

6-1-2012

Using Rigidity Analysis to Probe Mutation-Induced Structural Changes in Proteins

Filip Jagodzinski

University of Massachusetts Amherst

Jeanne Hardy

University of Massachusetts Amherst

Ileana Streinu

University of Massachusetts Amherst, istreinu@smith.edu

Follow this and additional works at: https://scholarworks.smith.edu/csc_facpubs



Part of the [Computer Sciences Commons](#)

Recommended Citation

Jagodzinski, Filip; Hardy, Jeanne; and Streinu, Ileana, "Using Rigidity Analysis to Probe Mutation-Induced Structural Changes in Proteins" (2012). Computer Science: Faculty Publications, Smith College, Northampton, MA.

https://scholarworks.smith.edu/csc_facpubs/275

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

USING RIGIDITY ANALYSIS TO PROBE MUTATION-INDUCED STRUCTURAL CHANGES IN PROTEINS

FILIP JAGODZINSKI^{*,§}, JEANNE HARDY^{†,¶} and ILEANA STREINU^{*,‡,||}

^{*}*Department of Computer Science, 140 Governors Drive
University of Massachusetts Amherst, Amherst, MA 01002, USA*

[†]*Department of Chemistry, 104 Lederle Graduate Research Tower
University of Massachusetts Amherst, Amherst, MA 01002, USA*

[‡]*Department of Computer Science, Smith College
Northampton, MA 01063, USA*

[§]filip@cs.umass.edu

[¶]hardy@chem.umass.edu

^{||}streinu@cs.smith.edu

^{||}streinu@cs.umass.edu

Received 13 February 2012

Revised 9 April 2012

Accepted 11 April 2012

Published 12 June 2012

Predicting the effect of a single amino acid substitution on the stability of a protein structure is a fundamental task in macromolecular modeling. It has relevance to drug design and understanding of disease-causing protein variants. We present KINARI-Mutagen, a web server for performing *in silico* mutation experiments on protein structures from the Protein Data Bank. Our rigidity-theoretical approach permits fast evaluation of the effects of mutations that may not be easy to perform *in vitro*, because it is not always possible to express a protein with a specific amino acid substitution. We use KINARI-Mutagen to identify critical residues, and we show that our predictions correlate with destabilizing mutations to glycine. In two in-depth case studies we show that the mutated residues identified by KINARI-Mutagen as critical correlate with experimental data, and would not have been identified by other methods such as Solvent Accessible Surface Area measurements or residue ranking by contributions to stabilizing interactions. We also generate 48 mutants for 14 proteins, and compare our rigidity-based results against experimental mutation stability data. KINARI-Mutagen is available at <http://kinari.cs.umass.edu>.

Keywords: Protein rigidity; mutations; stability prediction.

1. Introduction

In this paper, we present KINARI-Mutagen, a web application for performing computational mutation experiments using rigidity and flexibility analysis.

^{||} Corresponding author.

Mutations in proteins. A mutation in a protein's amino acid sequence can have deleterious effects on its stability and function. A number of diseases result from single point mutations. Hence knowing their effect can be used to guide the design of drugs aimed at combating those disorders. To predict and better understand the roles of mutations, the genetic information that codes for the amino acid sequence of a protein can be altered, and the expressed mutant proteins can be analyzed to infer the impact of the specific mutation. Such studies are aided by several widely used molecular biology techniques, such as site-directed mutagenesis.¹ Unfortunately, such experiments are often labor- and time-intensive. The possible number of mutants that can be made from even the smallest proteins makes exhaustive mutagenesis studies impractical. For example, 20^{100} mutants can in principle be engineered for a 100-residue protein using the 20 naturally occurring amino acids.

Rigidity analysis of proteins. Flexibility information can be obtained through several computational methods. Here we focus on rigidity analysis as implemented in our software KINARI,² which calculates the rigid regions of a protein structure. The premise is that the protein's function is directly correlated with its distribution and sizes of rigid clusters, and destabilizing any of them will have an observable effect. Rigid clusters are groups of atoms whose pair-wise distances are determined by inter-atomic interactions, such as covalent bonds, angle constraints, and other types of interactions (e.g. hydrogen bonds). KINARI uses an efficient combinatorial algorithm to quickly compute the rigid clusters, and does not rely on expensive all-atom energy calculations. Figure 1 shows the identified rigid regions in the protein lysozyme from bacteriophage T4.

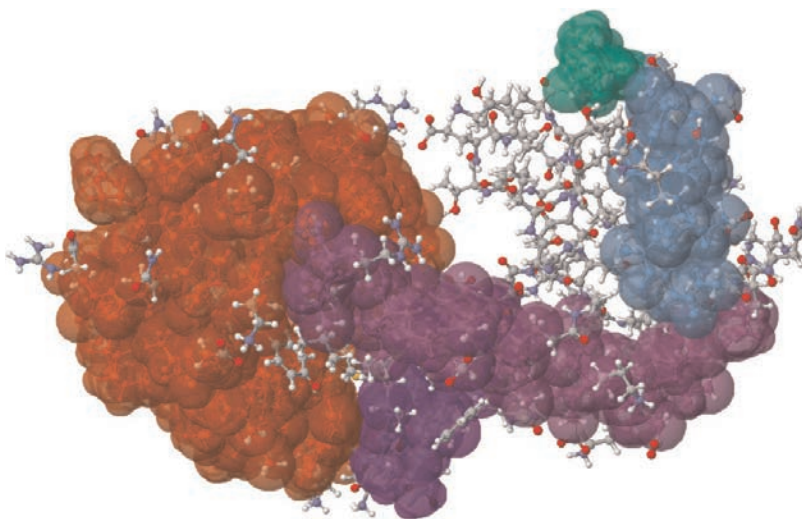


Fig. 1. Rigidity analysis performed on Lysozyme from bacteriophage T4 (PDB ID 2zlm). The color bodies indicate clusters of atoms that are rigid, as identified by an efficient pebble game algorithm.

