StoneHinge: Hinge Prediction by Network Analysis of Individual Protein Structures

Kevin S. Keating[1], Samuel C. Flores[2,3,†], Mark B. Gerstein[1,3,4], Leslie A. Kuhn[5]*

[1]Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut
[2]Department of Physics, Yale University, New Haven, Connecticut
[3]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut
[4]Department of Computer Science, Yale University, New Haven, Connecticut
[5]Departments of Biochemistry and Molecular Biology, Computer Science and Engineering, and Physics and Astronomy, Michigan State University, East Lansing, Michigan

*Correspondence to: Leslie A. Kuhn, Protein Structural Analysis and Design Lab, 502C Biochemistry Building, Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824-1319; phone: (517) 353-8745; fax: (517) 353-9334; KuhnL@msu.edu; http://www.bmb.msu.edu/~kuhn/

† Current address: Department of BioEngineering, Stanford University, Stanford, California

Running title: Hinge Prediction by Network Analysis

Manuscript Pages:	25
Supplemental Material Pages:	2
Tables:	2
Figures:	4
Supplemental Material Figures:	1

ABSTRACT

Hinge motions are important for molecular recognition, and knowledge of their location can guide the sampling of protein conformations for docking. Predicting domains and intervening hinges is also important for identifying structurally self-determinate units and anticipating the influence of mutations on protein flexibility and stability. Here we present *StoneHinge,* a novel approach for predicting hinges between domains using input from two complementary analyses of non-covalent bond networks: *StoneHingeP,* which identifies domain-hinge-domain signatures in *ProFlex* constraint counting results, and *StoneHingeD*, which does the same for *DomDecomp* Gaussian network analyses. Predictions for the two methods are compared to hinges defined in the literature and by visual inspection of interpolated motions between conformations in a series of proteins. For *StoneHingeP*, all the predicted hinges agree with hinge sites reported in the literature or observed visually, although some predictions include extra residues. Furthermore, no hinges are predicted in six hinge-free proteins. On the other hand, *StoneHingeD* tends to over-predict the number of hinges, while accurately pinpointing hinge locations. By determining the consensus of their results, *StoneHinge* improves the specificity, predicting 11 out of 13 hinges found both visually and in the literature for nine different open protein structures, and making no false-positive predictions. By comparison, a popular hinge detection method that requires knowledge of both the open and closed conformations finds 10 of the 13 known hinges, while predicting four additional, false hinges.

Flexibility is critical to both the structure and function of proteins and impacts areas from protein folding (Hayward et al. 1997) to prion propagation (Scheibel and Lindquist 2001) to structure-based drug design (Schneidman-Duhovny et al. 2005; Cozzini et al. 2008). Hinge motions can be particularly important for interactions between proteins and small molecules by exposing the interaction surface to the ligand (Gerstein et al. 1994). Therefore, predicting hinges has the potential to enhance structure-based ligand discovery.

Large-scale conformational changes such as hinge motion can be divided into three classes on the basis of size: fragment, domain, and subunit (Gerstein et al. 1999; Krebs 2001). Fragment motion refers to movements of small regions such as surface loops, whereas domain motion generally involves large conformational changes between covalently linked domains. Subunit motion refers to movement between polypeptide chains, often associated with allostery (Gerstein et al. 1999; Krebs 2001). Here, we focus on identifying hinges between domains, which can be considerably harder to analyze and predict than fragment motions. Even within this class of motion, hinges have been defined in different ways. Here, we define a hinge as a region of localized, internal motion of the main chain between two domains of a protein, rather than between a domain and a flexible region, such as a loop.

The *StoneHinge* method, presented here, computes the consensus of two network-based hinge predictors, *StoneHingeP* and *StoneHingeD*. *StoneHingeP* identifies overconstrained and underconstrained regions in the bond network, corresponding to mutually rigid or flexible regions. All residues intervening between a pair of rigid domains are defined as the hinge. These hinges may have rigid inclusions such as β turns or short helices, as the movement of these substructures can also contribute to hinge motion. This definition allows for disseminated motions, where domain movement is caused by small changes spread over a number of residues.

3

Such hinges are known for a number of proteins, including Bence-Jones protein (Maiorov and Abagyan 1997), lysine/arginine/ornithine (LAO) binding protein (Maiorov and Abagyan 1997), and T4 lysozyme (Zhang et al. 1995; Hayward and Berendsen 1998). For instance, a twelve-residue, strap-like hinge connects domains in the LAO binding protein (Maiorov and Abagyan 1997). Other hinge-prediction methods such as *FlexProt* (Shatsky et al. 2002) and *StoneHingeD* (presented here) define hinges as pivot points between two consecutive residues. *StoneHingeD* defines each hinge as a fixed point between connected regions undergoing large-scale opening and closing modes in *DomDecomp* (Kundu et al. 2004). Such pivot-like hinges are known for proteins including adenylate kinase (Gerstein et al. 1993), inorganic pyrophosphatase (Ahn et al. 2001), and ribose binding protein (Bjorkman and Mowbray 1998).

