11-26-2012

Towards Sequence-Based DNA Flexibility Analysis

Emily Flynn
Smith College

Filip Jagodzinski
Central Washington University

Ileana Streinu
Smith College, istreinu@smith.edu

Towards Sequence-Based DNA Flexibility Analysis

Emily Flynn
Smith College
Northampton, MA, USA
eflynn@smith.edu

Filip Jagodzinski
Central Washington University
Ellensburg, WA, USA
jagodzinski@cwu.edu

Ileana Streinu*
Smith College & UMass
Northampton & Amherst, MA
istreinu@smith.edu

ABSTRACT

In this poster, we present an extension to our freely available KINARI-Web server to identify rigid and flexible regions of nucleic acids and protein-nucleic acid complexes contained in the Protein Data Bank (PDB). The goal is to explore the effect of DNA and RNA on the rigidity and stability of these structures. We also propose an approach for determining DNA rigidity based solely on sequence. Currently, only the rigidity of DNA molecules whose structures have been deposited in the PDB (approx. <4,000 files) can be analyzed. Once fine-tuned and validated, this new coordinate-free method for investigating DNA flexibility could be applied to the more than 135 million sequences in GenBank, and to nanostructure design.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Biology and genetics; I.6.3 [**Simulation and Modeling**]: Applications

General Terms

Rigidity, Flexibility, DNA, RNA, Proteins, Pebble Game

1. INTRODUCTION AND BACKGROUND

Understanding the rigidity and flexibility of protein-nucleic acid complexes may shed light on their stability and function. The Protein Data Bank (PDB) includes X-ray solved structures (such as viruses and ribosomes) containing both proteins and nucleic acids. KINARI-Web [1] is a server for protein rigidity analysis, developed in our group. It is freely available at `http://kinari.cs.umass.edu`. This poster reports on a current project aimed at extending KINARI to work with DNA and RNA, in order to understand their effect on the flexibility of nucleic acid-protein complexes.

We proceed with a brief summary of the pebble game paradigm underlying rigidity analysis software, and a short introduction to KINARI-Web, our free online server implementing this method for protein studies.

The Pebble Game Paradigm and Rigidity Analysis of Proteins. While physics-based simulations such as molecular dynamics are useful for gaining insight into motions of molecules at the atomic-level, they are computationally too expensive. Rigidity analysis is an alternative, fast, graph-based method that identifies rigid clusters of atoms in molecules modeled as mechanical structures. It relies on an algorithm (the *pebble game*) whose origins can be found in a simple counting rule identified in 1864 by James Clerk Maxwell [11]. The validity of Maxwell's counts was proven by Laman [9] for dimension 2, and extended to the analysis of 3-dimensional structures called body-bar-hinge frameworks by Tay [16]. The pebble game algorithm is an adaptation by Hendrickson and Jacobs [7] of an earlier method of Hendrickson [4], based on bipartite matching. It was first applied to analyze glass networks, then proteins, by Jacobs, Thorpe, and their collaborators [6, 5, 17].

KINARI-Web: Protein Rigidity Analysis. KINARI-Web [1] is a second generation software for rigidity analysis of molecules, developed in our group. It is freely available as a web server and provides options for larger-scale computational experiments. In particular, it streamlines the curation of input protein data, including the addition of missing hydrogen atoms not present in the X-ray solved structures from the Protein Data Bank (using the *Reduce* [21] software), as well as the calculation of the important stabilizing interactions within the molecule (covalent bonds, hydrogen bonds and hydrophobic interactions). From this information, a mechanical model of the molecule is constructed. Atoms along with their covalently bonded neighbors form bodies, while covalent bonds between bodies are modeled as hinges. Other stabilizing interactions are modeled as hinges or bars. From the mechanical framework, an associated multi-graph is constructed, with a node for each body, an edge for each bar, and 5 edges for each hinge. The pebble game algorithm [10] decomposes this graph into clusters corresponding to rigid components in the framework, and hence clusters of atoms in the biomolecule (Figure 1). The rigidity results can be visually investigated using an integrated JMol viewer.