Our contribution: KINARI-Mutagen. We extend KINARI to generate mutant protein structures and analyze their rigidity. We present here the first release of this new tool, KINARI-Mutagen. Its ultimate goal is to identify destabilizing mutations. This first version demonstrates that even the simplest type, a mutation to a glycine (called here an *excision*), yields valuable information.

In an excision, a mutated residue has the hydrogen bonds and hydrophobic interactions of its side-chain removed from the molecular model on which rigidity analysis is performed. This is a simplified version of a mutation to a glycine. KINARI-Mutagen can be used to answer two types of questions: (1) will mutating a residue in a protein destabilize it, or (2) given a protein, which residues could destabilize a protein if mutated? The stability of a protein is modeled in our software with rigidity theory. We demonstrate our software’s usefulness in two case studies, which show that the mutated residues identified by KINARI – Mutagen as critical correlate with experimental data, and would not have been identified by other methods such as Solvent Accessible Surface Area measurements or residue ranking by contributions to stabilizing interactions.

2. Background and Related Work

Here we review previous work that addressed the effect of mutations on the structure of a protein. We briefly introduce rigidity analysis, and we summarize previous rigidity-stability studies. We sketch how the KINARI software models proteins.

2.1. Mutations affect protein structure and function

Deoxyribonucleic acid (DNA) contains the instructions on how amino acids should be joined during protein synthesis to make a protein. If there is an error in the process, the resulting amino acid sequence may differ from the most common sequence of amino acids, which is designated the **wild-type** version of that protein. A protein with mutations is called a **mutant**. Mutant proteins contribute to many genetic diseases. For example, single point mutations in the cystic fibrosis transmembrane conductance regulator protein lead to the development of cystic fibrosis. In the protein α -galactosidase, there are over 190 single point mutations that lead to development of Fabry Disease.³ Thus understanding the effect of point mutations is of biomedical importance.

A mutation in the amino acid sequence can inhibit the protein’s function. Alber *et al.*⁴ have found that temperature-sensitive mutations often occur at residues which are structurally important. Similarly, a mutation at a residue location that plays a crucial role can render a protein inoperative.⁵ However, because not all mutations are equally disruptive, it is important to know how a mutation will affect the protein.

2.2. Predicting the effects of mutations

One way in which the role of a residue substitution can be directly studied is by mutation experiments in the physical protein. Matthews *et al.* have designed and

analyzed many mutants of lysozyme from the bacteriophage T4. When core residues in lysozyme were substituted by alanine, an analysis of the crystal structures revealed that the unoccupied volume in some of the mutants underwent a collapse, while other mutants formed an empty cavity.⁶ Residues of T4 lysozyme with high mobility or high solvent accessibility were shown to be much less susceptible to destabilizing substitutions. The authors concluded that residues that are held relatively rigidly within the core of the protein make the largest contribution to the protein's stability.⁷ The magnitude of the contributions of various substitutions to the thermodynamic stability of proteins can be directly measured. Substitutions result in destabilization of up to 2.7 to 5.0 kcal/mol.⁸

Although the studies by Matthews and others provide precise, experimentally verified insight into the role of a residue based on its mutation, such studies are time-consuming and often cost-prohibitive. Moreover some mutant proteins cannot be expressed, due to dramatic destabilization caused by the mutation, and so only a small subset of all possible mutations can be studied explicitly. To address this, computational and analysis techniques have been proposed.

In computational experiments by Lee and Levitt,⁹ the side-chains in each of 78 structures of mutant proteins were perturbed. A heuristic energy measure, E_{calc} , was used to predict the stability of each protein, and compared to known activity data. Gilis and Rooman¹⁰ estimated the folding free energy changes upon mutations using database-derived potentials, and concluded that hydrophobic interactions contribute most to the stabilizing of the protein core. Similarly, Prevost *et al.*¹¹ have used molecular dynamics simulations to study the effect of mutating Barnase residue Isoleucine 96 to alanine, and predicted that the major contributions to the free energy difference arose from non-bonded interactions.

Machine learning and statistical methods have also been developed to help predict the effects of mutations. Cheng *et al.*¹² used Support Vector Machines to predict with 84% accuracy the sign of the stability change for a protein induced by a single-site mutation. However, their online tool MUpro only outputs whether a mutation is expected to stabilize or destabilize a protein, and does not provide data that can be used to rank residue mutations based on their impact on the protein's stability. Also, data of amino acid replacements that are tolerated within families of homologous proteins has been used to devise stability scores for predicting the effect of residue substitutions,¹³ which has been extended and implemented into an online web server.¹⁴ It is not clear, however, how the use of environmental substitution data to devise a score for the effect of a mutation is appropriate if no such data exists, or if a newly discovered protein has few homologues.

Thus, progress has been made in predicting the effects of mutations on protein stability. However, many such methods rely on computationally intensive energy calculations or are not able to infer the role of a single amino acid in stabilizing a protein's structure. To complement these already existing methods, we seek to apply rigidity concepts to the computational prediction and analysis of the stability of mutant protein structures.

2.3. Rigidity based protein flexibility: Background

Jacobs *et al.*¹⁵ and Thorpe *et al.*¹⁶ were the first to apply rigidity analysis to proteins using a different type of pebble game (bar and joint). They studied HIV-1 protease and correlated the rigidity results with the known mechanical properties of the molecule. Rader *et al.*¹⁷ simulated a thermal unfolding of rhodopsin, and introduced the dilution analysis method. It identifies a folding, which was correlated with experimental results. More recent work with rigidity theory investigated possible motions and structural stability. However, none of the existing studies used rigidity analysis to try to predict the effects of mutations.