To account for backbone flexibility during ligand docking, programs such as *FlexDock* have been developed to sample hinge rotations (Schneidman-Duhovny et al. 2005). *FlexDock* partitions a protein into rigid regions with intervening hinges and docks the other molecule against each rigid region. The success of *FlexDock* for protein-protein docking at CAPRI (http://capri.ebi.ac.uk/) demonstrates the utility of incorporating hinge motion (Schneidman-Duhovny et al. 2005). However, this approach requires that the hinges first be defined by another method.

Identifying Hinges

Hinge *detectors*, such as *FlexProt* (Shatsky et al. 2002)*, require two conformations of the protein and analyze which residues stay mutually rigid between the conformers or change in conformation; the latter are identified as hinges. On the other hand, hinge *predictors* such as the *StoneHinge* algorithms presented here identify the locations of hinges given a single structure as

input. Such methods are more widely applicable and focus more on the intrinsic flexibility of the protein rather than on ligand-induced changes.

A number of other domain and hinge prediction methods are available, using a variety of criteria to identify the domains bordering a hinge, such as compactness (Zehfus 1994; Holm and Sander 1998), structural redundancy (Holm and Sander 1998), abundance of inter-domain contacts (Holm and Sander 1994; Islam et al. 1995; Siddiqui and Barton 1995), presence of hydrophobic cores (Swindells 1995a; b), distribution of electron density (Painter and Merritt 2006), and inter- versus intra-domain potential energy (Flores and Gerstein 2007). Alternatively, hinge prediction can be based on sequence statistics (Flores et al. 2007) or assessed by normal mode analysis (Hinsen 1998; Kundu et al. 2004).

StoneHinge

StoneHinge predicts hinges based on the consensus between *StoneHingeP* and *StoneHingeD* results. *StoneHingeP* is unique in using constraint counting from rigidity theory, as implemented in *ProFlex,* as the basis for predictions. *ProFlex* (successor to the *FIRST* method (Jacobs et al. 2001)) analyzes flexibility using a three-dimensional constraint counting algorithm that decomposes a protein structure into rotatable and non-rotatable bonds. This analysis is based on bond rotational constraints placed by covalent and non-covalent bonds in the network. The non-covalent constraints are reminiscent of the elastic interactions in normal mode analysis or the contacts analyzed in geometric approaches, both of which also represent the tertiary interactions stabilizing the protein. *StoneHingeP* analyzes the *ProFlex* results to identify the energy at which the protein structure first decomposes into two rigid regions (domains) of significant size, connected by a flexible hinge. If two domains containing at least 20 residues are

not found, then hinge motion is not predicted.

StoneHingeD uses the *DomDecomp* Gaussian Network Model (GNM) normal mode analysis to identify domains (Kundu et al. 2004). A reduced protein representation is used, modeling favorable contacts as an elastic network of springs connecting pairs of alpha carbons (Bahar et al. 1997). *StoneHingeD* defines hinges by identifying residues that are fixed points, or undergo the least motion, along the direction of maximal lowest-frequency (largest-scale) motion in the *DomDecomp* analysis.

The consensus predictor, *StoneHinge*, assigns the residues of a *StoneHingeD* prediction as a consensus hinge if they fall within five residues of a *StoneHingeP* prediction. *StoneHingeD* is used to assign hinge residues because it was observed during training (see Methods) to be more sensitive to the precise location of hinges, whereas requiring consensus with *StoneHingeP* greatly reduces false-positive hinge predictions. *StoneHinge* consensus results are presented for a series of protein structures solved both in open and closed conformations, allowing assessment of the effects of input conformation on prediction. *StoneHinge* predictions are then compared to hinges reported in the literature, as well as hinges selected by visual inspection of protein motion between the open and closed states (Flores et al. 2008) using the *Molecular Motions Database* morph viewer (Krebs and Gerstein 2000; Flores et al. 2006). Open conformations (normally a ligand-free form of the protein) may allow more accurate predictions, as the domains are usually separated in space. This makes distinguishing the domains easier, both visually and in the bond network analysis. However, *StoneHinge* predictions from the open and closed states are typically comparable, and the method yields more specific prediction of inter-domain hinges than either of its component methods. *StoneHinge* predictions are then compared with those of the popular hinge detector, *FlexProt* (not to be confused with the *ProFlex* constraint-counting method),

which performs domain superposition between open and closed states to locate any hinges (Shatsky et al. 2002; 2004).

