

11-26-2012

## Towards Biophysical Validation of Constraint Modeling for Rigidity Analysis of Proteins

Filip Jagodzinski  
*Central Washington University*

Ileana Streinu  
*Smith College, [istreinu@smith.edu](mailto:istreinu@smith.edu)*

Follow this and additional works at: [https://scholarworks.smith.edu/csc\\_facpubs](https://scholarworks.smith.edu/csc_facpubs)



Part of the [Computer Sciences Commons](#)

---

### Recommended Citation

Jagodzinski, Filip and Streinu, Ileana, "Towards Biophysical Validation of Constraint Modeling for Rigidity Analysis of Proteins" (2012). Computer Science: Faculty Publications, Smith College, Northampton, MA. [https://scholarworks.smith.edu/csc\\_facpubs/273](https://scholarworks.smith.edu/csc_facpubs/273)

This Conference Proceeding has been accepted for inclusion in Computer Science: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact [scholarworks@smith.edu](mailto:scholarworks@smith.edu)

# Towards Biophysical Validation of Constraint Modeling for Rigidity Analysis of Proteins

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

Ileana Streinu\*  
Smith College &  
University of Massachusetts  
Northampton, MA & Amherst, MA, USA  
streinu@{cs.smith.edu, cs.umass.edu}

## ABSTRACT

Proteins are dynamic molecules, and understanding how they flex and bend provides fundamental insights to their functions. Methods such as molecular dynamics are computationally expensive, and can simulate protein motions on limited timescales. Rigidity analysis is an alternative method, in which a protein structure is analyzed to infer which portions of the molecule are flexible. To perform rigidity analysis, a model is first constructed in which various inter-atomic stabilizing interactions are modeled according to their strength. No detailed study has been conducted as to what is the most plausible, chemically validated modeling scheme. All previous implementations have relied on heuristics, which allowed for extracting relevant observations but only for a very limited set of proteins.

We used our recently released KINARI-Web server for protein rigidity analysis to systematically vary how stabilizing interactions are modeled. Computational experiments that vary how hydrogen bonds and hydrophobic interactions are modeled to test which of them gives rigidity results that best correlate with experimental data has *not* been performed until this study. We collected a dataset of 159 Protein Data Bank files representing the wild-type and 158 variants of Lysozyme from bacteriophage T4, for which we retrieved experimentally derived stability data from the literature. We present here a systematic study seeking a possible correlation between some rigidity parameters and this experimental data. In particular, we compare rigidity results obtained from several methods for modeling inter-atomic interactions.

## Categories and Subject Descriptors

J.3 [Life and Medical Sciences]: Biology and genetics;  
I.6.5 [Simulation and Modeling]: Model Development—  
*Modeling methodologies*

\*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM-BCB'12, October 7-10, 2012, Orlando, FL, USA  
Copyright 2012 ACM 978-1-4503-1670-5/12/10 ...\$15.00.