2.4. Mechanical modeling of proteins

Originally, rigidity theory was applied to protein models obtained by associating a network of nodes (atoms) connected by rigid bars, corresponding to bonds and other stabilizing interactions. The study of rigidity and flexibility of these *bar-and-joint frameworks* has a long history going back to the 19th century, when a simple counting rule was identified by James Clerk Maxwell.¹⁸

We use a different modeling with three-dimensional structures called *body-bar-hinge frameworks*.¹⁹ This is associated to molecules as follows: an atom together with its covalently bonded neighbors forms a body. A rotatable covalent bond shared between two such bodies acts as a hinge. Hydrogen bonds can be modeled with combinations of hinges and bars. An example is shown in Fig. 2.

2.5. Pebble game rigidity analysis

To a body-bar-hinge framework we associate a graph, with a node for each body, an edge for each bar, and five bars for each hinge. The pebble game algorithm decomposes this graph into clusters which correspond to rigid components in the framework.

The algorithm starts with six pebbles on each vertex of the graph, considers the edges one at a time, and accepts or rejects them. To be accepted, an edge must have

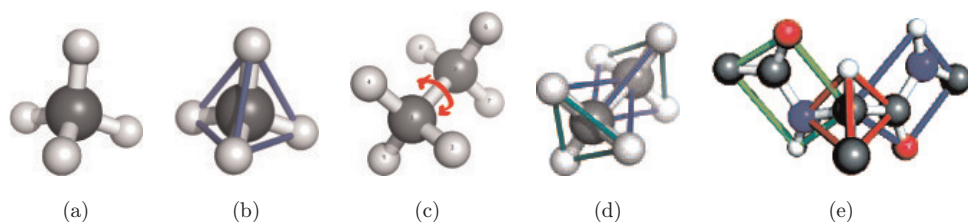


Fig. 2. Methane (a) is rigid because all pair-wise distances between atoms are fixed (b). In ethane (c), a carbon atom (gray) and its bonded neighbor atoms form a rigid body. The two bodies share a hinge along the center C–C bond, shown as an abstract *body-bar-hinge framework* (d). A protein's peptide units are modeled as rigid bodies (e).

at least seven pebbles distributed somehow on its two endpoints. If not enough pebbles are present, an attempt is made to collect them using a depth-first search approach. An accepted edge consumes one pebble. As more edges are accepted, they are grouped into rigid components. The algorithm ends when all edges have been considered. Formal proofs can be found in Ref. 20, and a Java applet demonstrating the algorithm is available at <http://linkage.cs.umass.edu/pg>.

3. System Description and Analysis Tools

KINARI-Mutagen investigates how different residues affect the rigidity and stability of a protein. Analyzing a protein involves four phases: (1) downloading and curating a PDB file, (2) performing excision to generate mutants, (3) analyzing the rigidity of each mutant, and (4) aggregating the results to help the user identify critical residues. KINARI-Mutagen provides a direct link to KINARI-Web,² for downloading a PDB file. Chains, ligands and water molecules in the protein can be retained or removed, and covalent and non-covalent interactions are identified.

The KINARI Mutation Engine generates mutants. It performs a simple computational mutation, where a residue is converted to a glycine. For the purpose of performing the rigidity analysis, it is not necessary to alter the positions of, or remove, atoms. Instead, it suffices to remove the side-chain's hydrogen bonds and hydrophobic interactions from the protein's molecular framework. This functions in our model like the removal of a side-chain. Subsequent versions of the Mutation Engine will permit increasingly advanced mutation functions. Because rigidity analysis is efficient, many generated mutant protein structures can be analyzed quickly.

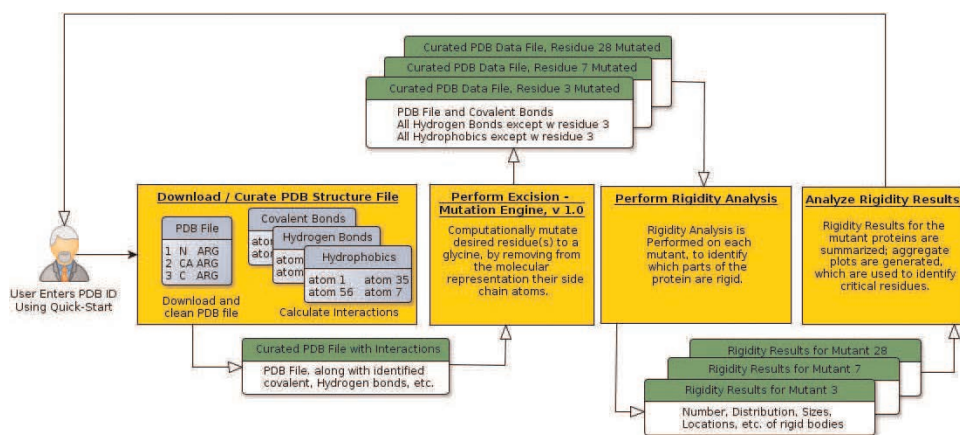


Fig. 3. KINARI-Mutagen downloads a PDB file, perform excision to generate mutants, analyze their rigidity, and aggregates the rigidity results. The generated plots and metrics provide information about which residues are critical in maintaining the protein's rigidity. Shown here is the procedure for performing excision on residues 3, 7 and 28 to generating three mutants.

