AMMCS-2011, Laurier Centennial:
International Conference on
Applied Mathematics, Modeling &
Computational Science,
Waterloo, Canada, July 25-29

Predicting protein hinge motions and allostery using rigidity theory

Adnan Sljoka[1] and Alexandr Bezginov[2]

[1] *York University, Toronto, Canada, adnanslj@yorku.ca*
[2] *University of Toronto, Toronto, Canada, alex.bezginov@utoronto.ca*

Abstract

Understanding how a 3-D structure of a protein functions depends on predicting which regions are rigid, and which are flexible. One recent approach models molecules as a structure of fixed units (atoms with their bond angles as rigid units, bonds as hinges) plus biochemical constraints coming from the local geometry [6, 13]. This generates a 'molecular graph' in the theory of combinatorial rigidity. The $6|V|-6$ counting condition for 3-dimensional body-hinge structures (modulo recently solved 30 year old molecular conjecture), and a fast 'pebble game' algorithm which tracks count in the multigraph, have led to the development of the program FIRST [3], for rapid predictions of the rigidity and flexibility of proteins.

In our prior work we have extended the pebble game algorithm to specify regions that are *relevant* as constraints with respect to a specified region (core) [10]. In this study we further extend this algorithm and offer a novel protein hinge prediction algorithm. We have tested our hinge prediction algorithm on several proteins, chosen from the dataset of manually annotated hinges available on the MOLMOV server [1]. Many of our predictions are in very good agreement with this data set. Our algorithms can also predict 'allosteric' interactions in proteins - where binding on one site of a molecule changes the shape or binding at a distance 'active site' of the molecule. We give some recent promising results which support the sliding piston-like movement of helices with respect to one another as a plausible mechanism by which receptors propagate conformational changes across membranes.

Introduction and Background

Rigidity theory has a rich history, but it is only in the last thirty years that it has started to find the applications in basic sciences [3, 6, 10, 9, 11]. A natural question was asked by James Clark Maxwell: Can we count vertices and edges in a framework in order to make predictions about its rigidity and flexibility? The first complete counting (combinatorial) generalization for (generic) bar and joint frameworks in 2-dimension was confirmed by Laman in 1970 [12]. Using Laman's theorem it is enough to count the vertices and edges and their distributions in the framework to give full answers about rigidity and avoid altogether the difficult geometry (i.e. rigidity becomes only a property of the underlying graph). One useful family of structures (known to structural and mechanical engineers and in robotics community) which also have nice counting results for rigidity and a well develop theory (via rigidity matrices) are the body-bar, body-hinge structures [12]. An important application of these structures is that molecules can be modelled as a special case of body-hinge structures, where atoms become rigid bodies connected by molecular hinges (bonds) in 3-space [13]. The recently solved molecular theorem [4] offers a remarkable graph theoretical criteria for determining the rigidity of molecular structures. In this count an atom has 6 DOF and each rotatable bond removes 5 DOF.

Theorem 1 (Molecular Theorem) *A generic molecular structure (generic positions of atoms) on a graph $G_M = (V, H_M)$ is rigid if and only if, each molecular hinge H_M is replaced by 5 edges, in the corresponding multigraph G^*_M, $|E| = 6|V| - 6$, and on all subsets of edges E', $|E'| \leq 6|V'| - 6$.*

In the stated form this count gives a poor algorithm, as it requires counting the number of edges on all subgraphs, which is an exponential number. However, we can use an efficient ($O(|V||E|)$) greedy pebble game algorithm to track this count on a multigraph and its subgraphs (for details see [7, 10]). The implementation of the pebble game on molecules has led to the development of FIRST. This program starts with a snapshot of a protein (PDB), and creates a multigraph where vertices are atoms and edges represent the distance constraints corresponding to the molecular interactions (covalent and hydrogen bonds and hydrophobic interactions). FIRST

then applies the pebble game to give fast predictions of all the rigid regions (regions that only have trivial rigid-body motions - translations and rotations) in the protein and flexible connections (corresponding to rotatable bonds). FIRST is free and available as an online server [3, 6].