## General Terms

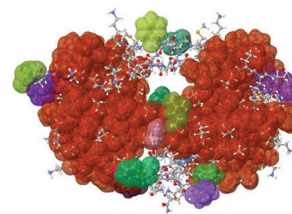
Verification

## Keywords

protein modeling, biological validation, rigidity analysis

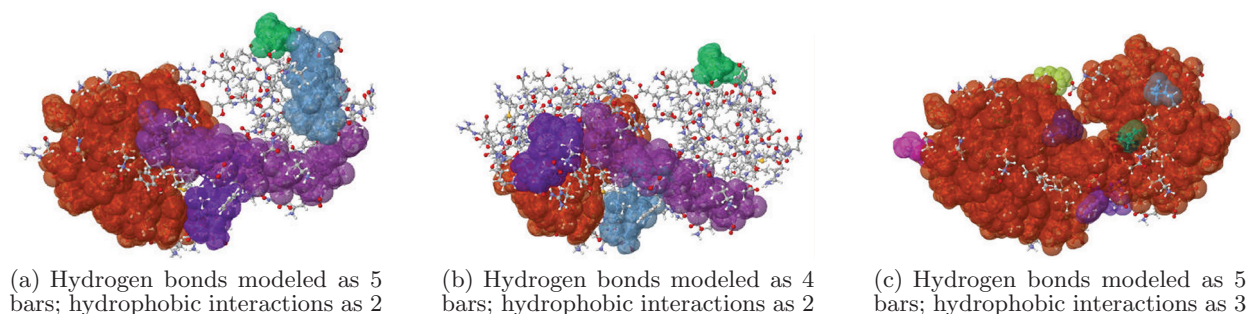
## 1. INTRODUCTION

Proteins perform their biologically roles by bending and flexing, and by interacting with other molecules. Knowing how they function is a fundamental goal in molecular biology. Because proteins cannot be observed directly, simulation methods such as molecular dynamics have been developed. A major limitation of such methods is that they rely on computationally intensive calculations of physics-based energy functions. As a result it is difficult to simulate a protein's motion on timescales that are long enough to gain insight about their functions.



**Figure 1: Rigidity analysis identifies rigid parts of a protein. Groups of colored atoms indicate rigid clusters. For HIV-1 Protease (PDB file 1hrv), most of the atoms are in a dominant rigid cluster (orange).**

Rigidity analysis is an alternative to physics-based simulation methods, that instead analyzes a single static structure of a protein. Its goal is not to predict how a molecule bends and flexes, but instead which parts of it are rigid. In performing the analysis, a molecule is modeled as a mechanical framework with bodies representing covalently bonded atoms and bars between them corresponding to the chemical constraints in the protein. The framework is used to build an associated graph, with nodes as atoms, and the bonds modeled as a number of bars depending on the strength of the bond, with a hinge representing covalent bonds. Multiple edges can exist between two nodes, with a higher number of edges indicating fewer degrees of freedom between the atoms in the corresponding protein, and hence a strong



**Figure 2: The choice of modeling of hydrogen bonds and hydrophobic interactions affects the rigidity results. When hydrophobics are modeled as 2 bars, and hydrogen bonds as 5 bars, in the mechanical framework of a protein, the dominant rigid cluster (a) is larger than when hydrophobic interactions are modeled as 2 bars, and hydrogen bonds as 4 bars (b). When the hydrophobic interactions are modeled as 3 bars, the rigidity of the protein changes drastically, resulting a much larger dominant rigidity cluster (c).**

bond. An efficient pebble game algorithm is used to analyze the graph and decompose it into rigid clusters. That information is used to infer the location of flexible parts of the protein. The rigidity results of HIV-1 Protease are shown in Figure 1, in which the majority of the atoms in the protein’s core are in a single dominant rigid cluster. We explain the modeling in more detail Section 2.

Rigidity analysis of proteins is different from physics-based simulation methods in several ways. First, it is not dependent on a force field function describing the protein-solvent system and the potential energies of all of the involved atoms. Because of this, rigidity analysis of a protein upwards of several thousand residues is performed in less than one minute, whereas even short, several nanosecond MD simulation might require days of computation time and require a computer cluster made up of hundreds or thousands or processors [11]. Moreover, rigidity analysis does not output a motion trajectory. Instead, only rigid components of a single conformation of a molecule are identified.

Rigidity analysis of proteins was initially implemented in MSU-First [5, 4] and the first online tool was *FlexWeb* [16]. These tools were used to correlate rigidity results with physical properties of several proteins, but they required case-by-case visual inspections of the biomolecules that were involved. Moreover, the choice of modeling of hydrogen bonds and hydrophobic interactions in *FlexWeb* has been determined based on the analysis of a small set of proteins. *Up until now, no systematic study has been performed with the intent to determine a universal modeling for hydrogen bonds and hydrophobic interactions.* In fact, recent work [18] has demonstrated that there is no general agreement about what the correct modeling should be, so that rigidity results for a large protein dataset correlate with biological phenomena.

We have recently developed and made available KINARI-Web [2]. It is a second generation free online server for protein rigidity analysis that allows for different ways of modeling chemical interactions. Previous developed software MSU-First and *FlexWeb* do not provide options for adjusting how hydrogen bonds and hydrophobic are modeled. As is shown in Figure 2, even small changes to how hydrogen bonds and hydrophobic interactions are modeled can drastically alter the rigidity results.