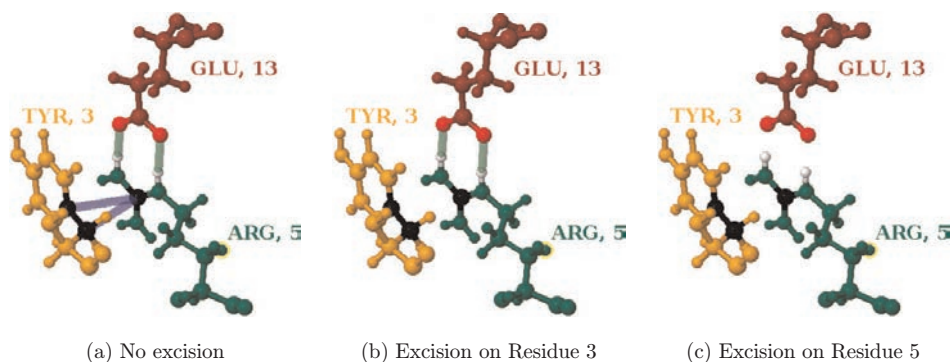


Fig. 4. KINARI-Mutagen simulates mutating a residue to glycine by removing its side-chain hydrogen bonds and hydrophobic interactions from the molecular model. (a) In the wild-type of PDB file 2PM1, two hydrogen bonds (light green bars) and two hydrophobic interactions (blue bars) exist among residues 3, 5, and 13. (b) Excising residue 3 removes the hydrophobic interactions that it forms with residue 5. (c) Excising residue 5 removes the hydrogen bonds between residue 5 and 13 and the hydrophobic interactions between residue 5 and 3.

We demonstrate the excision process on a fragment of human α -defensin 1 (Fig. 4). When excision is performed on residue 3, the hydrophobic interactions between it and residue 5 are removed from the molecular framework [Fig. 4(b)]. When excision is performed on residue 5 [Fig. 4(c)], then the hydrogen bonds and hydrophobic interactions that it engages in are removed.

In the third phase, the KINARI software is invoked to perform rigidity analysis on each mutant. Detailed descriptions of the rigidity calculation and modeling options are described in Ref. 2. When rigidity analysis is complete, an integrated Jmol-based visualizer is used to inspect the rigid regions of each mutant.

In the final stage of KINARI-Mutagen, the rigidity results for each of the mutants are aggregated. Information about critical residues can be inferred from several of the generated plots. Although this version of KINARI-Mutagen does not automatically predict which residues are critical, the Largest Rigid Cluster vs. Excised Residue plot (Fig. 6) designates a critical residue threshold, which is the average size of the largest rigid body for all of the analyzed mutants. Residues whose *in silico* mutation to a glycine causes the largest rigid body to decrease in size to below this threshold, are easily identified.

4. Case Study — Crambin

To demonstrate KINARI-Mutagen, for the first case study we generated and analyzed mutants of Crambin [PDB file 1crn, Fig. 5(a)], a 46 amino acid plant seed protein, whose crystals diffract to ultra-high resolution.²¹

The cartoon representation and rigidity results for two generated mutants of Crambin are shown in Fig. 5. The wild-type protein has a large rigid cluster [brown,

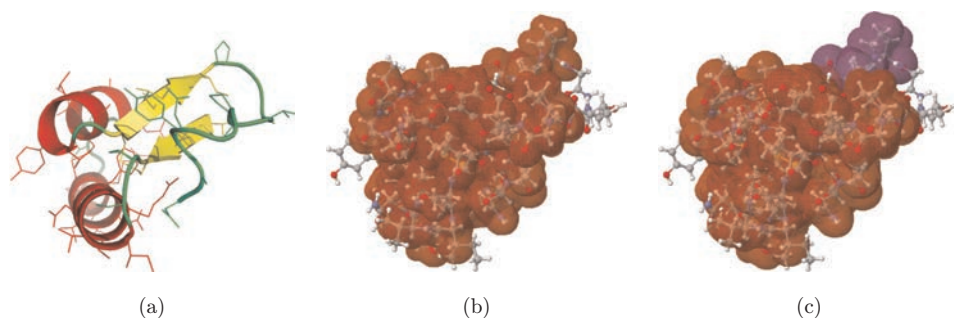


Fig. 5. KINARI-Mutagen was used to analyze Crambin (a) The largest rigid cluster of the wild-type protein (b) can be compared to the rigidity results of a mutant (c) to determine the effect of the mutation.

Fig. 5(b)]. Viewing the rigidity results of a mutant can be used to infer the impact of the mutation on the protein's rigidity. When excision was performed on residue 10, an arginine, [Fig. 5(c)], the size of the largest cluster decreased, and the number of clusters increased, when compared to the wild-type.

We wanted to know if KINARI-Mutagen could identify critical residues. KINARI-Mutagen uses the *SurfRace* program²² to calculate the Solvent Accessible Surface Area (SASA)²³ of each residue. A residue that is not exposed to the solvent has a low SASA value, measured in \AA^2 . Residues closer to the surface of a protein have higher SASA values, and completely buried residues have a SASA value of 0.

Several residues in the core of Crambin had a pronounced effect on the protein's predicted rigidity when they were mutated (residue 3 for example). Similarly, many residues (7, 15, and 28) that are solvent accessible, when mutated, had little effect on the largest rigid cluster. These findings were not surprising, because residues on the surface of a protein are not expected to help maintain a protein's stability.⁴ However, the software was able to identify critical residues on the surface of the protein that affected the protein's rigidity when mutated to a glycine.

We inspected the *Largest Rigid Cluster and SASA vs. Excised Residue* plot (Fig. 6), to identify critical residues that could not be located by using the SASA calculations alone. Of the 11 mutants that had largest rigid clusters below the critical residue threshold, eight of them (residue 2, 10, 17, 35, 36, 40, 41, and 44) had SASA values in the wild-type protein that were well above zero. Of these eight, 4 are known to be identical among viscotoxin A3 and α_1 -purothionin,²¹ while another 3 of them were conserved among two of these three homologous proteins. Only residue 44, with a SASA value of 70, was identified as critical, but which is not conserved among the three homologues.

