp

To run it the command in linux is

```
perl charmm2lammps.pl all22_prot_HYP GPOsM
```

The above command is broken down and explained.

- `perl charmm2lammps.pl` runs perl script named charmm2lammps.pl
- `all22_prot_HYP` name of both topology and parameter files. (e.g. all22_prot_HYP.top, all22_prot_HYP.prt
- `GPOsM` name of generated .data and .in file for lammps.
- `-charmm` makes charmm format comments (useful to check generated with original)
- `-border=8` adds border to all sides of simulation box (angstroms)
- `-lx,-ly,-lz` x y z dimensions of simulation box
- `-center` recenter atoms
- `-pdbctrl` generates new pdb
- `-water` adds TIP3P water molecules
- `-ions` adds counter ions using Na+ or Cl-

This results in a .in (input file) and a .data file that can be used to run in LAMMPS.

With these steps one can generate a collagen module with any sequence, provided that the topology and parameter file are provided, that can be simulated through LAMMPS.
Appendix C  FASTA File Alpha1.fasta

>alpha 1

GPMGPSGPRGLPGPPGAPGQPQGFQGPPGEPEPGASGPMGPRGPP
GPPGKNGDDGEAGKPRPGERGPPGOPQGARGLPTAGLPGMKGH
GFSGLDGAKGDAGPAGPKGEPSPGAGFAGAGKMR
GPOGPOGPOGPOGPOGPOGPOGPOGPOGPOPO
GARGPAGPQGPRGDKGETGEGQGDRIKGHGFSGLQGPPGPPGSP
GEQGSPASGPAGPRGPQGASAGPKDGKGLNGLPQPGPRGRT
GDAGPVGPAGPQGPPGPPGPPGPP
Appendix D  Topology File Hydroxyproline

RESI HYP    0.00
GROUP       !     HD1 HD2
ATOM N      N  -0.2900 !     |   \ /
ATOM CA     CP1  0.0200 !     N---CD HG
ATOM HA     HB  0.0900 !     |   \ /
ATOM CD     CP3  0.1800 !     \     CG
ATOM HD1    HA  0.0000 !     |   / \
ATOM HD2    HA  0.0000 !  HA-CA--CB  OG2--HG2
GROUP       !     |   / \
ATOM C      C   0.5300 !     | HB1 HB2
ATOM O      O  -0.5300 !   O=C
GROUP       !     |
ATOM CB     CP2  -0.1800
ATOM HB1    HA   0.0900
ATOM HB2    HA   0.0900
GROUP
ATOM CG     CP2  0.1400
ATOM HG     HA   0.0900
ATOM OG2    OH1   -0.6600
ATOM HG2    H    0.4300
BOND C CA   C +N
BOND N CA   CA CB CB CG CG CD N CD
BOND CA   HB1 CB HB2 CD HD1 CD HD2 CG HG
BOND CG   OG2 OG2 HG2
DOUBLE O C
IMPR N -C CA CD
IMPR C CA +N O
DONOR HG2 OG2
ACCEPTOR O C
IC CD  -C *N CA  1.4468 122.5296 -179.2170 122.5076 1.4550
IC N  CA  C +N  1.4550 111.6376 -31.0428 120.0469 1.3161
IC +N CA  *C O  1.3161 120.0469 -179.5571 120.4223 1.2316
IC N  CA  CB CG  1.4550 101.8808  29.6080 103.6605 1.5362
IC N  CA  CB HB1  1.4550 101.8808 152.0681 112.9350 1.0925
IC N  CA  CB HB2  1.4550 101.8808 -86.5729 108.9663 1.0930
IC CA  CB CG HG  1.5360 103.6605  79.0616 109.5554 1.0926
IC CD  N  CA CB  1.4468 114.9587 -11.2918 122.5076 1.3162
IC CD  N  CA HA  1.4468 114.9587 -129.2290 110.5162 1.5360
IC CD  N  CA CB  1.4468 114.9587  109.5875 111.6376 1.5380
IC HA  CA  N  -C  1.0924 110.5162  50.0429 122.5076 1.3162
IC CA  N  CD CG  1.4550 114.9587 -11.8599 102.4794 1.5305
IC CA  N  CD HD1  1.4550 114.9587 -131.8901 111.1069 1.0924
IC CA  N  CD HD2  1.4550 114.9587 106.0401 109.5406 1.0925
IC N  CA  C  O  1.4550 111.6376 149.4002 120.4223 1.2316
IC CA  CB CG OG2  1.5360 103.6605 -158.7449 112.5784 1.4300
IC CD  CG OG2 HG2  1.5305 112.2089  31.3937 109.5000 1.0300
IC N  CD CG CB  1.4468 102.4794  29.9288 103.3139 1.5362
PATCHING FIRS PROP

Molecular Cooperativity/Compatibility  K. Kwan, 2015
Appendix E  LAMMPS Input file for Ideal Chains

#----------------------------------------------
# Coarse-grain model; cross-linked chains
# Kenny Kwan
#----------------------------------------------