*Our goal is to correlate rigidity results of proteins to experimentally derived biological data from the literature, in*

*the form of  $\Delta\Delta G$ .* This experimental data value designates whether a variant of a protein is stable, and we use this score as the ground-truth. In this paper, we investigate which set of modeling options for hydrogen bonds and hydrophobic interactions produce rigidity results that correlate to this biological data. To do this, we have retrieved from the PDB the structure files of 158 variants and the wild-type of Lysozyme from bacteriophage T4. For each variant, we searched the ProTherm [7] database for its  $\Delta\Delta G$  score. We systematically varied how hydrogen bonds and hydrophobic interactions are modeled during rigidity analysis of the 159 protein structures.

## 2. BACKGROUND

Here we briefly review rigidity analysis, and summarize how proteins are modeled in KINARI. A more thorough description of how proteins are modeled in KINARI is presented in [2]. We give a brief historical perspective of rigidity analysis of proteins, and summarize several previous studies. We also sketch a brief overview of  $\Delta\Delta G$ .

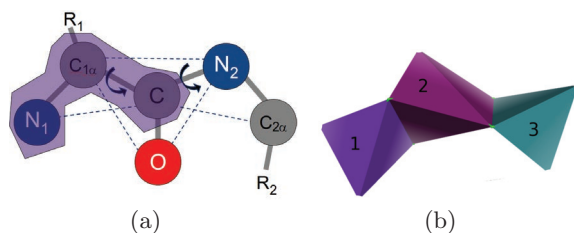
### 2.1 Rigidity Analysis

A geometric and combinatorial approach to the analysis of protein rigidity associates atoms and their chemical interactions to the nodes and rigid bars of a network. The study of rigidity and flexibility of these bar-and-joint frameworks was developed by 19th century engineers attempting to analyze cross-bracing of steel structures. In 1864 James Clerk Maxwell [10] identified a simple counting rule to determine the rigidity of such structures. This counting rule was proven correct in 2 dimensions by Laman [8], and subsequently was modified for the analysis of 3-dimensional structures [14]. These structures are analyzed with efficient algorithms, based on the pebble game paradigm [5, 9], and applied to mechanical models.

### 2.2 Mechanical Modeling of Proteins

The KINARI software relies on 3-dimensional structures called *body-bar-hinge frameworks* to model the mechanics of proteins. Atoms along with their covalently bonded neighbors form bodies. Covalent bonds between bodies are modeled as a hinge, and other stabilizing interactions such as hydrogen bonds and hydrophobics are modeled as hinges or

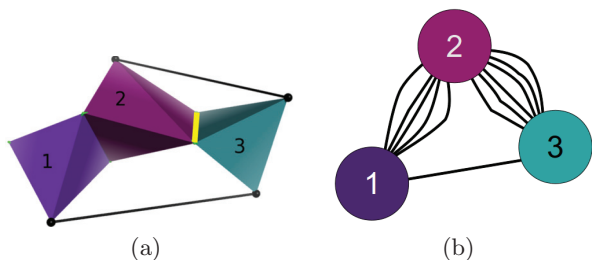
bars. In Figure 3 we show a schematic of a protein, and how the mechanical model is constructed.



**Figure 3:** Solid lines designate covalent bonds, and dashed lines represent distance constraints that arise due to angle constraints imposed by atom interactions (a). Covalently bonded atoms form bodies (for example the purple region). The *body-bar-hinge framework* for (a) is shown in (b), where body 2 is composed of  $C_{1\alpha}$ , C, O, and  $N_2$ , while body 3 is composed of C,  $N_2$ , and  $C_{2\alpha}$ . A hinge between two bodies allows for a one-degree-of-freedom rotation of one body about the other, along the hinge axis.

### 2.3 Pebble Game Rigidity Analysis

Using the body-bar-hinge framework we build an associated graph, with a node for each body, an edge for each bar, and 5 bars for each hinge. The pebble game algorithm decomposes this graph into clusters which correspond to rigid components in the framework (Figure 4).



**Figure 4:** In (a) we show an abstract model illustrating how bars would be placed between bodies, if hydrogen bonds or other stabilizing interactions existed between them. The graph (b) is built by associating each rigid body to a node, each hinge to 5 bars, and each bar to an edge. Bodies 2 and 3 have 6 bars between them, 5 for the hinge (shown in yellow in (a)), and 1 for the bar.