One may hypothesize that KINARI's results may have a simpler explanation. A residue engaged in many stabilizing interactions (hydrogen bonds and hydrophobics) is likely to have an effect on the protein's stability and rigidity. To investigate this, we inspected the strengths of the hydrogen bonds of Crambin, which are calculated

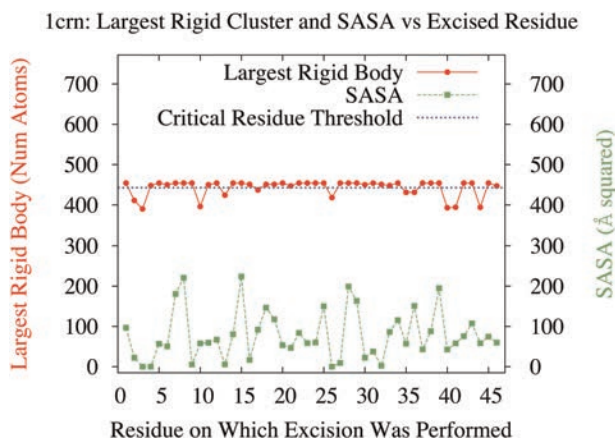


Fig. 6. The *SASA and Size of Largest Rigid Cluster vs Excised Residue* plot is used to locate solvent exposed amino acids that play a crucial role in stabilizing the protein. The blue dotted line, the critical residue threshold, designates the size of the average largest rigid body among all of the generated mutants, and can be used to select residues whose mutation affects the protein's rigidity.

by KINARI-Mutagen using an energy function.²⁴ Residues 46, 21 and 30, have side chains that engage in strong hydrogen bonds with energies of -5.2 , -5.9 , and -5.07 kcal/mol. KINARI-Mutagen did not identify them as critical, and they are not conserved among homologues of Crambin. We similarly confirmed that critical residues could not have been found by merely identifying amino acids that engage in many hydrophobic interactions. Residue 19, a proline, engages in five hydrophobic interactions. It is neither conserved among Crambin homologues, nor did KINARI-Mutagen identify it as critical.

KINARI-Mutagen is thus a method that supplements other approaches that study protein stability due to mutations and residue conservation. The set of critical residues identified by our method is different than the set of amino acids that are ranked by just the strength of hydrogen bonds or number of stabilizing hydrophobic interactions. Moreover, KINARI-Mutagen can identify conserved surface exposed residues that could not be detected using Solvent Accessible Surface Area measurements alone.

5. Case Study — Lysozyme from Bacteriophage T4

In the second case study, we evaluate whether rigidity analysis can identify destabilizing mutations. From the literature^{4,25–27} we retrieved stability data for 163 different point mutations in lysozyme from bacteriophage T4 (PDB ID 2lzm for Wild-type, Fig. 1). The experimentally derived value $\Delta\Delta G$, the free energy of unfolding, measures 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. From the available $\Delta\Delta G$ dataset, we selected the eight mutations that involved a substitution to a glycine. We compared $\Delta\Delta G$ values of these mutations that had been performed in the physical protein to the rigidity calculation predictions of KINARI-Mutagen.

Table 1 lists for each lysozyme mutant several rigidity measures, that we evaluated as predictors of protein stability. For the amino acid which was mutated at a particular sequence location, we list its Solvent Accessible Surface Area, the volume of the wild-type amino acid, the change of the volume of the residue when mutated to glycine, as well as the stability data from the literature ($\Delta\Delta G$), which we consider the “ground truth” stability measurement. The loss in the number of hydrogen bonds and hydrophobic interactions that were caused by the mutation to glycine are listed, as well as the change of the largest rigid body relative to the wild-type.

When KINARI-Mutagen was used to mutate residues 96, 105, 157 and 124, the size of the largest rigid body decreased in size in parallel to a decrease in the $\Delta\Delta G$ value. For example, residue 96, an arginine, when mutated to a glycine, caused the largest rigid body of the mutant to decrease by 130 atoms relative to the largest rigid body in the wild-type protein. When residue 96 was mutated to a glycine in the physical lysozyme, the stability of the protein decreased significantly, as indicated by the low $\Delta\Delta G$ value. Similarly, for residues 105, 157 and 124, the size of the largest rigid body decreased in size relative to the wild-type protein. The $\Delta\Delta G$ values for mutations at residues 105, 157 and 124, indicate that their destabilizing effect is not as great as when a mutation is performed on residue 96. In these cases KINARI-Mutagen was able to predict a change in the protein’s stability.

KINARI-Mutagen was not able in all instances to predict the effect of a mutation on the protein’s stability. For residues 99, 3, 59 and 55, the lack of loss of hydrogen bonds and/or hydrophobic interactions when these residues were mutated to a glycine explains why KINARI-Mutagen’s could not be used as a discerning measure of protein stability. For these mutations, the loss of hydrogen bonds and hydrophobic interactions were not as great as when mutations were performed on residues 96, 105, 157 and 124. The predictive ability of KINARI-Mutagen relies on the change in the molecular model due to a loss of these interactions. For these mutation instances, the change in the protein’s stability is caused by phenomena that KINARI-Mutagen does not currently capture. We suspect that the change in volume of the wild-type amino acid to a glycine causes a large-enough collapse or reorientation of the protein’s structure in the vicinity of the substitution, which affects the protein’s stability.

Lastly, we compared KINARI-Mutagen’s predictions to the Cluster Configuration Entropy (CCE) measurement.²⁸ 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

Table 1. Rigidity data of 8 lysozyme mutants.

