WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

HingeProt: Automated prediction of hinges in protein structures

Ugur Emekli,[1] Dina Schneidman-Duhovny,[2] Haim J. Wolfson,[2] Ruth Nussinov,[3,4] and Turkan Haliloglu[1]*

[1] Polymer Research Center and Chemical Engineering Department, Bogaziçi University, 34342 Bebek, Istanbul, Turkey

[2] School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

[3] Basic Research Program, SAIC-Frederick Inc., Center for Cancer Research Nanobiology Program, NCI-Frederick, Frederick, Maryland 21701

[4] Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

ABSTRACT

Proteins are highly flexible molecules. Prediction of molecular flexibility aids in the comprehension and prediction of protein function and in providing details of functional mechanisms. The ability to predict the locations, directions, and extent of molecular movements can assist in fitting atomic resolution structures to low-resolution EM density maps and in predicting the complex structures of interacting molecules (docking). There are several types of molecular movements. In this work, we focus on the prediction of hinge movements. Given a single protein structure, the method automatically divides it into the rigid parts and the hinge regions connecting them. The method employs the Elastic Network Model, which is very efficient and was validated against a large data set of proteins. The output can be used in applications such as flexible protein–protein and protein–ligand docking, flexible docking of protein structures into cryo-EM maps, and refinement of low-resolution EM structures. The web server of HingeProt provides convenient visualization of the results and is available with two mirror sites at http://www.prc.boun.edu.tr/appserv/prc/HingeProt3 and http://bioinfo3d.cs.tau.ac.il/HingeProt/.

Proteins 2008; 70:1219–1227.

Key words: hinges; protein flexibility; NMA; protein function; docking; EM.

INTRODUCTION

Since conformational changes are an inherent part of protein functionality, the study of molecular motions has numerous applications. Getting an insight into conformational flexibility is expected to assist in the comprehension of the mechanisms of protein functions in the respective biochemical processes. Further, it can assist in the prediction of 3D structures of protein interactions (docking). The task of protein docking is immensely complicated by protein flexibility.[1] Most docking methods fail even in the case of limited backbone flexibility. Predicted protein movements can be used to generate a set of protein conformations for docking studies.[2–5] In addition, flexible docking methods can benefit from an a-priori identification of rigid parts and hinges.[6–8] Moreover, prediction of protein movements can help in fitting of high-resolution protein structures into cryo-electron microscopy (cryo-EM) density maps.[9–13]

It is common to classify protein movements into hinge and shear.[14] Hinge movements involve rotation of protein parts (mostly domains) about a region called a hinge (in most of the cases a loop or linker). This region usually involves several residues that undergo significant conformational changes, but most of the rotating protein parts remain unchanged. Shear movements involve a sliding movement of proteins parts with respect to each other. This movement is usually restricted, with small conformational changes across the movement interface plane. Recent structural studies show that the movement picture is even more complicated

Grant sponsor: Turkish Academy of Sciences (Young Scientist Award Program); Grant number: EA-TUBA-GEBIP/2001-1-1; Grant sponsor: State Planning Organization of Turkey; Grant numbers: 03K120250, 03A501D; Grant sponsor: European Commission; Grant number: EU_FP6-2004-ACC-SSA-2 (Project No: 517991); Grant sponsor: NIAID, NIH; Grant number: 1UC1AIO67231; Grant sponsor: Israel Science Foundation; Grant number: 281/05; Grant sponsor: National Cancer Institute, National Institutes of Health; Grant number: NO1-CO-12400; Grant sponsors: US–Israel Binational Science Foundation, Hermann Minkowski-Minerva Center for Geometry at TAU; Center for Cancer Research, National Cancer Institute, NIH (Intramural Research Program).
*Correspondence to: Turkan Haliloglu, Polymer Research Center and Chemical Engineering Department, Bogaziçi University, 34342 Bebek, Istanbul, Turkey. E-mail: halilogt@boun.edu.tr
Received 31 December 2006; Accepted 2 May 2007
Published online 10 September 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21613

and some proteins may present a large repertoire of conformational changes.[15]

Several approaches are available for the identification of molecular flexibility. These can be divided into two major groups.[16] The first group identifies structural flexibility through analysis of two or more different conformations of the same or homologous protein.[17–19] Methods belonging to the second group start from a single protein structure. These methods are not limited by the experimental availability of different conformations. Molecular Dynamics (MD) and Monte-Carlo (MC) are classical representatives of methods in this group. However, these methods are computationally expensive and therefore not capable of capturing large-scale protein rearrangements. Other efficient approaches attempt to estimate protein rigidity and flexibility using graph[20] or constraint theory.[21] Methods based on the Normal Mode Analysis (NMA) fall in-between the accurate computationally expensive MD/MC methods and the less accurate, however, very efficient graph theoretic approaches. Most of the approaches available to date for the hinge prediction task belong to the first group of methods, which require two protein structures and can not be used when only a single structure is available. In this work, we present a method for the explicit prediction of hinges that belongs to the second methods category. Our method, HingeProt, is an NMA-based approach for the identification of hinge regions and rigid parts given a single protein structure.