2. METHODS

We describe now the extensions to KINARI needed to analyze nucleic acids from X-ray solved PDB files. We then sketch a method for sequence-based DNA rigidity analysis.

2.1 Extending KINARI-Web to Nucleic Acids

In order for KINARI to analyze nucleic acids, modifica-

*Corresponding author

ACM-BCB'12, October 7-10, 2012, Orlando, FL, USA
ACM 978-1-4503-1670-5/12/10

tions were made to the curation steps of KINARI-Web to recognize, besides proteins, the presence of RNA, and DNA molecules in a PDB file. The algorithm for calculating the covalent bonds was adapted to work for nucleic acids. Identifying non-covalent interactions required more specific modifications, discussed below.

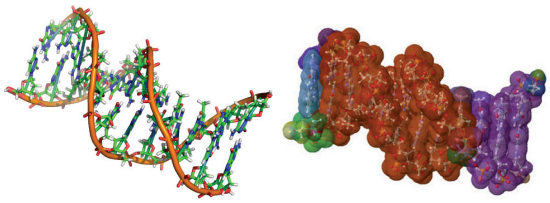

Figure 1: A segment of a 12-base DNA sequence (PDB file 119d) is shown on the left. The rigidity of the DNA strand, as calculated using the KINARI-Web software, is shown on the right. Same-colored regions represent clusters of atoms that are rigid.

Identifying Hydrogen Bonds in Nucleic Acids. Since no single fully agreed-upon method for calculating hydrogen bonds exists, we have explored multiple options that are based on different principles; not surprisingly, they produce varying results. KINARI-Web integrates two external pieces of software, HBPlus [12] and bndlst [8], as options for hydrogen bond calculation. The output of both methods was compared to FR3D [15], a program for finding 3D structural motifs in RNA. KINARI does not use FR3D to identify hydrogen bonds because it finds interactions between nitrogen bases, not atoms, and our software requires atom-atom constraints for rigidity analysis. Since FR3D is an RNA-specific program, it allows us to compare its results to what the options available in KINARI computed in terms of hydrogen bonds for DNA and protein-nucleic acid complexes. Bndlst, a Kinemage Lab software for identifying covalent and hydrogen bonds, finds many more interactions than FR3D, and it is not easy to automatically prune out bonds that should not be there. As a result, our default method for DNA and RNA was selected to be HBPlus, which produces results comparable to FR3D. HBPlus is also the default option for proteins, and was used in the profiling and validation of KINARI [1, 13]. For HBPlus to work with nucleic acids, DNA residues were added as parameters and the PDB files were pre-processed to convert the nomenclature of ribose atoms and phosphate oxygens to those found in HBPlus. A close investigation of HBPlus (at source code level) reveals that the program does not look for base-phosphate hydrogen bonds with carbon donors, which are identified by FR3D. Although these hydrogen bonds are weak (-0.1 to -1.1 kcal/mol) [22], they may have an effect on nucleic acid rigidity. An on-going task is the investigation of how the inclusion of these weak interactions affects KINARI rigidity analysis of DNA and RNA structures.

Identifying Hydrophobic Interactions. Properly and automatically identifying hydrophobic interactions in proteins is a challenging, and not entirely validated task, as it relies on various heuristics. It is even more important to do it properly in nucleic acids, because their stability is so dependent on stacking interactions. KINARI identifies hydrophobics using a heuristic approach that adds constraints when the Van der Waals radii of two sulfur or carbon atoms lie within a distance of 0.25 Å. The energy of these hydrophobic interactions is found by calculating the Lennard-Jones potential using atom types from AMBER [20]. To calculate the energies for these constraints in DNA and RNA, nucleic acid AMBER atoms types were added to KINARI. Additional modifications to the heuristic method remains to be explored, as it has been shown that using a similar method for identifying hydrophobics results in overly rigid RNA structures [2].