Sequence number	WT amino acid, volume (\AA^3)	SASA (\AA^2) of WT amino acid	AA volume change for mutation to GLY	$\Delta\Delta G$ from literature	Hydrogen bonds lost in mutant	Hydrophobics lost in mutant	Change to largest rigid body in mutant	CCE ²⁸
99	LEU, 166	0	-106	-6.3	0	0	0	0.43
96	ARG, 173	73.04	-113	-2.6	2	0	130	0.61
3	ILE, 166	43.72	-106	-2.1	0	1	0	0.43
59	THR, 116	128.75	-56	-1.6	1	0	0	0.39
105	GLN, 143	64.8	-83	-1.5	2	0	11	0.44
157	THR, 116	100.97	-56	-1.1	2	1	63	0.49
55	ASN, 114	113.45	-54	-0.6	0	0	0	0.43
124	LYS, 168	103.19	-108	-0.1	1	2	15	0.44

Note: For each mutant, the experimental $\Delta\Delta G$ value, the change in volume of the amino acid when mutated to glycine, the loss of hydrogen bonds and hydrophobic interactions, the change of the largest rigid body, and the Cluster Configuration Entropy (CCE) values are listed. For the wild-type of the protein, the CCE value is 0.43, and the largest rigid body contains 830 atoms. Mutants are ordered by $\Delta\Delta G$ values, and rows shaded gray indicate amino acid mutations that KINARI-Mutagen correctly identified as destabilizing.

given as the following:

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

$$\text{CCE} = - \sum_s w_s \ln w_s$$

For two conformations of a protein, the one with the higher CCE value is more disordered, and hence is more unstable. The CCE values for several variants correlated well with the largest rigid body metrics for those mutants. For mutants that had residue substitutions at amino acids 105, 124, 157 and 96, the CCE values were 0.43, 0.44, 0.49 and 0.61, respectively, while the change in the size of the largest rigid body for those variants were 11, 15, 63 and 130.

To investigate why we did not predict the T59G mutant to be destabilizing, we referred to the *Distribution of Rigid Bodies, By Residue* (DRBR) plot (Fig. 7). It was used to distinguish between mutations that have only a local effect on the rigidity of a protein and mutations that drastically affect a protein's stability. The row **2lzm.A.0059** indicates that a mutant was generated by excising residue **59** of chain **A** of protein **2lzm**. For the residue 59 mutation, the change of the protein's rigidity is not localized to the largest rigid cluster, which explains why using the size of the largest rigid cluster was not a good predictor of protein stability.

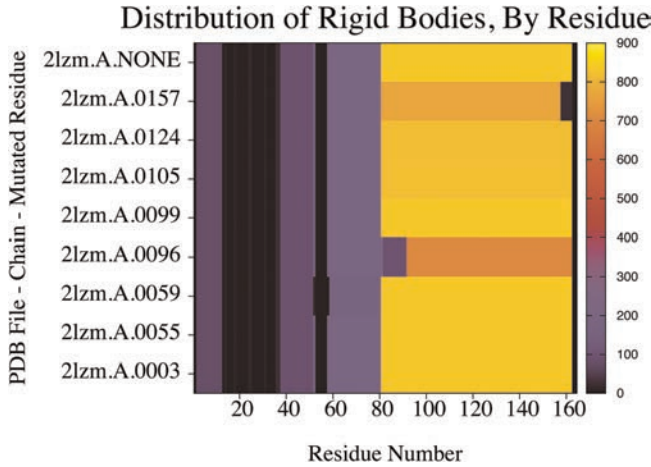


Fig. 7. *Distribution of Rigid Bodies, By Residue*: The left axis lists mutants that were analyzed. The vertical color legend on the right-hand side assigns colors to the rigid body sizes found among the mutants. The color at each x - y position in the plot indicates the size of the largest cluster that residue x belongs to for the mutant in row y .

6. Validation — 48 Mutants

To further determine if KINARI-Mutagen could correctly identify destabilizing mutations in a wider range of proteins, we searched the ProTherm Database,²⁹ which catalogues $\Delta\Delta G$ values for substitutions that have been performed in the physical protein. A total of 167 entries had mutations to glycine. Of those, 48 mutants among 14 proteins had single-point substitutions. We also chose PDB files that had all core residues resolved.

We used KINARI-Mutagen to generate the 48 *in silico* mutants and analyze their rigidity. Along with the SASA value for each wild-type residue at the location where the mutation was performed, we tallied the change to the largest rigid body of the protein caused by the point mutation, and the degree of hydrophobicity of each wild-type residue, using the Kyte and Doolite hydrophobicity scale.³⁰ The output of KINARI was also used to tally how many hydrogen bonds and hydrophobic interactions were lost due to the mutation. To facilitate analysis, the 48 mutants were grouped according to whether the substituted residue engaged in hydrogen bonds and hydrophobic interactions (Table 2). Detailed rigidity results for the 48 mutants are shown in Tables S1–S4, in the Supplementary file which is available from the journal website.

KINARI-Mutagen relies on the loss of hydrogen bonds and hydrophobic interactions upon a residue’s change to glycine, to predict the effects of a mutation. Thus we did not expect to accurately predict a substitution as destabilizing, if KINARI found that in the wild-type protein the amino acid engaged in neither hydrogen bonds nor hydrophobic interactions (Group 1). Group 2 has entries for which the residue of the wild-type protein was solvent exposed (more than 50% of the residue was exposed). Because these residues are on the periphery of the protein, their being mutated to glycine would not be expected to have a large effect on the size of the largest rigid cluster, especially if the side chain of the residue was protruding fully into the solvent [Fig. 8(a)].