NMA is a unique tool for providing an analytical solution for the modes of movement accessible to a given structure in its global minimum.[22,23] With the computationally simpler coarse grained elastic network (EN) models, NMA has recently shown its success for predicting the intrinsic collective movements.[24–28] It has been shown that known protein movements can be predicted by perturbing the original structures along the direction of the two slowest frequency normal modes.[25,29]

There are several web servers that employ NMA to predict protein movements. MoViEs[30] and NOMAD-Ref[31] apply NMA on the atomic level of the structure. WEBnm@[32] and ElNemo[33] work on the residue level. Moreover, ElNemo can handle large protein structures using the RTB model (the building block approximation that groups several residues into a single super-residue), instead of residue level representation. The oGNM/iGNM server[34] also works on the residue level and calculates the equilibrium dynamics of an input structure using the Gaussian Network Model.[24,35] The AD-ENM/DC-ENM server can perform normal modes-based prediction of protein conformational changes guided by distance constraints.[36] The ANM server performs systematic evaluation of the Anisotropic Network Model.[37] The MolMovDB server focuses on assessing the significance of the movement direction vectors of NMA using a large dataset of protein movements.[38] The ProMod database provides the NMA results for a large dataset of molecules using the full atom model.[39]

Despite the large number of available tools, interpreting NMA output that is usually in the form of eigenvalues and eigenvectors is not trivial.[40] Most of the servers provide visualization of the predicted movement and/or graphical representation of mean-square fluctuations and cross correlations. Here, we employ NMA for explicit identification of hinge regions in protein structures. The novelty of the HingeProt method is in using NMA for fully automated detection of hinge regions from a single protein structure. The method focuses on partitioning the input structure into rigid parts and the hinge regions connecting them based on the NMA output. HingeProt is available as a web server. The input to the server is the protein structure, which may consist of several protein chains. The output is conveniently given as a list of the rigid parts and the hinge regions as well as the visualization of the fluctuation for the two slowest modes. The method was successfully tested and validated on more than 1000 protein structures.

Below, we describe the method and the validation procedure. In the Results section, we present a few selected test cases and large scale validation on two datasets of known hinge movements.

METHODS

The HingeProt method consists of two main stages: (i) Analysis of the input structure using Elastic Network Models and (ii) Processing the results for prediction of the hinges and the rigid parts.

Analysis of the input structure using Elastic Network Models

Given an input PDB structure, we apply the Gaussian Network Model (GNM)[24,35] using only the Cα atoms of the structure. We compute the eigenvectors corresponding to the slowest first and second modes of the GNM. The square of each eigenvector describes the mean square fluctuations (the autocorrelations) of residues from equilibrium positions along the kth principal coordinate (k = 1, 2 for the first and second modes). A correlation between two residues reflects their tendency to move in the same direction. Thus, we compute the correlation matrix of the structure by computing the correlation between each pair of residues in a given mode (see Fig. 1 for the correlation matrix of the glutamine binding protein for the first slowest mode). Each axis in the figure corresponds to the residue numbering of the input protein. The lines point to residues where the change of sign occurs. "+" suggests positive correlation, implying that residues tend to move in the same direction, while "–" stands for negative correlation, i.e. residues that move in opposite directions. Since the GNM is a one dimensional

DOI 10.1002/prot

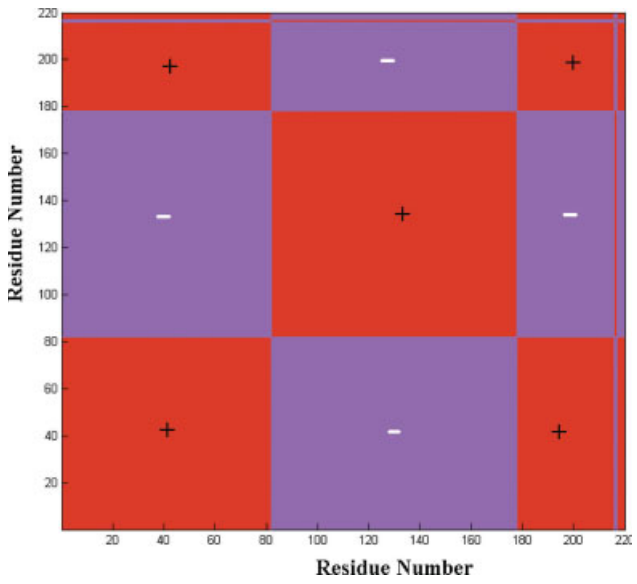

Figure 1

Correlation between the fluctuations of residues for the first slowest mode of glutamine binding protein (PDB 1gggA). Regions marked with + have correlation values of +1 and those marked with − have values of −1. Each rectangular region indicates cooperative motion between the residues inside the rectangle. Initially, HingeProt divides this structure into three rigid fragments: residues 5–86, 87–183, and 184–224 (note that the numbering in the figure starts from 1, while in the PDB the first residue is number 5). A change in the sign for short fragment of residues 217–218 is ignored. In the clustering stage, regions 1–86 and 184–224 are clustered into one rigid part.

model in its description of the fluctuations, the change in the sign of the correlation values between − and + suggests a flexible joint that connects the rigid structural units. We use the change of sign in autocorrelation as an initial guess for hinge positions.