In our prior work we have developed a *Relevant Regions Detection Algorithm* [10]. In this algorithm we select a nonempty (induced) subgraph G_c from the multigraph G^*_M, which we call the *core*. We play the pebble game on G^*_M and do special pebble reversing operations to identify the *relative* degrees of freedom (DOF) and the *relevant region* of the core. *Relative* DOF quantify the flexibility of the core. In this selective course graining, *relevant regions* constrain the possible motions of the core. In this paper we further modify the *Relevant Regions Detection Algorithm* and use it develop a hinge prediction algorithm with applications to allostery and signal transduction in proteins. These new algorithms as an add-on to FIRST with documentation and source codes are available at a request.

Hinge Predictions

Hinge motions in proteins are essential for large repertoire of functions, including binding to other proteins and ligands, catalysis and drug docking. Hinges occur between two well-defined (rigid) domains connected by a flexible linker that move relative to each other. The easiest hinges to predict occur when the structure of a protein exists in two or more different conformations (i.e. open and closed). One can then visually inspect the pairs of structures and manually annotate the hinge location. These proteins often serve as test cases for hinge prediction programs. One collection of well annotated hinges is Hinge Atlas Gold (HAG) data set available at MOLMOV [1].

In the *Relevant Region Algorithm* if we let the *core* be composed of two large rigid clusters (extracted from the output of FIRST), we can locate the hinge points by finding the relevant region. Furthermore, relative (DOF) of the core is DOF available in the hinge motion. We can think of this as freezing a rigid cluster in the hinge, and see how many DOF are available to the other cluster. The answer could be zero (they are in the same rigid region), 1, 2, 3, 4, 5 or 6 DOF, where six could be anything from a linker with six rotatable bonds, up to complete disconnection.

Algorithm 2 – ***Hinge Prediction algorithm:***

Input*: PDB file.*

Output*: Hinge location and DOF of hinge*

File preparation: For a given protein (PDB) add missing protons (usually to all x-ray structure) (i.e. using WHATIF or REDUCE programs). If needed, remove water molecules, ligands and other heteroatoms.

1. *Run FIRST algorithm and choose an energy cutoff corresponding to the two largest rigid clusters and set them as* core.
2. *Find* relevant region *of the* core *and relative DOF of the core (pebbles reversed to the core).*

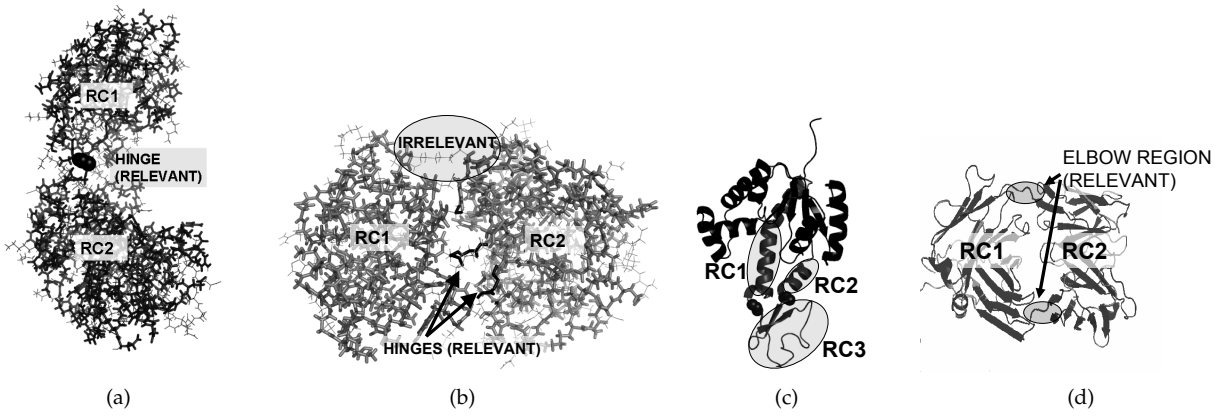
The relevant region (atoms and bonds) is the predicted hinge location.