The algorithm starts with 6 pebbles on each vertex of the associated graph, and reasons about the edges one at a time, and accepts or rejects them. To be accepted, an edge must have at least 7 pebbles distributed somehow on its two endpoints. If not enough pebbles are present, they are collected using a depth-first search approach. An accepted edge consumes one pebble. As more and more edges are accepted, they are combined into rigid components. The algorithm ends when all edges have been considered. A formal proof of correctness for this algorithm can be found [9].

### 2.4 Rigidity Based Protein Flexibility

Rigidity analysis of protein structures was demonstrated by Thorpe, *et al.* [16]. They studied different states of HIV-1 protease and showed that the rigid clusters in open and closed conformations of the protease are correlated with the known mechanical properties of the molecule.

Rader *et al.* [12] simulated the unfolding of rhodopsin, a trans-membrane receptor, by performing a *rigidity dilution analysis* using the FIRST software [15]. In this method, hydrogen bonds are removed from the molecular model one after another, from weakest to strongest, and rigidity analysis is performed on the model after each removal. A “folding core” is identified when there exists only one rigid cluster with at least three residues of two or more secondary structures. The computed core was correlated with experimental results, and confirmed via a visual inspection of the protein using insights of its physical properties.

Rigidity theory has also been used to investigate the possible motions and to gain insights into the structural stability of proteins. Rader *et al.* [13] have combined elastic network models with rigidity analysis of constraint networks for freely rotating rods to predict protein folding nuclei. They verified their method against data that was attained from native state hydrogen-deuterium exchange experiments.

In a comparative study of the rigidity analysis of 62 protein structures from six different protein families, Wells *et al.* [18] demonstrated that the main-chain rigidity of a protein is very sensitive to small structural variations. In that study, Wells concluded that the modeling of hydrogen bonds needs to be chosen carefully so that specific hypotheses about the rigidity of particular proteins can be formed.

Recently, Fox, *et al.* [3], used a benchmark dataset of 32 PDB structure files to validate the modeling in KINARI against a dataset that was analyzed using the Gerstein Lab’s RigidFinder algorithm [1]. Fox introduced a metric called the cluster decomposition score to compare KINARI’s rigidity results against Gerstein’s structural predictions. They found that the sensitivity of the cluster decomposition score is dependent on the choice of the hydrogen bond energy cut-off value, which designates a threshold at which these bonds are retained in the molecular model.

In recent work, the KINARI-Web software has also been used to help formulate hypotheses on the role of individual residues, and whether their mutation to glycine is destabilizing [6]. In that work, *in silico* mutations to glycine were performed when making the model of the protein, and the rigidity results were correlated against the stabilities of the proteins which had equivalent mutations performed in the physical protein.

### 2.5 $\Delta\Delta G$ as a Measure of Protein Stability

Concentrations of a protein in different states can be experimentally calculated in several ways. One such method is circular dichroism (CD) spectroscopy, which measures the differences in absorption patterns of left and right-handed polarized light. The CD spectrum of a protein in the near ultraviolet spectral region can be sensitive to certain aspects of tertiary and secondary structures. Thus, an analysis of the CD spectrum can be used to determine the presence of secondary structures. If a protein is known to have secondary structures such as alpha-helices, and if an analysis of the CD spectrum reveals that there are no alpha-helices in the sample being studied, then the protein can be assumed



to be denatured, or unfolded. The equilibrium constant can be used to determine the conformational stability  $\Delta G$  of a protein.

If  $\Delta G$  data is available for different conformations of the same protein, that information can be used to infer the relative stabilities of the two protein structures. The delta of the Gibbs free energy,  $\Delta\Delta G$ , can be used to measure the stability of a variant against a reference protein (nearly always the wild-type protein). The lower the  $\Delta\Delta G$  value, the more unstable is the variant. Thus,  $\Delta\Delta G$  is an experimentally derived score, which we use as the ground-truth when comparing the stability of variants of a protein structure.

### 3. METHODS

We present our setup, starting with a description of the dataset of 158 variants and wild-type protein data files from the PDB, and their  $\Delta\Delta G$  measurements. We present several methods used for modeling hydrogen bonds and hydrophobic interactions. Finally, we show the results obtained in seeking a correlation between the rigidity analysis results and the  $\Delta\Delta G$  values.