In principle, we can use several modes for hinge prediction. Here, we focus on the two slowest modes that are the most cooperative global modes of movements and are related to functional movements in a structure.[29]

Processing the results for prediction of hinges and rigid parts

This stage is performed separately for each GNM mode. Each mode gives a different set of hinge regions. The hinges from the two modes may be completely different or may be associated with the same regions. The input to this stage is the correlation matrix for a given mode. First, we divide the structure into sequential rigid fragments based on the change of sign of the correlation. Next, we cluster the fragments into structurally compact rigid parts and score the obtained prediction. Below, we describe these stages in detail. The output of this stage is the final partition into rigid parts and hinge regions with the partition score for each GNM mode.

Partitioning of the structure into sequential rigid fragments

The protein chain is divided into sequential rigid fragments based on the change of the sign of the correlation. For example, the (i–j) fragment starts with residue i and ends with residue j. All residues in the same fragment initially have the same sign. The endpoints i and j are considered as hinge residues. Usually, a hinge in the protein involves several residues, not necessarily being close in sequence, since several torsion angles have to change for large domain movement. Here, we mark the hinge by one residue, since we need to break the chain at the specific point. Some of the derived fragments are very short and represent a moving loop or simply noise. Therefore, rigid fragments consisting of less than 15 residues are merged with their neighboring rigid fragments. These short fragments can be neglected in the search for large rigid parts. For example, the glutamine binding protein (Fig. 1, first slowest mode) will initially be partitioned into five sequential fragments: 1–86, 87–183, 184–216, 217–218, and 219–224. Because of the very short length of the fourth fragment 217–218, it is merged into one three sequential fragments.

Spatial fragment clustering

Next, we apply a spatial clustering of the rigid fragments. Two rigid fragments (i–j) and (i'–j') are joined if: (i) their correlation sign is the same; (ii) their endpoints are within 16 Å from one another, i.e. the distance between the Cα atoms of residue i and residue i' and the distance between the Cα atoms of residue j and residue j' are less than 16 Å. We also attempt to join the fragments in the opposite direction, by checking the distance between the Cα atoms of residue i and residue j' and the distance between the Cα atoms of residue j and residue i'. If one of the endpoints is the N- or C-termini, this endpoint is not checked. The clustering is performed using the CAST clustering method.[40] Each cluster produced by the CAST algorithm represents one rigid part and may include several initial rigid fragments. These rigid parts are the output of the HingeProt method. For example, in the glutamine binding protein [Fig. 2(c,d)], the first and the third sequential fragments are clustered by the clustering algorithm, dividing the structure into two rigid parts. After the clustering, each rigid part may have more than two endpoints. We define a *hinge region* as a set of spatially clustered endpoints of a rigid part. For example, in the glutamine binding protein, the hinge region includes two *hinge residues* 87 and 183, which are far in sequence but close in space, with a distance of 6 Å between their Cα atoms [Fig. 2(c,d)].

Scoring the quality of the partition

Once we partition the protein into hinges and rigid parts, we would like to score its quality. The motivation

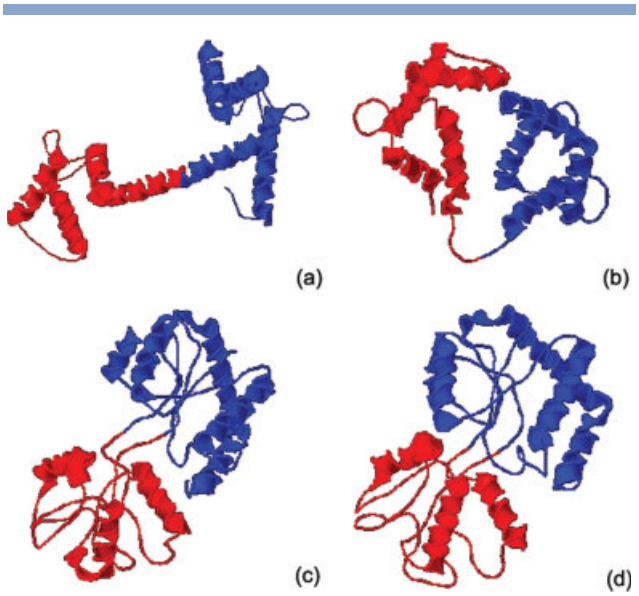

Figure 2

*Single chain examples: (**a**) Calmodulin protein (PDB 4cln). The slowest first mode divides the protein into two rigid parts: (1–80) and (81–148), with one hinge residue (80) between them; (b) The bound form of calmodulin protein (PDB 2bbm). The hinge residue (80) can still be observed; (c) Glutamine binding protein (PDB 1gggA). The protein is divided into two rigid parts: (5–87,184–224) and (88–183) by the first slowest mode. The two parts are intertwined and therefore connected by two hinge residues: 87 and 183; (d) The HingeProt output for the closed form of glutamine binding protein (PDB 1wdnA). Hinges can still be recognized by the first slowest mode.*