2.2 Analyzing DNA Rigidity from Sequence

While there are over 135 million sequences in GenBank, fewer than 4,000 structures with nucleic acids have been solved. The previously described method for rigidity analysis requires atom coordinates. Our next goal is to make progress toward investigating DNA flexibility from sequence only. We require a method for placing non-covalent interactions without needing atom coordinates. We propose a three-step method for sequence-based rigidity analysis of DNA: (1) generate a complementary strand and covalent bonds to the molecular framework based on general knowledge of nucleotide structure, (2) place non-covalent interactions (hydrogen bonds and hydrophobics), and (3) perform rigidity analysis. 3D-DART [18], a web server for DNA structure modeling, is used to convert the sequence to a list of coordinates for visualization purposes.

Identifying hydrogen bonds from DNA sequence. We rely on knowledge of the standard Watson-Crick model to place hydrogen bonds between residues of the template and complementary strands. For benchmarking purposes, PDB files of DNA molecules were analyzed using HBPlus, and hydrogen bond outputs compared to those generated by our sequence-based approach.

Identifying hydrophobics from DNA sequences. We focused on hydrophobic stacking interactions between consecutive base pairs, as they are critical in DNA stability [19]. A data set of DNA PDB structures [3] was used to identify the most frequent interactions between each of the 16 types of base pairs.

3. DISCUSSION AND FUTURE WORK

We discuss now a few issues that remain to be addressed for effective nucleic acid rigidity analysis.

3.1 Nucleic Acid X-ray data

For validating rigidity analysis of nucleic acids, a comparison of the rigidity predictions for resolved nucleic acid structures should be performed against known structural properties of the molecules. We are collecting a benchmarking data set, anticipated to find future use in the fine tuning of the placement and modeling of non-covalent interactions. Possible issues in identifying these constraints are formulated next.

Examining the Contributions of Hydrogen Bonds. HBPlus does not look for base-phosphate hydrogen bonds with carbon donors. While preliminary results suggest that similar rigidity results are produced whether or not these constraints are included, their impact on rigidity should be examined. In addition, HBPlus identifies multiple ribose-ribose hydrogen bonds in RNA that are not listed in the FR3D database. The effects of these constraints on rigidity must be explored to determine if they should be included in the molecular model of nucleic acids.

Towards Modeling Hydrophobic Interactions. KINARI's own method for identifying and modeling hydropho-

bic interactions has to be further validated using data sets of RNA and DNA molecules. Rigidity results should be compared to experimental data using varying parameters, including limiting the number of constraints and varying the cutoff distance, to improve the accuracy of modeling these interactions. In addition, it is of interest to examine how changing the type of constraint used to model hydrophobic interactions affects rigidity.

Modeling Phosphate-Phosphate Repulsions. The repulsion between the negatively charged phosphates in nucleic acid backbones stabilizes their structures, making them more rigid. Bending DNA leads to the unfavorable crowding of phosphates on the inner face of the bend, which is partially but not entirely compensated for by increasing the separation of phosphates on the outer face [14]. Although stacking interactions have a dominant influence on DNA rigidity, phosphate-phosphate repulsions also affect stability. In future work, methods should be developed for modeling these interactions.

3.2 Sequence-Based DNA Rigidity Analysis

KINARI analysis of DNA molecules with known structures should be compared to the results of the sequence-based approach to assess the accuracy of computing rigidity from sequence only. Since the proposed coordinate-free method for placing hydrophobics is based on patterns in a small set of known DNA structures, a much larger data set of DNA molecules should be examined to determine the most frequent constraints.

Conclusions. We have adapted KINARI-Web to analyze the rigidity of nucleic acids with known structures, and proposed a method to perform DNA sequence-based rigidity analysis. Remaining work includes fine-tuning our approach for examining DNA and RNA flexibility and validating our results against experimental data. Finally, our coordinate-free method for analyzing DNA rigidity from sequence should be compared to the rigidity analysis of the corresponding DNA PDB files.

Authors' contributions. IS conceived and supervised the research. EF researched and implemented the tools, in consultation with FJ. All three authors wrote the paper.

Acknowledgments. This work was funded by NSF DMS-0714934, UBM-1129194 'Four Colleges Biomathematics Consortium', DARPA 23 Mathematical Challenges grants, and a Smith College Summer fellowship (Schultz Foundation).