#### 3.1 Selection of PDB files and their $\Delta\Delta G$ data

We used the ProTherm Database [7], which catalogs  $\Delta\Delta G$  values for variants (also called mutants) of proteins. We identified 158 PDB entries for variants of Lysozyme from bacteriophage T4, for which there is available  $\Delta\Delta G$  data. For all of these 158 entries, the reference protein (the wild-type, non-mutated form) was the protein structure in PDB file 2lzm. The  $\Delta\Delta G$  values of these protein mutants ranged from very negative, meaning that a mutant was much less stable than the wild-type, to positive, indicating that the mutant was more stable than the wild-type.

#### 3.2 Rigidity Analysis of Lysozyme Structures

Using the KINARI software, we analyzed the rigidity of each of the 158 variants. For each protein, single and double covalent bonds were modeled in the associated graph as 5 bars and 6 bars, respectively. This modeling represents that single covalent bonds impose one degree of freedom between the corresponding atoms in the mechanical model of the protein, which is equivalent to allowing rotation along the bond, and that the double covalent bonds retain zero degrees of freedom and do not permit rotation.

Because there is no agreed upon way of performing the mechanical modeling of hydrogen bonds and hydrophobic interactions, we used KINARI's customizable modeling feature to systematically vary how the modeling of these constraints. In our software, hydrophobic interactions and hydrogen bonds can each be modeled 7 different ways: 1 through 6 bars, or as hinges. In addition, we have also implemented in the KINARI software a method to model hydrogen bonds according to their energies. We designated several ways of modeling these bonds, so that the stronger ones (having a lower kCal/mol value) were modeled as more bars in the graph of the molecule's mechanical model. We tried these different modeling schemes with the goal to identify which of them would be effective in modeling what happens in nature. Table 1 lists the 7 ways that hydrogen bonds were modeled according to energy, in our computational experiments.

Because hydrophobic interactions can be modeled 7 different ways, and hydrogen bonds can be modeled 14 ways (7 according to energy, and 7 without), the total modeling

**Table 1: Modeling of hydrogen bonds according to their energies**

Method	Energy of hydrogen bond (kcal/mol)	How modeled
ByEnergy1	strength < -5	6 Bars
	-5 < strength < -4	Hinges
	-4 < strength < -3	4 Bars
	-3 < strength < -2	3 Bars
	-2 < strength < -1	2 Bars
ByEnergy2	-1 < strength	1 Bar
	strength < -4	Hinges
	-4 < strength < -1	4 Bars
ByEnergy3	-1 < strength	3 Bars
	strength < -2	4 Bars
ByEnergy4	-2 < strength	Hinges
	strength < -3	4 Bars
ByEnergy5	-3 < strength	Hinges
	strength < -4	4 Bars
ByEnergy6	-4 < strength	Hinges
	strength < -5	4 Bars
ByEnergy7	-5 < strength	Hinges
	strength < -6	4 Bars
	-6 < strength	Hinges

combinations for hydrogen bonds and hydrophobic combinations is  $7 \times 14 = 98$ . Thus, each protein was analyzed 98 different ways.

#### 3.3 Correlating Rigidity Results to $\Delta\Delta G$

Now we come to the main method in our paper, where we correlate our rigidity results to the experimentally derived  $\Delta\Delta G$  values that we retrieved from the literature.

To evaluate the 98 different ways of modeling hydrogen bonds and hydrophobic interactions, we used two rigidity metrics as indicators of a molecule's stability. The first rigidity metric that we calculated for each variant is the size of the dominant rigid cluster (DRC), which is the number of atoms in the largest rigid cluster. The second is the Cluster Configuration Entropy (CCE) [17]. CCE is a function of the probability that a vertex in the mechanical model is part of a cluster of size  $s$ . To compute CCE, a normalized cluster number,  $n_s$  is defined as the number of clusters of size  $s$  divided by the total number of vertices in the mechanical model. The probability that a vertex belongs to an  $s$ -cluster,  $w_s$ , and the CCE value of the entire mechanical model are given as the following:

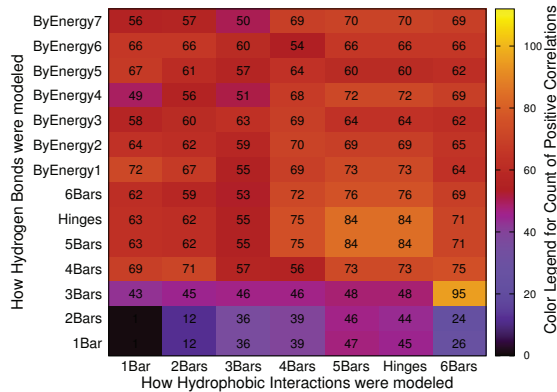
$$w_s = \frac{sn_s}{\sum_s sn_s} \quad CCE = - \sum_s w_s \ln w_s \quad (1)$$

For two conformations of a protein, the one with the higher CCE value is more disordered, and hence is more unstable.

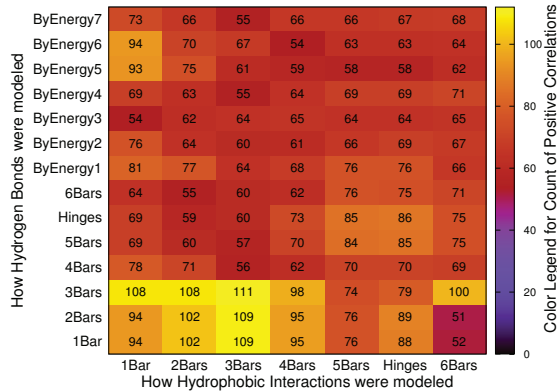
We compared the size of the dominant cluster of the files in our dataset to the size of the dominant cluster of the wild-type protein structure in PDB file 2lzm. Each variant had thus a larger or smaller dominant rigid cluster relative to the wild type. For example, if the dominant rigid cluster of the variant contained more atoms than the dominant rigid cluster of the wild-type protein, then we designated that variant as more stable than the wild-type protein. If the variant's

dominant rigid cluster had fewer atoms (was smaller) than the wild-type’s dominant rigid cluster, then the variant was designated as less stable than the wild-type. Likewise, we compared the CCE value of the variants and the wild-type.

To correlate our choices of modeling for hydrogen bonds and hydrophobic interactions to experimental biological data, for both the DRC and CCE metrics, we tallied how many of the 98 rigidity analyses of each of the 158 variants qualitatively correlated with their  $\Delta\Delta G$  values. Using this count, we ranked our 98 modeling methods. For example, a modeling method that positively correlated with the  $\Delta\Delta G$  data in 140 of the 158 variants was designated as better than a modeling method that positively correlated with only 80 of the 158 variants. Our general method for evaluating each of the 98 ways that hydrogen bonds and hydrophobic interactions were modeled is sketched in Algorithm 1.



(a) Correlating Dominant Rigid Cluster with  $\Delta\Delta G$



(b) Correlating Cluster Config. Entropy with  $\Delta\Delta G$

**Figure 5:** We performed rigidity analysis of our 158 variants 98 different ways (7 ways of modeling hydrophobic interactions and 14 ways of modeling hydrogen bonds (Section 3.2). In (a), the number in each of the grid’s  $x,y$  box indicates for how many of the 158 proteins did the size of the Dominant Rigid Cluster (DRC) positively correlate with the experimentally derived stability data. In (b), the number in each of the grid’s  $x,y$  box indicates positive correlation for Cluster Configuration Entropy. Boxes whose color tends towards yellow designate the modeling choices whose rigidity results correlated best with  $\Delta\Delta G$  data.

---

**Algorithm 1:** Measuring Quantitative Correlation of Rigidity DRC with Experimental  $\Delta\Delta G$ . HP=Hydrophobic Interaction, HB=Hydrogen Bond, wt=wild-type.