for this stage is to be able to recognize wrong partitions to rigid parts and cases where the input protein does not exhibit hinge movement. Hinge movements are similar to rotations around an articulated joint and therefore can be very large; however, only a small number of residues are involved since even one bond can provide the required rotational freedom.[14,41] Based on this observation, we developed a scoring function that favors partition to compact rigid parts with small interface between the parts, i.e. many intrapart contacting residues and as few as possible interpart contacts. Two residues are considered as contacting if the distance between their Cα atoms is below a threshold (here we used 13 Å). We define the score of the rigid part as the number of contacts within the given part divided by the total number of contacts the residues of this rigid part make within the whole protein structure. For each rigid part R_k in a structure of N residues:

$$\text{Score}(R_k) = \sum_{i,j \in R_k} S_{ij} \Big/ \sum_{i=1\ldots N, j \in R_k} S_{ij} \qquad (1)$$

S_{ij} is 1 if the distance between the Cα atoms of the two residues is below a threshold and 0 otherwise. The numerator favors the compactness of residues within the rigid part and the denominator decreases the score of parts with many intermolecular contacts. The score of the partition is defined as an average score of all its rigid parts:

$$\sum_{k=1}^{M} \text{Score}(R_k)/M \qquad (2)$$

M is the number of rigid parts. The score is in the range [0–1], while good partitions are usually closer to 1.0.

Movement visualization

The GNM predicts the relative magnitudes of the fluctuations but not the direction of the fluctuations, as it is isotropic by definition. The direction of the fluctuations can be characterized by ANM.[42] Thus, here we used GNM, which has been shown as more robust in the calculation of mean square fluctuations in the slowest modes,[42] to predict the hinge regions; whereas, the ANM is used to predict the direction of the fluctuations in the corresponding modes of GNM from which the hinge regions are determined.

ANM with $3N - 6$ eigenvectors describes the fluctuations of N residues in the x, y, and z directions from the average structure according to the kth mode. The ANM modes are mapped to the GNM modes by comparing the mean square fluctuations of residues between the resulting modes in the two models. This allows us to give a predicted orientation to the fluctuations of residues in the two most dominant GNM modes. As the fluctuations are symmetric with respect to the equilibrium positions, ANM-predicted deformed structures could readily be obtained by adding and subtracting the fluctuations of each residue to/from its equilibrium position. It should be noted that the absolute magnitudes of the fluctuation vectors can not be predicted by GNM or ANM. It depends on the force constant in the harmonic potential assumed between the interacting residues. Thus, the extents of deformations from the average structure are meaningful only when compared with respect to the residues/structural units of the same protein in a given mode.

Web server interface

We provide a web server interface for HingeProt. The server has a very simple user interface and output, which is easy to interpret. The input to the server is the coordinate file of a protein in PDB format. The user may enter a four-letter PDB code or upload a PDB file. The user can choose whether to carry out the analysis on a single chain or the whole structure. If the input file includes more than one chain, the user uses the "Select chain" option to select one or more chains on the next page. The user may optionally enter an e-mail address, to

DOI 10.1002/prot

which the results will be sent once the job has finished. In either case, the results, i.e. the rigid parts and the hinge residues, will be displayed in the browser window. The fluctuations of the structure at each mode are shown in the *Jmol* applet. The structure is colored according to its partition into rigid parts. The PDB files used for animation can be downloaded by the user.

RESULTS AND DISCUSSION

Selected case studies

Calmodulin

There is a very large conformational change in the calmodulin molecule upon ligand binding. This involves the splitting of one large helix. Figure 2(a) shows the output obtained by HingeProt for calmodulin (PDB 4cln). The first slowest mode divides the molecule into two rigid parts in the middle of the helix that was proposed to be broken as seen in Figure 2(b). We have also applied HingeProt to the complexed structure of the calmodulin (PDB 2bbm). The RMSD between the two structures is 14.78 Å. Nevertheless, HingeProt detected the same hinge location [Fig. 2(b)].

Glutamine binding protein

The protein undergoes a conformational change from the open (PDB 1wdnA) to the closed form (PDB 1gggA) upon the binding of glutamine. The RMSD between the two conformations is 5.33 Å. Figure 2(c) shows the HingeProt partition of the open form structure and Figure 2(d) of the closed form. The first slowest mode correctly recognizes the two domains that move one relative to other, as two rigid parts. Note that the domains are intertwined, i.e. one of the rigid parts consists of two sequential rigid fragments.