In Group 3, four mutants had wild-type amino acids (Valine, Leucine, Methionine, Phenylalanine) that do form hydrophobic interactions that can be observed by visual inspection. However, because of the packing of these core residues in this structure which were slightly less tight than in many protein cores, the algorithm in KINARI to detect hydrophobic interactions detected far too few of them. Figure 8(b)

Table 2. The rigidity results for the 48 mutants was grouped based on whether the mutated residue engaged in stabilizing interactions.

Group	Description of wild Type AA at mutation point	# mutants	Identified as destabilizing
1	No hydrogen bonds or hydrophobics detected	13	0
2	Solvent exposed (>50%)	8	0
3	Too few hydrophobic	4	0
4	Stabilizing interactions found	23	22

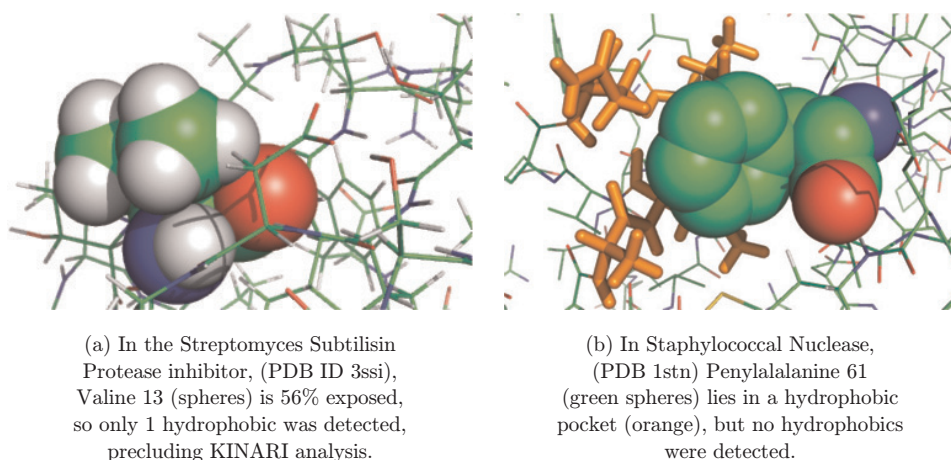


Fig. 8. Some amino acids that are highly solvent exposed were not identified as destabilizing, because they did not engage in stabilizing interactions [Fig. 8(a)]. Some residues like those that are completely or largely solvent inaccessible lie more than 3.5 Å from the nearest heavy atom, so hydrophobic interactions [orange sticks, Fig. 8(b)] in the range of 3.6 Å to 4.5 Å are not found by the hydrophobic detection algorithm in KINARI, preventing quantitative analysis of the impact of the mutation to glycine.

shows a phenylalanine that upon visual inspected should have been stabilized via several hydrophobic interactions, but no atoms in the residue were within 3.5 Å of a heavy neighbor atom, so no hydrophobic interacts were detected. Had the atoms of that part of the structure been oriented slightly differently to allow closer packing, KINARI's hydrophobic detection algorithm would have placed more hydrophobic interactions there, which could have caused that residue to be labeled as critical when it was mutated.

Group 4 contains 23 mutants that had more reasonable numbers of hydrophobic interactions which were identified by KINARI, and many of them have hydrogen bonds. Of the these, 22 were identified as critical, based on the fact that these mutants had largest rigid clusters that were smaller than the largest rigid cluster of the wild-type protein.

From the analysis of these 48 mutants, this first implementation of KINARI-Mutagen is able to make qualitative stability predictions. In the cases when residues are highly solvent exposed, KINARI-Mutagen is not as accurate, because such residues do not engage in as many stabilizing interactions as would be expected of them. Similarly, the pre-existing algorithm to detect hydrophobic interactions is not always accurate, when compared to the predicted hydrophobic interactions from a visual inspection. In future work, we plan to address this hydrophobic interaction algorithm.

7. Conclusions and Future Work

We have presented KINARI-Mutagen, which simulates mutating a residue to a glycine, and computes the mutant's rigidity. This first release of our software

performs the simplest kind of mutation. For future releases, we plan to include analysis of mutations that do not lead to major backbone conformational changes. The full range of mutations appears at this time to be a challenging task.

The results of this paper are already an indication that rigidity analysis provides valuable information about the stability of a mutated protein, that could not have been inferred by other methods, such as SASA measurements, or by ranking of contributions to stabilizing interactions.

Acknowledgments

These studies were funded by the National Institute of General Medical Sciences DMS-0714934 as part of the Joint program in mathematical biology supported by the Directorate for Mathematical and Physical Sciences of the National Science Foundation and the National Institute of General Medical Sciences of the National Institutes of Health.

References

1. Hutschison CA, Philipps S, Edgell MH, Gillham S, Jagnke P, Smith M, Mutagenesis at a specific position in a DNA sequence, *J Biol Chem* **18**:6551–6560, 1978.
2. Fox N, Jagodzinski F, Li Y, Streinu I, KINARI-Web: A server for protein rigidity analysis, *Nucleic Acids Res* **39** (Web Server Issue):W177–W183, 2011.
3. Garman SC, Garboczi DN, Structural basis of Fabry disease, *Mol Genetics Metab* **77**(1–2):3–11, 2002.
4. Alber T, Dao-pin S, Wozniak JA, Cook BW Matthews, Contributions of hydrogen bonds of Thr 157 to the thermodynamic stability of phage T4 lysozyme, *Nature* **330**:41–46, 1987.
5. Granzin J, Puras-Lutzke R, Landt O, Grundert H-P, Heinemann U, Saenger W, Hahn U, RNase T1 mutant Glu46Gln binds the inhibitors 2'GMP and 2'AMP at the 3' subsite, *J Mol Biol* **225**(2):533–542, 1992.
6. Xu J, Baase WA, Baldwin E, Matthews BW, The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect, *Protein Sci* **7**(1):158–177, 1998.
7. Alber T, Dao-pin S, Nye JA, Muchmore DC, Matthews BW, Temperature-sensitive mutations of bacteriophage T4 lysozyme occur at sites with low mobility and low solvent accessibility in the folded protein, *Biochemistry* **26**(13):3754–3758, 1987.
8. Eriksson AE, Baase WA, Zhang XJ, Heinz DW, Baldwin EP, Matthews BW, Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect, *Science* **255**:178–183, 1992.
9. Lee C, Levitt M, Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core, *Nature* **352**:448–451, 1991.
10. Gilis D, Rooman M, Predicting protein stability changes upon mutation using database-derived potentials: Solvent accessibility determines the importance of local versus non-local interactions along the sequence, *J Mol Biol* **272**(2):276–290, 1997.
11. Prevost M, Wodak SJ, Tidor B, Karplus M, Contribution of the hydrophobic effect to protein stability: analysis based on simulations of the Ile-96-Ala mutation in barnase, *Proc Nat Acad Sci USA* **88**(23):10880–10884, 1991.