---

```

Input: Rigidity Results of proteins
Input:  $\Delta\Delta G$  values for all protein-mutant pairs
Initialize correlationCorrectCounts[][] matrix to all 0s
foreach  $i^{th}$  protein analyzed do
  foreach  $j$  modeling of HP do
    foreach  $k$  modeling HB do
       $DRC_{var}$  = size of DRC for  $i$  when modeling  $j, k$ 
       $DRC_{wt}$  = size of DRC for wt when modeling  $j, k$ 
       $ddg$  =  $\Delta\Delta G$  for wt,  $i$  pair
      if  $DRC_{var} < DRC_{wt}$  then
        if  $ddg < 0$  then
          correlationCorrectCounts[ $j$ ][ $k$ ] ++
        end
      else
        if  $ddg > 0$  then
          correlationCorrectCounts[ $j$ ][ $k$ ] ++
        end
      end
    end
  end
end
Output: correlationCorrectCounts matrix

```

---

Using the results of Algorithm 1 allowed us to rank the different ways that we modeled stabilizing interactions in our software. We show for how many of the 158 analyzed proteins did Dominant Rigid Cluster (Figure 5, top), and Cluster Configuration Entropy (Figure 5, bottom) positively correlate with the experimental biological data.

## 4. DISCUSSION

Based on the analysis of our method of systematically varying how hydrogen bonds and hydrophobic interactions are modeled, we observed several behaviors for both of the DRC and CCE rigidity metrics.

There is not one single choice of modeling of hydrogen bonds and hydrophobic interactions that would have given DRC and CCE metrics that correlated positively for all 158 proteins with  $\Delta\Delta G$  data. This is corroborated by the results of Wells, *et al* [18], who found that the choice of modeling of hydrogen bonds and selecting of the hydrogen bond cutoff needs to be chosen on a protein case-by-case basis so that rigidity results can be verified against biological data. One possible explanation for this is that the size of the dominant rigid cluster or the Cluster Configuration Entropy are not the best metrics as an indicator of a protein’s stability. Instead, correlating rigidity results to  $\Delta\Delta G$  data might require a multi-dimensional analysis dependent on these two, or even more, rigidity metrics.

However, there are several modeling choices for stabilizing interactions that would have generated rigidity results that would have correctly predicted the stability of at least 50% of the 158 protein variants. For the DRC metric, when hydrogen bonds were modeled as 3 bars, and hydrophobic interactions as 6 bars, then the rigidity results would have positively correlated with  $\Delta\Delta G$  data in 95 of the 158 vari-

ants. Note also that when hydrogen bonds were modeled as 5 bars, and hydrophobic interactions as 2 bars (the way that these interactions are equivalently modeled in the *FlexWeb* server), then the stability of only 62 out of the 158 variants would have been predicted correctly. In the case of CCE, there are 8 combinations of modeling hydrogen bonds and hydrophobic interactions so that the rigidity metric would have positively correlated with  $\Delta\Delta G$  data in at least 100 of the 158 variants. The stability of the most number of variants, 111, would have been correctly identified using the CCE metric when hydrogen bonds and hydrophobic interactions were modeled as 3 bars.

Although we did not identify a single choice of modeling of hydrogen bonds and hydrophobics which generated rigidity results that positively correlated with  $\Delta\Delta G$  data in all of our variants, we have demonstrated the use of our method (Algorithm 1) in correlating rigidity metrics to experimental data. In addition, we have shown that there are several combinations of modeling hydrogen bonds and hydrophobic interactions so that the Cluster Configuration Entropy metric correlates better with experimental data than the Dominant Rigid Cluster metric. Moreover, our method is not dependent on a case-by-case analysis of the studied proteins, but instead requires only experimental data (here  $\Delta\Delta G$ ), and rigidity metrics. As such, we believe that our general method can be used with other rigidity metrics and experimental data, as well as on other protein datasets.

## 5. CONCLUSIONS

Rigidity analysis of protein structures identifies regions of a molecule that are flexible. In rigidity analysis, a protein's atoms and chemical interactions are used to build a mechanical model, which is associated to a graph composed of nodes that represent atoms, and edges that correspond to chemical constraints. A systematic study to determine how hydrogen bonds and hydrophobic interactions should be modeled so that the rigidity results correlate with experimental data has not been performed prior to this work.