# Initialization

units real
atom_style molecular
timestep 2.0

# Atom Definition - Indicate input geometry file

read_data chain_250.dat

# Neighbor Settings

neighbor 100 bin
neigh_modify every 1000 delay 1000

# Force Fields and Interactions

bond_style harmonic
bond_coeff 1  50. 10.00
bond_coeff 2  100. 20.00

angle_style harmonic
angle_coeff 1  100. 180.00

pair_style lj/cut 10.0
pair_coeff * * 100.0 8.9090

# Basic Output

thermo  5000
dump        myDump all xyz 10000 chain250.xyz
compute       com1 chain1 com
compute       com2 chain2 com
compute       gyr0 all gyration

Molecular Cooperativity/Compatibility  K. Kwan, 2015
compute  gyr1 chain1 gyration
compute  gyr2 chain2 gyration
compute  peratom all pe/atom pair
compute  pe_reax all reduce sum c_peratom

fix  peout all ave/time 1 1 5000 c_pe_reax file chain_pe250.data
velocity  all create 300.0 351713
fix  3 all nvt temp 300. 300. 100.

run  5000000

# ---------------------------------------------------------
# END OF SIMULATION
# ---------------------------------------------------------
Appendix F  MATLAB Gyrations Ideal Chains

%%%Appendix F MATLAB Gyrations Ideal Chains
%---------------------------------------------------------------
%Compute Gyration based Cooperativity Metrics
%
%Kenny Kwan
%---------------------------------------------------------------

```matlab
fls = dir( 'chain_gyr***.data' );
for fi=1:numel(fls)
    G=importdata( fls(fi).name, ' ', 2 );
    GYR=G.data;
    GYR=GYR(:,2:13);
    [nr,nc]=size(GYR);

    %Lammps orders the gyration tensor like this xx, yy, zz, xy, xz, yz.
    for i =1:6
        dGYR(:,i)=GYR(:,i)-GYR(:,i+6);
    end

    for i=1:nr
        %Assembling the gyration tensor
        g1a=diag(GYR(i,1:3));
        g1b=zeros(3);
        g1b(1,2:3)=GYR(i,4:5);
        g1b(2,3)=GYR(i,6);
        g1=g1b'+g1b;
        G1=g1a+g1;

        g2a=diag(GYR(i,7:9));
        g2b=zeros(3);
        g2b(1,2:3)=GYR(i,10:11);
        g2b(2,3)=GYR(i,12);
        g2=g2b'+g2b;
        G2=g2a+g2;

        dG=(G1-G2);

        %Calculating the trace for each tensor/Slip
        %The trace is basically the size of a polymer
        T1= trace(dG);
        Slip(i)=T1;
        TT1=trace(G1);
        TT2=trace(G2);
```

Molecular Cooperativity/Compatibility

K. Kwan, 2015
deltaT=TT1-TT2;
nDT=norm(deltaT);

%For ashperecity

dGhat=(dG-1/3*(T1)*eye(3));
dGhat2= dGhat^2;

T2= trace(dGhat2);

A = (3/2).*(T2)/(T1*T1);

aspHERE(i) = A;

%Testing Asphere single chain
TG1=trace(G1);
Ghat=G1-1/3*TG1*eye(3);
TG12=trace(Ghat^2);
AA1(i)=(3/2).*((TG12))/(TG1*TG1);

TG2=trace(G2);
Ghat2=G2-1/3*TG2*eye(3);
TG22=trace(Ghat2^2);
AA2(i)=(3/2).*((TG22))/(TG2*TG2);

deltaA=AA1-AA2;
nDA=norm(deltaA);

%Distance between the matrices
norm2(i)=norm(dG);
Nnorm(i)=norm2(i)/norm(G1);

[eigvec,eigval]=eig(dG);
DG{i}=dG;

AnglesG(i)=subspace(G1,G2);
end

tt(fi)= nDT;
nn(fi)=nDA;
end
function [Eval1, Eval2, Eval3] = GyrationInput(x,numG);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Organizing radius of gyration
%Kenny Kwan Yang 2015
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%to use
%[Eval1, Eval2, Eval3] = GyrationInput(x,numG)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Lammps orders the gyration tensor in the following manner
% xx, yy, zz, xy, xz, yz.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%OUTPUTS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%Eval1,Eval2,Eval3 = Eigenvalues of Gyration Tensors for Chain1,2,3
%The Radius of Gyration
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%INPUTS
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%x = Gyration data from lammps.
%numG = 1 or 3, takes the whole gyration or the gyration of 3 chains. 2
%Chains problems still need to be addressed (just another if)

fls = dir( x );
for fi=1:numel(fls)
    G=importdata( fls(fi).name, ' ', 2 );
    GYR=G.data;
    if numG == 1
        GYR=GYR(:,2:7);
    else
        GYR=GYR(:,2:end);
    end