RESULTS

Following initial tests on training data (see Methods), *StoneHinge* was run on nine protein structures in both open and closed conformations (Table 1). The resulting predictions are summarized alongside the literature-reported and visually-assigned (Flores et al. 2008) hinge residues, along with hinges predicted using each of the component methods, *StoneHingeP* and *StoneHingeD,* and those detected by the *FlexProt* domain superposition method (Table 2 and Figure 1). Sensitivity of the hinge predictions was assessed against the intersection of the literature and visually-assigned hinges, referred to as the gold standard hinges (see Methods).

Cyclic AMP dependent protein kinase (CAPK)

The hinge motion of CAPK involves multiple stretches of the backbone and was defined in the literature by molecular dynamics analysis (Tsigelny et al. 1999). *StoneHinge* predictions in both the open and closed states match the gold standard hinge. *FlexProt* predicts residues 330-331, which are ten residues from a visually-defined hinge that is spatially adjacent to the literature hinges.

Bence-Jones protein

The two domains involved in hinge motion are easily distinguished by eye. There is a single stretch of the backbone passing between them, and the domains remain slightly separated in the closed conformation. The predictions of all programs on the open and closed states match

the literature hinge, which was assigned using a variant of Siddiqui and Barton's method (Maiorov and Abagyan 1997).

Lysine/arginine/ornithine (LAO) binding protein

The literature hinge comprises two backbone segments (Maiorov and Abagyan 1997). *StoneHinge* predictions agree with the literature and visually-assigned hinges for both the open and closed states. For the closed state, in which the *StoneHingeP* component made no prediction, *StoneHingeD* predictions were used (following the protocol described in Figure 3). *FlexProt* reports a hinge close in sequence to the first literature hinge, but based on visual inspection, it is located in a rigid region distal from the hinge.

Adenylate kinase

Here the hinge region again incorporates two segments of the backbone, with two additional hinges identified in the literature but not by visual inspection (Gerstein et al. 1993). *StoneHinge* predictions from both the open and the closed states miss the gold standard hinges, while detecting one of the literature-only hinges. *FlexProt* identifies both of the gold standard hinges. Adenylate kinase presents a challenging case in which the literature and visual hinge definitions disagree on two hinges, and in which the *StoneHinge* component predictors also do not show clear consensus.

Glutamine binding protein

During this hinge motion, which involves two segments of the backbone, the second domain swings approximately 90°. In both the open and closed states, a face of this domain lies

against the first domain. By observing this motion, the two domains can be distinguished clearly. However, because the two domains fold together, *StoneHingeP* predicts the majority of the protein as belonging to a single large domain, and therefore predicts no hinges. This also occurs for *StoneHingeD* predictions on the closed state, but it correctly predicts hinges in the open structure. Because *StoneHingeP* makes no predictions on the open structure, *StoneHingeD* predictions are used alone for *StoneHinge* (Fig. 3). Both the consensus *StoneHinge* and *FlexProt* hinge identifications agree well with the visually and crystallographically defined hinges for the open state (Hsiao et al. 1996), but *StoneHinge* misses both hinges in the closed state.

DNA polymerase β

A protease sensitive hinge region between residues 82 and 86 comprises the single literature hinge (Kumar et al. 1990). However, visual inspection identified three domains that move relative to each other, with less motion in the second hinge at residues 263-264 (Flores et al. 2008). Because *StoneHingeP* identifies hinges between two rigid domains, it misses the secondary hinge. The length of the primary hinge is significantly over-predicted in the open and closed forms, due to the inclusion of α helices bordering the hinge. *StoneHingeD* correctly predicts both hinges but also predicts two hinges in the middle of the second domain. This case particularly indicates the strengths of using *StoneHingeP* and *StoneHingeD* together to pinpoint the location of hinges that are in agreement, and cancel extraneous predictions. *FlexProt* does not detect the secondary hinge, but it accurately detects the first.

Calmodulin

This is an atypical hinge that consists of an α helix which partially unwinds in the center,

as determined using NMR (Ikura et al. 1992) and computational analyses (Meador et al. 1992; van der Spoel et al. 1996) as well as visual examination of the conformational transition (Flores et al. 2008). *StoneHinge* predictions on both forms of the protein match this hinge, whereas *FlexProt* identifies four hinges, one of which is adjacent to the correct hinge.