4. REFERENCES

[1] N. Fox, F. Jagodzinski, Y. Li, and I. Streinu. KINARI-Web: A server for protein rigidity analysis. *Nucleic Acids Research*, 39:W177–W183, 2011.

[2] S. Fulle and H. Gohlke. Constraint counting on RNA structures: Linking flexibility and function. *Methods*, 49(2):181–188, 2009.

[3] M.A. El Hassan and C.R. Calladine. Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Phil. Trans.: Math. Phys. and Eng. Sciences*, 355:43–100, 1997.

[4] B. Hendrickson. *The molecule problem: determining conformation from pairwise distances.* PhD thesis, Cornell University, 1991.

[5] D.J. Jacobs, A.J. Rader, M.F. Thorpe, and L.A. Kuhn. Protein flexibility predictions using graph theory. *Proteins 44*, pages 150–165, 2001.

[6] D.J. Jacobs and M.F. Thorpe. Generic rigidity percolation: the pebble game. *Physics Review Letters*, 75:4051–4054, 1995.

[7] D. J. Jacobs and B. Hendrickson. An algorithm for two-dimensional rigidity percolation: the pebble game. *Journal of Computational Physics*, 137:346–365, 1997.

[8] Richardson Laboratory, Duke University. Bndlst. http://kinemage.biochem.duke.edu/software/utilities.php.

[9] G. Laman. On graphs and rigidity of plane skeletal structures. *Jour. Eng. Mathematics*, 4:331–340, 1970.

[10] A. Lee and I. Streinu. Pebble game algorithms and sparse graphs. *Discr. Math.*, 308(8):1425–1437, 2008.

[11] J.C. Maxwell. On the calculation of the equilibrium and stiffness of frames. *Philosophical Magazine Series 4*, 27:294–299, 1864.

[12] I.K. McDonald and J.M. Thornton. Satisfying hydrogen bonding potential in proteins. *Journal of Molecular Biology*, 238:777–793, 1994.

[13] N. Fox and I. Streinu Towards accurate modeling for protein rigidity analysis. In Proc. *2nd IEEE Int. Conf on Computational Advances in Bio and Medical Sciences (ICCABS'12). Feb. 23-25*, 2012.

[14] K. Range, E. Mayaan, L.J. Maher III, and D.M. York. The contribution of phosphate-phosphate repulsions to the free energy of DNA bending. *Nucleic Acids Research*, 33:1257–1268, 2005.

[15] M. Sarver, C.L. Zirbel, J. Stombaugh, A. Mokdad, and N.B. Leontis. FR3D: Finding local and composite recurrent structural motifs in RNA 3D structures. *Journal of Mathematical Biology*, 56:215–252, 2008.

[16] T.-S. Tay. Rigidity of multigraphs I *J. Comb. Theory B*, 36:95–112, 1984.

[17] M. Thorpe, M. Lei, A. Rader, D.Jacobs, L.Kuhn. Protein flexibility and dynamics using constraint theory. *J. Mol. Graph. Modeling*, 19(1):60–9, 2001.

[18] M. van Dijk and A.M.J.J. Bonvin. 3D-DART: a DNA structure modelling server. *Nucleic Acids Research*, 37 (Web Server Issue):W235–W239, 2009.

[19] J. Šponer, H.A. Gabb, J. Leszczynski, and P. Hobza. Base-base and deoxyribose-base stacking interactions in b-DNA and z-DNA: A quantum-chemical study. *Biophysical Journal*, 73:76–87, 1997.

[20] J. Wang, P. Cieplak, and P.A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21:1049–1074, 2000.

[21] M. Word, S.C. Lovell, J.S. Richardson, and D.C. Richardson. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Bio.*, 285:1735–1747, 1999.

[22] C.L. Zirbel, J.E. Šponer, J. Šponer, J. Stombaugh, and N.B. Leontis. Classification and energetics of the base-phosphate interactions in RNA. *Nucl. Acids Res.*, 37:4898–4918, 2009.