    [nr,nc]=size(GYR);
    for i=1:nr
        %Assembling the gyration tensor
        gla=diag(GYR(i,1:3));
        glb=zeros(3);
        glb(1,2:3)=GYR(i,4:5);
        glb(2,3)=GYR(i,6);
        gl=glb'+glb;
        Gl=gl+a+gl;
        [vec1,vall]=eig(Gl);

    end

    %end

end

end
Rgs1(i,1:3)=diag(val1);
GG1{i}=G1;

if numG == 3

g2a=diag(GYR(i,7:9));
g2b=zeros(3);
g2b(1,2:3)=GYR(i,10:11);
g2b(2,3)=GYR(i,12);
g2=g2b'+g2b;
G2=g2a+g2;

g3a=diag(GYR(i,13:15));
g3b=zeros(3);
g3b(1,2:3)=GYR(i,16:17);
g3b(2,3)=GYR(i,18);
g3=g3b'+g3b;
G3=g3a+g3;

[vec2,val2]=eig(G2);
[vec3,val3]=eig(G3);
Rgs2(i,1:3)=diag(val2);
Rgs3(i,1:3)=diag(val3);
GG2{i}=G2;
GG3{i}=G3;
end
end

if numG == 1;
   Eval1=Rgs1;
   Eval2=''
   Eval3='';
else
   Eval1= Rgs1;
   Eval2=Rgs2;
   Eval3=Rgs3;
end

end
Appendix H  MATLAB Anisotropy and Orientation Metrics

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
%Anisotropy and Orientation Metrics.
%Kenny Kwan Yang 2015
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

load Section3M.mat
load Section3H.mat

Rgs1M=Rgs1M3;
Rgs2M=Rgs2M3;
Rgs3M=Rgs3M3;

Rgs1=Rgs1H3;
Rgs2=Rgs2H3;
Rgs3=Rgs3H3;

nr=length(Rgs1M);

%Setting up the gyration deviatoric (for up to 5 sections)
Shat1H=mean(Rgs1,2);
Shat2H=mean(Rgs2,2);
Shat3H=mean(Rgs3,2);
% Shat4H=mean(Rgs4,2);
% Shat5H=mean(Rgs5,2);

Shat1M=mean(Rgs1M,2);
Shat2M=mean(Rgs2M,2);
Shat3M=mean(Rgs3M,2);
% Shat4M=mean(Rgs4M,2);
% Shat5M=mean(Rgs5M,2);

%calculating individual shape anisotropy per chain
kappa1H=3/2*(sum((Rgs1-[Shat1H Shat1H Shat1H]).^2,2))./(sum(Rgs1,2)).^2;
kappa2H=3/2*(sum((Rgs2-[Shat2H Shat2H Shat2H]).^2,2))./(sum(Rgs2,2)).^2;
kappa3H=3/2*(sum((Rgs3-[Shat3H Shat3H Shat3H]).^2,2))./(sum(Rgs3,2)).^2;
% kappa4H=3/2*(sum((Rgs4-[Shat4H Shat4H Shat4H]).^2,2))./(sum(Rgs4,2)).^2;
% kappa5H=3/2*(sum((Rgs5-[Shat5H Shat5H Shat5H]).^2,2))./(sum(Rgs5,2)).^2;

kappa1M=3/2*(sum((Rgs1M-[Shat1M Shat1M Shat1M]).^2,2))./(sum(Rgs1M,2)).^2;
kappa2M=3/2*(sum((Rgs2M-[Shat2M Shat2M Shat2M]).^2,2))./(sum(Rgs2M,2)).^2;
kappa3M=3/2*(sum((Rgs3M-[Shat3M Shat3M Shat3M]).^2,2))./(sum(Rgs3M,2)).^2;
% kappa4M=3/2*(sum((Rgs4M-[Shat4M Shat4M Shat4M]).^2,2))./(sum(Rgs4M,2)).^2;

Molecular Cooperativity/Compatibility  K. Kwan, 2015
% kappa5M=3/2* (sum((Rgs5M-[Shat5M Shat5M Shat5M]).^2,2))./(sum(Rgs5M,2)).^2;

%Orientation mismatch
alpha1HM=sum(Rgs1(1:nr,:).*Rgs1M,2)./(((sum(Rgs1(1:nr,:).*Rgs1(1:nr,:),2 ))).^((1/2)).*(sum(Rgs1M.*Rgs1M,2)).^((1/2)))
alpha2HM=sum(Rgs2(1:nr,:).*Rgs2M,2)./(((sum(Rgs2(1:nr,:).*Rgs2(1:nr,:),2 ))).^((1/2)).*(sum(Rgs2M.*Rgs2M,2)).^((1/2)))
alpha3HM=sum(Rgs3(1:nr,:).*Rgs3M,2)./(((sum(Rgs3(1:nr,:).*Rgs3(1:nr,:),2 ))).^((1/2)).*(sum(Rgs3M.*Rgs3M,2)).^((1/2)))
alpha4HM=sum(Rgs4(1:nr,:).*Rgs4M,2)./(((sum(Rgs4(1:nr,:).*Rgs4(1:nr,:),2 ))).^((1/2)).*(sum(Rgs4M.*Rgs4M,2)).^((1/2)))
alpha5HM=sum(Rgs5(1:nr,:).*Rgs5M,2)./(((sum(Rgs5(1:nr,:).*Rgs5(1:nr,:),2 ))).^((1/2)).*(sum(Rgs5M.*Rgs5M,2)).^((1/2)))