Inorganic pyrophosphatase

This hinge includes one segment of backbone, and the two domains are easily distinguished in the open conformation. The literature hinge is based on a comparison of crystal structures (Ahn et al. 2001). *StoneHinge* predictions from the open and closed structures agree with the literature and visually-defined hinge. *FlexProt* detects this hinge but also identifies an extra one.

Ribose Binding Protein

Here, the hinge involves multiple stretches of backbone. The literature hinges are based on the comparison of the closed structure with several open structures (Bjorkman and Mowbray 1998). *StoneHinge* predictions agree with the two gold standard hinges. *FlexProt* detects the first hinge region, but misses the second. The predictions for ribose binding protein are mapped onto the three-dimensional structure of the protein in Figure 2, with the literature and visually-defined hinges appearing in panel (G) for comparison.

Summary of StoneHinge Consensus Predictions

All gold standard (consensus of literature and visually-assigned) hinges were predicted by *StoneHinge* in the open states of the nine proteins, except for two in adenylate kinase. The

same hinges were missed in the closed state of adenylate kinase, and two were missed in the closed glutamine binding protein; however, in the open state of this protein, both were predicted. No extra hinges were predicted in any protein.

Negative Controls

To assess whether *StoneHinge* would falsely predict intra-domain loops as hinges, the algorithm was tested on six proteins with low main-chain RMSD values between ligand-bound and ligand-free structures, indicating no significant backbone motion (Raymer et al. 1997; Mayer et al. 2003). Both the apo and ligand-bound structures were tested for three of the proteins. As with all predictions, ligands were removed first. These negative controls are listed in Table 2, along with their *StoneHingeP* rigid cluster (predicted domain) sizes and *StoneHingeD* domain count. For all six proteins, *StoneHingeP* predicted the second rigid domain to have fewer than twenty residues, thus correctly predicting no hinge motion (see Methods). Similarly, *StoneHingeD* predicted no hinges in all cases except for Protein A, due to identifying at most one domain in these proteins. Therefore, the consensus *StoneHinge* prediction for all proteins except Protein A indicated no hinges.

DISCUSSION

The results of applying *StoneHingeP*, *StoneHingeD*, and *StoneHinge* to 27 structures (7 of which were used in training, with the other 20 reserved for testing), indicate they are effective predictors. *StoneHingeP* was shown to be highly sensitive: nearly all hinges in the gold standard set were identified for the structures in which *StoneHingeP* made a prediction. As shown in Figure 1, any additional predictions corresponded to a hinge either reported in the literature or

determined by visual inspection. However, in both DNA polymerase β structures and the open form of LAO binding protein, *StoneHingeP* significantly over-predicted the lengths of hinges. These over-predictions were caused by predicting several smaller rigid regions in place of the second rigid domain. Similarly, in glutamine binding protein and closed LAO binding protein, *StoneHingeP* missed the hinges due to not identifying the second rigid domain. Predicting more flexibility than is apparent from experimental results is typically due to irregular secondary structure or suboptimal hydrogen-bond stereochemistry, which weakens the network of hydrogen bonds analyzed by *ProFlex* constraint counting. *StoneHingeD*, the other component algorithm, was shown to be a precise predictor of known hinges. Each predicted hinge was specified as a pair of residues, although additional hinges were predicted for DNA polymerase β, ribose binding protein, and the closed structures of CAPK and adenylate kinase.

StoneHingeP is also notable as the only algorithm using rigidity theory to locate hinges. Thus, it is likely to provide useful and independent information when used in consensus predictions, as shown by *StoneHinge*. Its consensus of *StoneHingeD* and *StoneHingeP* results was both sensitive and specific. All predictions corresponded to a gold standard hinge, and all gold standard hinges were correctly predicted, excepting adenylate kinase and the closed form of glutamine binding protein. Additionally, *StoneHinge* runs quickly due to the integer constraint counting algorithm of *StoneHingeP* and the reduced protein representation of *StoneHingeD*. It takes roughly two minutes to analyze a 300-residue protein on a Pentium IV, including determining the optimal energy level. The algorithm is also completely automated and does not require human intervention to interpret the results. As such, it is suitable for high-throughput use.

Using StoneHinge to Guide Flexible Docking

As with other hinge predictors, *StoneHinge* results can guide the sampling of protein main-chain flexibility during ligand docking. One approach is to use *StoneHinge* together with *FlexDock*, which can dock ligands into proteins with pre-specified hinges (Schneidman-Duhovny et al. 2005). As only the location of the hinge must be specified, not the direction of motion, predictions from *StoneHinge* are suitable for use with *FlexDock*. Additionally, the network of rotatable bonds derived by *StoneHinge* can also be used with *ROCK* (Lei et al. 2004), which generates a panel of protein conformations by sampling the hinge angles within stereochemically favorable ranges. Ligands can then be independently docked into each protein conformation using *SLIDE* (Zavodszky et al. 2002), as was demonstrated for fully flexible docking of the cyclic peptide cyclosporin with its receptor, cyclophilin A (Zavodszky et al. 2004). *StoneHingeP* already provides the input files needed by *ROCK*, making the process easy to automate.