In Table 1 and Figure 1 we have the output of our Hinge Prediction algorithm on several proteins chosen from HAG. The predicted hinge residues are in good agreement with HAG. We have also output the DOF count in the hinge. For Calmodulin, an important calcium binding protein, and Inorganic pyrophosphatase we get an exact match with HAG. In these cases there is a single point in the backbone chain separating the rigid clusters. In LAO Binding protein, there are two independent hinge regions between the clusters. Adenylate Kinase represents an interesting case as the second and third largest clusters are similar in size and proximity to the largest rigid clusters. For this protein we ran the algorithm two times, once for each pair of rigid clusters.

Due to the nature of the pebble game algorithm, this algorithm gives rapid predictions of hinges requiring only a single structure (pdb), which makes it suitable for high-throughput analysis. Other hinge predictors provide only hinge location to the residue accuracy, whereas our method can also detect any side chain interactions that might be constraining the motions of the two domains. This may be very useful information for mutation studies (i.e. Which mutations cause the change in the hinge motion and mutations are not important?). Finding relevant regions for hinges allows us to find only those hinges that constrain the motion between the rigid clusters, and ignores the longer *irrelevant* linkers that are not playing the role in the hinge motions. This avoids overpredicting hinge locations as is the case with program StonehingeP [5]. For instance, in LAO Binding protein one of the linkers joining the two clusters is found to be irrelevant, thus it is not a hinge point. Our Hinge Prediction algorithm is also capable of detecting other types of motions (i.e. shear) or with unknown mechanisms. The elbow in Immunoglobulin FAB region is one such example where we locate a relevant region between the two clusters, while it is not provided in HAG. The relative DOF between the two clusters not only identifies how 'tight' the hinge is, but this count (1 to 6) allows us to do further geometric classification and asses the likelihood that a motion is a pure rotation, translation (shear), or a general screw motion (a mix of translation and rotation).

Table 1: Outputs of hinge prediction algorithm

Protein	PDB ID	HAG residues	Predicted hinge residues	DOF
Adenylate Kinase	2AK3	125-126, 161-162	124-125, 158-159	3, 3
Calmodulin	3CLN	80-81	80-81	1
Inorganic pyrophosphatase	1K23	189-190	189-190	3
LAO Binding Protein	2LAO	90-91, 192-193	89-91, 192-194	1
Immunoglobulin (FAB ARM)	1IGT	N/A	C107-108, D112-113	5



(a) (b) (c) (d)

Figure 1: Output of Hinge Prediction Algorithm. Inorganic Pyrophosphatase hinge of 3 DOF corresponds to the relevant region (spheres) between Rigid Cluster 1 (RC1) and RC2 (a). LAO Binding protein has two hinges (b). Top linker is irrelevant and is not a hinge point. Adenylate Kinase has two hinges among three rigid clusters (c). Other types of motions such as elbow motion in the FAB arm of immunoglobulin can be detected (d).

Allostery

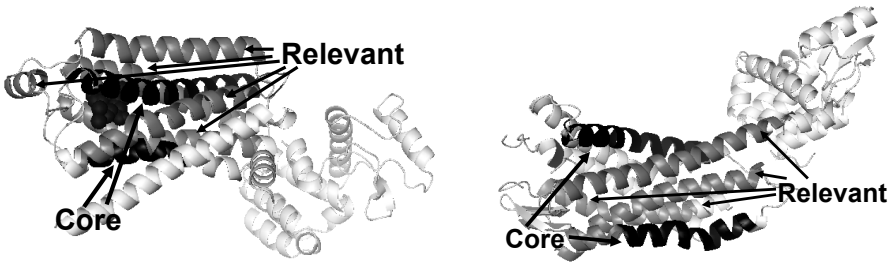
While most of the work on predicting allostery is an area of ongoing research, we give some promising and encouraging results to this important biological mechanism. In many proteins the binding of a ligand at the allosteric site triggers a conformational change that can propagate a substantial distance to cause a rearrangement and change in shape at the active site [8]. The exact mechanism of this allosteric structural change propagation is elusive and not fully understood. Working with the molecular model, we can ask: If there is a change in rigidity (shape) at one site when will there be a corresponding change at another site, in another words how are the DOF and motions of distinct sites coordinated? If two sites are allosterically linked, which regions are important for this allosteric transmission (i.e. allosteric pathways).