Hemoglobin

The protein undergoes significant conformational changes during the transition between different states. Hemoglobin is a tetramer consisting of two α and two β chains. Normal mode analyses have been shown to provide satisfactory results in the prediction of the movement of hemoglobin from the deoxy T-state to oxy R-state.[43] It is also known that the transition from T-state to R-state is initiated by the hinge movement at the interfacial area, which can be identified by the most global movement of hemoglobin.[44] HingeProt divides the molecule into two rigid parts, each rigid part representing one α and one β chain (Fig. 3). The hinge residues are not identified in this case, since the hinge region is located across the interface of the two α–β dimers.

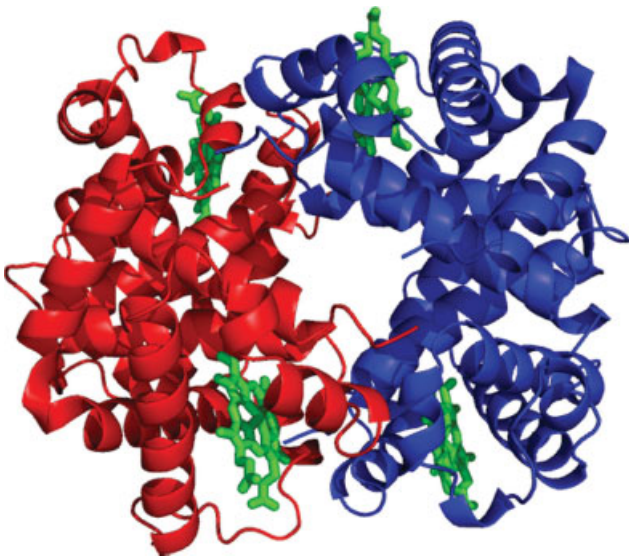

Figure 3

A multichain structure example—hemoglobin (1bz0: four Chains). HingeProt divides the protein structure into two rigid parts (red and blue), each rigid part representing one α and one β chain. The movement of rigid parts assists in orienting the four heme groups (green) toward oxygen binding.

Large scale statistical validation

Validation procedure

To test the accuracy and validity of the method, we devised the following procedure. Given two different conformations of the same protein, if there is a hinge movement between the two conformations, a rigid alignment of their structures will not be able to cover the full length of the protein structure. Only part of the structure will be aligned. For example, when aligning the calmodulin protein in two different conformation (PDB codes: 1rfj:A and 1up5:A), only half of the protein length can be aligned. Calmodulin consists of two domains with the hinge between them and only one of the domains is aligned in a rigid alignment. However, if we know the location of the hinge, we can split the structure into rigid fragments and align the full length of the protein. This fact was used for HingeProt validation. If the hinges recognized by HingeProt can describe the conformational change between the two structures, then the size of the alignment of the rigid fragments will be higher as compared to the rigid alignment of the whole structure. Our validation procedure follows this simple strategy. The procedure is schematically described in Figure 4.

Given a pair of structures of the same protein, HingeProt was applied to the first structure and rigid parts with hinges were identified for the first two GNM modes as described in Methods section. The structure was split into the identified rigid parts and each rigid part was

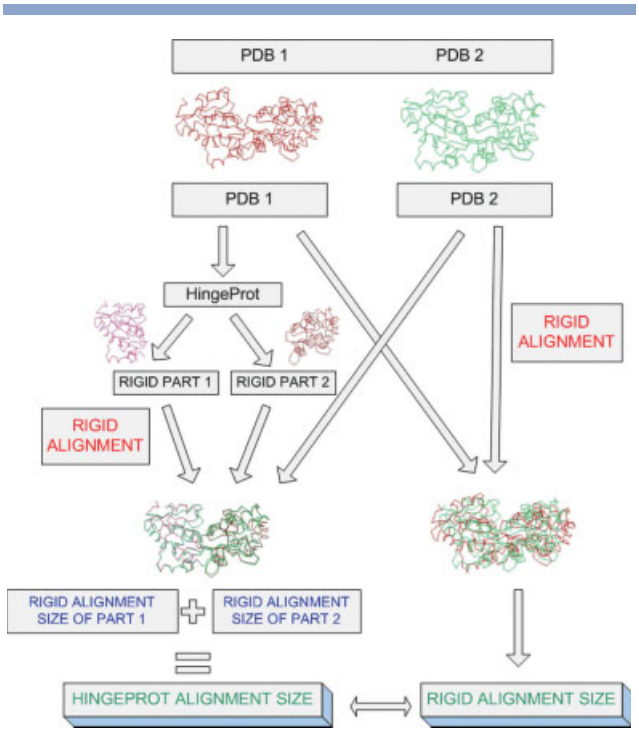

Figure 4

The test procedure scheme: The procedure devised to test the success of the methodology is depicted in the figure. The procedure is repeated for each protein. When a large dataset is considered, the procedure gives a good measure of the validity of the results obtained by HingeProt.

aligned to the second structure using MultiProt.[45] We computed a sequence order dependent alignment restricting the RMSD between the aligned residues to 1.5 Å. The alignment size and RMSD were recorded for each rigid part. The total alignment size was calculated as a sum of the aligned residues of each rigid fragment and the RMSD between all aligned residues of the first protein chain and the second protein chain was computed. The aim was to determine the increase in the alignment size before and after the partition into rigid parts according to HingeProt.