We have used our freely-available KINARI-Web server to systematically vary the way that hydrogen bonds and hydrophobic interactions were modeled in 158 structures of variants of Lysozyme from bacteriophage T4. In correlating two rigidity metrics for each of the proteins against  $\Delta\Delta G$  data, we have found that there is no one single "best" choice of modeling hydrogen bonds and hydrophobic interactions. However, for our dataset, there were a few modeling schemes so that the rigidity metrics for more than 100 of the 158 variants correlated against  $\Delta\Delta G$  data.

## 6. ACKNOWLEDGMENTS

This work was funded by NSF CCF-1016988, DMS-0714934, and DARPA 23 Mathematical Challenges grants of IS. The work was performed by FJ while a PhD student at UMass.

## 7. REFERENCES

- [1] A. Abyzov, R. Bjornson, M. Felipe, and M. Gerstein. RigidFinder: a fast and sensitive method to detect rigid blocks in large macromolecular complexes. *Proteins*, 78(2):309–324, February 2010.
- [2] N. Fox, F. Jagodzinski, Y. Li, and I. Streinu. KINARI-Web: A server for protein rigidity analysis. *Nucleic Acids Research*, 39 (Web Server Issue):W177–W183, 2011.
- [3] N. Fox and I. Streinu. Towards accurate modeling for protein rigidity analysis. In *2nd IEEE International Conference on Computational Advances in Bio and Medical Sciences*, Feb. 23-25, February 2012.
- [4] 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.
- [5] D.J. Jacobs and M.F. Thorpe. Generic rigidity percolation: the pebble game. *Physics Review Letters*, 75:4051–4054, 1995.
- [6] F. Jagodzinski, J. Hardy, and I. Streinu. Using rigidity analysis to probe mutation-induced structural changes in proteins. *Journal of Bioinformatics and Computational Biology*, 10(3), 2012.
- [7] M.D. Kumar, K.A. Bava, M.M. Gromiha, P. Prabakaran, K. Kitajima, H. Uedaira, and A. Sarai. Protherm and pronit: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Research*, 34(suppl 1):D204–D206, 2005.
- [8] G. Laman. On graphs and rigidity of plane skeletal structures. *Journal of Engineering Mathematics*, 4:331–340, 1970.
- [9] A. Lee and I. Streinu. Pebble game algorithms and sparse graphs. *Discrete Mathematics*, 308(8):1425–1437, 2008.
- [10] J.C. Maxwell. On the calculation of the equilibrium and stiffness of frames. *Philosophical Magazine Series 4*, 27:294 – 299, 1864.
- [11] V. Pande, I. Baker, J. Chapman, S.P. Elmer, S. Khaliq, S.T. Larson, Y.M. Rhee, M.R. Shirts, C.D. Snow, E.J. Sorin, and B. Zagrovic. Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers*, 68(1):91–109, 2003.
- [12] A.J. Rader, G. Anderson, B. Isin, I. Bahar, and J. Klein-Seetharaman. Identification of core amino acids stabilizing rhodopsin. *Proceedings of National Academy of Sciences, U.S.A.*, 101:7246–7251, 2004.
- [13] A.J. Rader and I. Bahar. Folding core predictions from network models of proteins. *Polymer*, 45(2):659–668, 2004.
- [14] T.-S. Tay. Rigidity of multigraphs I: linking rigid bodies in n-space. *Journal of Combinatorial Theory, Series B*, 36:95–112, 1984.
- [15] M.F. Thorpe, M.V. Chubynsky, B.M. Hesperheide, S. Menor, D.J. Jacobs, L.A. Kuhn, M.I. Zavodszky, M. Lei, and A.J. Rader W. Whiteley. *Flexibility in Biomolecules*, chapter 6, pages 97–112. Current Topics in Physics. Imperial College Press, London, 2005. R.A. Barrio and K.K. Kaski, eds.
- [16] M.F. Thorpe, M. Lei, A.J. Rader, and D.J. Jacobs L.A. Kuhn. Protein flexibility and dynamics using constraint theory. *Journal of Molecular Graphics and Modeling*, 19(1):60–9, 2001.
- [17] I.R. Tsang and I.J. Tsang. Cluster size diversity, percolation, and complex systems. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 60:2684–2698, 1999.
- [18] S.A. Wells, J.E. Jimenez-Roldan, and R.A. Romer. Comparative analysis of rigidity across protein families. *Physical Biology*, 6(4), 2009.