12. Cheng J, Randall A, Baldi P, Prediction of protein stability changes for single-site mutations using support vector machines, *Proteins: Struct Funct Bioinform* **62**:1125–1132, 2006.
13. Topham CM, Srinivasan N, Blundell T, Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitutions and propensity tables, *Protein Eng* **10**(1):7–21, 1997.
14. Worth CL, Preissner R, Blundell L, SDM-a server for predicting effects of mutations on protein stability and malfunction, *Nucleic Acids Res* **39** (Web Server Issue): W215–W222, 2011.
15. Jacobs DJ, Thorpe MF, Generic rigidity percolation: The pebble game, *Phys Rev Lett* **75**:4051–4054, 1995.
16. Thorpe MF, Lei M, Rader AJ, Jacobs DJ, Kuhn LA, Protein flexibility and dynamics using constraint theory, *J Mol Graphics Model* **19**(1):60–9, 2001.
17. Rader AJ, Hespeneide BM, Kuhn LA, Thorpe MF, Protein unfolding: Rigidity lost, *Proc Natl Acad Sci USA*, p. 3540–3545, 2002.
18. Maxwell JC, On the calculation of the equilibrium and stiffness of frames, *Philosophical Magazine Series 4* **27**:294–299, 1864.
19. Tay T-S, Rigidity of multigraphs I: linking rigid bodies in n-space, *J Combinatorial Theory B* **36**:95–112, 1984.
20. Lee A, Streinu I, Pebble game algorithms and sparse graphs, *Discrete Math* **308**(8):1425–1437, 2008.
21. Teeter MM, Mazer JA, L'Italien JJ, Primary structure of the hydrophobic plant protein crambin, *Biochem* **20**(19):5437–5443, 1981.
22. Tsodikov OV, Record MT, Sergeev YV, Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature, *J Comput Chem* **23**(6):600–609, 2002.
23. Lee B, Richards FM, The interpretation of protein structures: Estimation of static accessibility, *J Mol Biol* **55**(3):379–400, 1971.
24. Mayo SL, Dahiyat BI, Gordon DB, Automated design of the surface positions of protein helices, *Protein Sci* **6**(6):1333–1337, 1997.
25. Bell JA, Becktel WJ, Sauer U, Baase WA, Matthews BW, Dissection of helix capping in T4 lysozyme by structural and thermodynamic analysis of six amino acid substitutions at Thr 59, *Biochem* **31**(14):3590–3596, 1992.
26. Mooers B, Baase WA, Wray JW, Matthews BW, Contributions of all 20 amino acids at site 96 to the stability and structure of T4 lysozyme, *Protein Sci* **18**(5):871–880, 2009.
27. Nicholson H, Soderlind E, Tronrud D.E., Matthews BW, Contributions of left-handed helical residues to the structure and stability of bacteriophage T4 lysozyme, *J Mol Biol* **210**(1):181–193, 1989.
28. Gohlke H, Radestock S, Exploiting the link between protein rigidity and thermostability for data-driven protein engineering, *Eng Life Sci* **8**:507–522, 2008.
29. Kumar MD, Bava KA, Gromiha MM, Prabakaran P, Kitajima K, Uedaira H, Sarai A, Protherm and pronit: Thermodynamic databases for proteins and protein-nucleic acid interactions, *Nucleic Acids Res* **34** (suppl 1):D204–D206, 2005.
30. Kyte J, Doolittle RF, A simple method for displaying the hydropathic character of a protein, *J Mol Biol* **157**(1):105–132, 1982.



Filip Jagodzinski is a Ph.D. candidate in the Department of Computer Science at the University of Massachusetts, Amherst. He received his B.S. degree in Biomedical Engineering from Columbia University, and his M.S. in Computer Science from Villanova University, in 1999 and 2003, respectively. Previously he demonstrated using insights from robotics operational space control to simulate the large-scale motions of proteins.



Jeanne Hardy received her Ph.D. degree in Biochemistry and Molecular Biology from the University of California at Berkeley, in 2000. From 2002–2005, she was an NIH Postdoctoral Fellow at Sunesis Pharmaceuticals. She is currently an Assistant Professor at the Department of Chemistry at the University of Massachusetts Amherst.



Ileana Streinu received a Doctorate in Mathematics/Computer Science from the University of Bucharest, Romania, and a Ph.D. degree in Computer Science from Rutgers University, both in 1994. She is the Charles N. Clark Professor of Computer Science and Mathematics and director of the Biomathematics program at Smith College. Prof. Streinu is the recipient of the 2010 Robbins Prize of the American Mathematical Society and the 2004 Moisil Prize of the Romanian Academy.

Copyright of Journal of Bioinformatics & Computational Biology is the property of World Scientific Publishing Company and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.