Datasets

Two datasets of protein conformations were used for this type of validation. The first consisted of 29 protein chain pairs collected from the Molecular Movements Database[46] and the second dataset consisted of 1080 protein chain pairs collected from the DynDom database.[19] All pairs are chosen such that the alignment percentage of the two structures is 90% or less to guarantee that further increase in alignment percentage is possible. The first dataset was collected based on movements described in the literature. The second dataset was constructed automatically considering all the available PDB structures, and therefore includes considerably higher number of protein movements and is especially useful for statistical data extraction and large-scale verification.

Validation results

Alignment coverage. We measured the success of the prediction by the percentage of the alignment coverage, i.e. the total number of aligned residues divided by the size of the protein. The results are summarized in Table I.

For the first dataset, which consists of 29 protein pairs with rigid alignment coverage of 90% or less, on average the improvement in the alignment was 18.8% for the slowest mode and 15% for the second slowest mode. For the second dataset, which contains 1080 protein pairs with a rigid alignment coverage of 90% or less, the improvement in alignment percentage was even better, 21.1% for the slowest mode and 17.7% for the second slowest mode. To provide further detail, using the results of HingeProt for 81 protein pairs from the second dataset increases the alignment from an average 73.3% to 99% and higher in the first slowest mode, which means nearly all residues are aligned.

RMSD. The increase in percent alignment coverage is accompanied by a decrease in the RMSD. The RMSD for protein pairs decreases from an average of 0.67 Å to an average of 0.59 Å following the first slowest mode fragmentation and to an average of 0.57 Å after the second slowest mode fragmentation for the first dataset. A similar decrease is also observed for the second dataset, from an average of 0.68 Å to an average of 0.55 Å for both of the two slowest modes.

Number of rigid parts. The average number of rigid parts is 2.07 for the first slowest mode and 3.15 for the second mode. The number of hinge residues per hinge region is 1.86 and 3.7 for the first and second modes, respectively. This means that many of the rigid parts are made of several sequential fragments and as a result the hinge regions include several residues. We computed the average number of residues per hinge region as the number of hinge residues/(number of rigid parts − 1). This yielded 1.7 hinge residues per hinge region for both modes. We also examined a possible correlation between

Table I

Alignment Coverage Percentage

	Rigid alignment	HingeProt alignment mode 1	HingeProt alignment mode 2
Molecular movements DB	66.9%	85.7%	81.9%
DynDom DB	67.6%	88.7%	85.3%

DOI 10.1002/prot

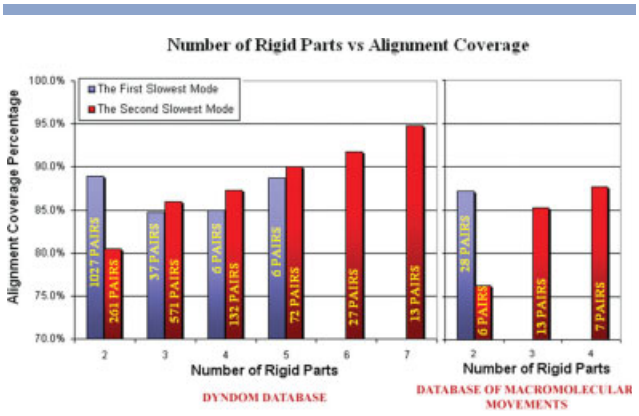

Figure 5

The correlation between coverage percentage and number of hinges determined by HingeProt. The graphs represent the percent alignment of the method for each dataset averaged according to number of rigid parts. Note that for the first slowest mode, cases with high number of rigid parts are not available although the size of the dataset is quite large. For DynDom database consisting of 1080 PDB pairs, it is not possible to talk about a positive correlation between the number of hinges and alignment percent for the first slowest mode, whereas for the second mode a positive correlation exists. For the cases in Database of MacroMolecular Movements, although the limited number of cases point to the same conclusions, it is not possible to examine the existence of a correlation due to lack of information. These considerations lead us to the conclusion that a more stable fragmentation can be reached for the first slowest mode.

the number of rigid parts and the alignment coverage percentage. Figure 5 shows the average alignment coverage percentage for each number of rigid parts. The number of cases used to calculate this average is noted on the plot. Most of the cases in the first mode (95%) are partitioned into two rigid parts and 90% of the cases in the second mode are partitioned to four or less rigid parts. The increase in alignment size is expected to be the consequence of correct hinge location identification. Correlation between the number of rigid parts and percent coverage is not anticipated. No correlation in the first mode and a slight correlation in the second mode compensated by a general increase in the number of rigid parts were observed. This behavior can be explained by the nature of the mode analysis. The first mode represents larger amplitude, lower frequency movement in which the cooperative domain movement is more dominant.