The phenomenon of allostery is prominently seen in transmembrane receptors, where the mechanical effect of ligand binding is mediated across the entire width of the plasma membrane [2]. G protein-coupled receptor (GPCR) proteins are characterized by a bundle of 7 transmembrane (TM) helices connected by extra- and intracellular loops. It is hypothesized that a piston-type of mechanism may be used by many GPCR proteins, where conformational changes in receptor are transmitted by sliding (shear) motions of the helices relative to each other [2].

In order to identify the regions involved in allosteric motions, we have performed the Relevant Region analysis on a set of high-resolution crystal structures of GPCR proteins. In Figure 2 we have the output on Human beta-2 adrenergic receptor (PDB: 2RH1) and chemokine receptor type 4 (CXCR4) (PDB: 3OE0). Beta-2 adrenergic receptor is implicated in multiple medical conditions such as cardiac arrhythmia, hypertension and bronchial asthma, while CXCR4 is linked to the growth and spread of cancer, mobilization of hematopoetic stem cells and HIV.

In Human beta-2 adrenergic receptor 5 of 7 TM helices exist as separate rigid clusters. When two helices (residues 102-142 and 304-321) are selected as the core, the relevant region spans a large area and encompasses four other helices, inhibitory compound carazolol (spheres) and part of the extra- and intracellular regions. In CXCR4 we find all TM helices are separate rigid clusters, and when we choose helices (residues 146-168 and 238-268) as

the core we find that three other helices form a relevant region. These findings suggest that the relative motions of two transmembrane helices occurs in a coordinated fashion and transmission of this communication involves other helices, the ligand and extra- and intracellular regions, and supports the proposed piston-like movement of helices as a mechanism by which receptors propagate conformational (shape) changes across membranes. This can have important implications for drug design, since a small-molecule compound could now be engineered to inhibit a receptor via interfering with the signal transmission regions (relevant regions), rather than directly targeting either the ligand binding or the active site.



(a) (b)

Figure 2: Human beta-2 adrenergic receptor (2RH1) (a) and C-X-C chemokine receptor type 4 (3OE0) (b) indicates that relative motion of two helices involves other helices supporting the sliding (shear) piston-like mechanism. Black helices are two separate rigid clusters forming the core, darker gray helices are the relevant regions.

Acknowledgements

We thank Walter Whiteley for many useful discussions and assistance with this manuscript.

References

[1] Database of Macromolecular Movements, http://www.molmovdb.org

[2] Gerstein, M. and Chothia, C., Signal Transduction: Proteins in Motion, Science 1999 285 (5434): 1682-1683.

[3] Flexweb Server (FIRST), http://flexweb.asu.edu

[4] Katoh, N. and Tanigawa, S., A Proof of the Molecular Conjecture. Discrete & Computational Geometry 45 (2011), 647-700.

[5] Keating, K., Flores, S., Gerstein, M. and Kuhn, L. (2009). StoneHinge: Hinge Prediction by Network Analysis of Individual Protein Structures. Protein Science 18: 359-371.

[6] Kuhn, L., Rader, D. and Thorpe, M., 2001, Protein flexibility predictions using graph theory, Proteins, 44:150-65.

[7] Lee, A. and Streinu, I., Pebble game algorithms and (k, l)-sparse graphs. accepted to eurocomb 05, 2005. arXiv:math/0702129v1

[8] Monon, J., Wyman, J., and Changeux, J., On the nature of allosteric transitions: A plausible model, Journal of Molecular Biology, vol. 12, pp. 88-118, May. 1965.

[9] Shai, O., Sljoka, A. and Whiteley, W., Directed graphs, Decompositions and Spatial Rigidity, arXiv:1010.5552.

[10] Sljoka, A., Counting for Rigidity, Flexibility and extensions vis the Pebble Game Algorithm, Masters Thesis, York University, 2006.

[11] Sljoka, A., Shai, O. and Whiteley W., Checking Mobility and Decomposition of Linkages via Pebble Game Algorithm, ASME 2011 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, IDETC/CIE 2011.

[12] Whiteley, W., Some Matroids from Discrete Applied Geometry, Contemporary Mathematics, AMS **197** (1996), 171–311.

[13] Whiteley, W., Counting out to the flexibility of molecules, Phys. Biol. 2, S116–S126, 2005.