Score. We calculated the partition score (see Methods section) for all the test cases in the datasets. The average score for the first slowest mode is 0.94 and for the second it is 0.82. This emphasizes the point that the first mode discovers hinge movement better than the second with less hinge residues and rigid parts. The partition score can also be used for exclusion of hinge movements in proteins. Partitioning to rigid parts with score below 0.8 with many hinge residues does not usually represent global hinge movement. For example, when applying HingeProt to bacteriorhodopsin (PDB code 1brd), which exhibits shear movement, the score of the first and second mode was 0.8 with two and three hinge residues. This low score indicates that the protein does not exhibit a global hinge movement and the partition of HingeProt is not relevant. For a number of cases (PDB codes 1e0sA, 1cbuB, 4dfrA, 1a8vA, 1k9pA, 1bjyA, 3tms) known to undergo shear movement at fragments smaller than domains, the score varied between 0.60 and 0.90 with an average of 0.84 and 0.71 for the first and the second slowest modes, respectively; on the other hand, this number varied between 0.85 and 1.00 for domains' hinge dominant movements. The low scores indicate that the proteins exhibit either shear or limited hinge movement and therefore the partition of HingeProt may not be relevant. We also noticed that in many of these cases, the hinge region usually includes a high number of spatially clustered hinge residues that restrict the movement of the protein.

Running times. The method was able to provide the results in reasonable time scales even for the largest proteins in the dataset. On a 3.2 GHz machine, for proteins of 1400, 900, 600, 300 amino acid lengths, the running times are 240 s, 90 s, 40 s, and less than 10 s, respectively.

Comparison to FlexProt

We compared our results with the FlexProt method[18] that detects hinges when two protein conformations are given. The method performs flexible alignment of protein structures, detecting hinges through the alignment process. FlexProt was applied to each pair of protein structures in the dataset with an RMSD threshold of the alignment of 1.5 Å. FlexProt outputs alignments for 0, 1, 2, 3, and 4 hinges. As expected, the alignment with zero hinges resulted in a 70% of the protein length coverage. We selected an alignment with the largest size from all alignments with all possible number of hinges. This resulted in 97% of the structure aligned with an average RMSD of 1.4 Å. HingeProt succeeds in covering 89% with the first slowest mode by using only a single protein structure. The RMSD results also reveal this difference. The RMSDs of the alignments are considerably higher for FlexProt, whereas with our method the increase in alignment size is accompanied by a decrease in RMSD.

CONCLUSION

We present a method for fast and reliable detection of hinge regions and rigid parts given a single protein conformation. The method employs elastic network models, GNM and ANM, for fully automated identification of hinge residues. HingeProt is available through a convenient user-friendly interface. The results are given within seconds to several minutes for proteins consisting of up to a few thousands of residues. The server provides easy

to interpret visualization of the results. We expect the method to be useful in a range of potential applications such as flexible protein–protein and protein–ligand docking, and fitting flexible hinge-bent protein structures into EM density maps and refining the EM structures, and also comprehension of functional mechanisms of macromolecular structures and assemblies. To this end, we will continue to improve the HingeProt webserver to make it particularly useful for the docking and EM-targeted applications.

ACKNOWLEDGMENTS

The content of this publication does not necessarily reflect the view or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organization imply endorsement by the U.S. Government. We would like to thank Ozge Kurkcuoglu for her help in preparing the server logo.

REFERENCES

1. Janin J. Assessing predictions of protein–protein interaction: the CAPRI experiment. Protein Sci 2005;14:278–283.
2. Delarue M, Sanejouand Y. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the Elastic Network Model. J Mol Biol 2002;320:1011.
3. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc 2005;127:9632–9640.
4. Tobi D, Bahar I. Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state. PNAS 2005;102:18908–18913.
5. May A, Zacharias M. Accounting for global protein deformability during protein–protein and protein–ligand docking. Biochim Biophys Acta 2005;1754(1/2):225–231.
6. Sandak B, Haim WJ, Nussinov R. Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. Proteins 1998;32:159–174.
7. Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. Geometry-based flexible and symmetric protein docking. Proteins 2005;60:224–231.
8. Bonvin AM. Flexible protein–protein docking. Curr Opin Struct Biol 2006;16:194–200.
9. Delarue M, Sanejouand YH. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. PNAS 2004;101:6957.
10. Suhre K, Navaza J, Sanejouand YH. NORMA: a tool for flexible fitting of high resolution protein structures into low resolution electron microscopy derived density maps. Acta Cryst D 2006;62:1098–1100.
11. Ming D, Kong Y, Lambert MA, Huang Z, Ma J. How to describe protein motion without amino acid sequence and atomic coordinates? PNAS 2002;99:8620.
12. Tama F, Wriggers W, Brooks CL. Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. J Mol Biol 2002; 321:297.
13. Tama F, Miyashita O, Brooks CL, 3rd. Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryo-EM. J Struct Biol 2004;147:315–326.
14. Lesk AM, Chothia C. Mechanisms of domain closure in proteins. J Mol Biol 1984;174:175–191.
15. Gerstein M, Echols N. Exploring the range of protein flexibility, from a structural proteomics perspective. Curr Opin Chem Biol 2004;8:14–19.
16. Kovacs JA, Chacon P, Abagyan R. Predictions of protein flexibility: first-order measures. Proteins 2004;56:661–668.
17. Boutonnet N, Rooman MJ, Wodak SJ. Automatic analysis of protein conformational changes by multiple linkage clustering. J Mol Biol 1995;253:633–647.
18. Shatsky M, Nussinov R, Wolfson HJ. Flexible protein alignment and hinge detection. Proteins 2002;48:242–256.
19. Qi G, Lee R, Hayward S. A comprehensive and non-redundant database of protein domain movements. Bioinformatics 2005;21: 2832–2838.
20. Jacobs DJ, Rader AJ, Kuhn LA, Thorpe MF. Protein flexibility predictions using graph theory. Proteins 2001;44:150–165.
21. Thorpe MF, Lei M, Rader AJ, Jacbos DJ, Kuhn LA. Protein flexibility and dynamics using constraint theory. J Mol Graph Model 2001;19:60–69.
22. Go N, Toguti T, Nishikawa T. Dynamics of small globular proteins in terms of low-frequency vibration modes. PNAS 1983;80:3696–3700.
23. Kitao A, Go N. Investigating protein dynamics in collective coordinate space. Curr Opin Struct Biol 1999;9:164–169.
24. Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181.
25. Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. Curr Opin Struct Biol 2005;15:586–592.
26. Ma J. Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Curr Opin Struct Biol 2005;13:373–380.
27. Tama F, Brooks CL. Symmetry, form, and shape: guiding principles for robustness in macromolecular machines. Annu Rev Biophys Biomol Struct 2006;35:115–133.
28. Damm KL, Carlson HA. Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures. Biophys J 2006;90:4558–4573.
29. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistics. Proteins 2002;48:682–695.
30. Cao ZW, Xue Y, Han LY, Xie B, Zhou B, Zheng CJ, Lin HH, Chen YZ. MoViES: molecular vibrations evaluation server for analysis of fluctuational dynamics of proteins and nucleic acids. Nucleic Acids Res 2004;32(Web Server issue):W679–W685.
31. Lindahl E, Azuara C, Koehl P, Delarue M. NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. Nucleic Acids Res 2006;34:W52–W56.
32. Hollup SM, Salensminde G, Reuter N. WEBnm@: a web application for normal mode analysis of proteins. BMC Bioinformat 2005;6:52.
33. Suhre K, Sanejouand YH. ElNemo: a normal mode web-server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res 2004;32:610–614.
34. Lee-Wei Y, Liu X, Jursa CJ, Holliman M, Rader AJ, Karimi HA, Bahar I. iGNM: a database of protein functional motions based on Gaussian network model. Bioinformatics 2005;21:2978–2987.
35. Haliloglu T, Bahar I, Erman B. Gaussian dynamics of folded proteins. Phys Rev Lett 1997;79:3090–3093.
36. Zheng W, Brooks BR. Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes. Biophysics 2006;90:4327–4336.
37. Eyal E, Yang LW, Bahar I. Anisotropic network model: systematic evaluation and a new web interface. Bioinformatics 2006;22:2619–2627.

 DOI 10.1002/prot

38. Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D, Gerstein M. Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. Protein Sci 2005;14:633–643.
39. Wako H, Kato M, Endo S. ProMode: a database of normal mode analyses on protein molecules with a full-atom model. Bioinformatics 2004;20:2035–2043.
40. Ben-Dor A, Shamir R, Yakhini Z. Clustering gene expression patterns. J Comput Biol 1999;6(3/4):281–297.
41. Gerstein M, Lesk AM, Chothia C. Structural mechanisms for domain movements in proteins. Biochemistry 1994;33:6739–6749.
42. Atilgan AR, Durell SR, Jernigan RL, Demirel MC, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
43. Mouawad L, Perahia D. Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures. J Mol Biol 1996;258:393–410.
44. Xu C, Tobi D, Bahar I. Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T → R2 transition. J Mol Biol 2003;333:153–168.
45. Shatsky M, Nussinov R, Wolfson HJ. A method for simultaneous alignment of multiple protein structures. Proteins Struct Funct Bioinformat 2004;56:143–156.
46. Flores S, Echols N, Milburn D, Hespenheide B, Keating K, Lu J, Wells S, Yu EZ, Thorpe M, Gerstein M. The database of macromolecular motions: new features added at the decade mark. Nucleic Acids Res 2006;34:296–301.