As an alternative sampling method, the *ProFlex* component of *StoneHingeP* can be combined with a rotations and translations of blocks normal mode analysis to sample conformation changes in proteins (Ahmed and Gohlke 2006) as input to docking. The output of *StoneHinge* can also be used as a starting point for *MBO(N)D* molecular dynamics simulations (Chun et al. 2000). By dividing the protein into rigid bodies and flexible regions (e.g., using the *StoneHinge* predictions), this algorithm reduces the computational time required for molecular dynamics simulations. We also foresee other advances and applications of the *StoneHinge* algorithm, including recognizing and sampling active-site loop motion.

StoneHinge determines the consensus of the *StoneHingeP* and *StoneHingeD* automated hinge predictions, which are described in greater detail in this section.

StoneHingeP Overview

StoneHingeP uses *ProFlex* (Jacobs et al. 2001) to predict a protein's rigid clusters, which are groups of atoms that are constrained by the bond network and do not move relative to each other. However, one rigid cluster may move relative to another, like two stones linked by a tether. These clusters may range in size from a few atoms to nearly the size of the entire protein. *StoneHingeP* prepares the input, runs the rigidity analysis module of *ProFlex*, and determines the appropriate energy level cut-off for hydrogen bonds and salt bridges to include in the network. The residue ranges within the largest two rigid clusters are then defined and filtered by *StoneHingeP*, based on whether they meet the domain size criterion, and the hinge residues between the domains are identified, possibly including rigid motifs such as turns. Figure 3 shows a flowchart of the algorithm, with each step explained below. *StoneHinge* performs all necessary preparation and analysis steps without intervention. This makes *StoneHinge* easy to use and enables automatic analysis of a database of structures.

Protein Preparation

Including buried waters. *ProFlex* works best when internal (entirely buried) water molecules are present in the structure but surface-bound water molecules are removed. The hydrogen bonds to internal waters can be important for the protein structure, whereas including surface waters tends to lead to overestimation of the rigidity of the protein (Jacobs et al. 2001).

All predictions presented here are made with internal water molecules only, as determined by PRO_ACT (Williams et al. 1994). Predictions were also run using no water molecules (data not shown), as is done on the automated *StoneHinge* server. While the overall flexibility analysis changed slightly, the hinge predictions were largely the same. Negligible differences have also been noted in a previous *ProFlex*-based analysis following removal of structural water from the Ras-Raf complex (Gohlke et al. 2004).

Removing ligands. All inhibitors and/or cofactors are removed from the structure before the prediction, as they tend to cross-link and rigidify the domains relative to each other. Ligand removal also enables identification of the intrinsic, rather than ligand-induced, flexibility of the protein. *StoneHinge* automatically removes all non-polypeptide inhibitors by stripping heteroatom records from the PDB file. However, the protein will remain in a ligand-bound conformation, at least in terms of side-chain orientations. As this can influence the network of interactions analyzed by *StoneHinge*, we recommend analyzing ligand-free, open structures when possible. *StoneHinge* also removes metal ions by default, as bonds between a metal ion and the protein are difficult to deduce automatically from the structure. However, for biologically relevant bound metals, neglecting them may result in increased flexibility, affecting the hinge predictions. It is possible to avoid this by preparing the bond network using the standalone version of *ProFlex* (available by contacting proflex@sol.bch.msu.edu), but this procedure has not yet been automated as part of the *StoneHinge* predictions.

Adding hydrogen atoms. *ProFlex* requires polar hydrogen atoms to be present in the structure, as they are used to calculate hydrogen-bond energies. *StoneHingeP* calls *GROMACS* (Lindahl et al. 2001) to add polar hydrogen atoms to the protein structure and optimize their positions by steepest descent energy minimization for 100 2-femtosecond steps. *GROMACS* is

freely available under the GNU Public License from http://www.gromacs.org.

      Calculating the hydrogen-bond dilution profile. *ProFlex* calculates a hydrogen-bond dilution profile (Hespenheide et al. 2002), which *StoneHingeP* analyzes to identify an energy level optimal for determining domains and intervening hinges (Supplemental Figure 1). To calculate the profile, the protein's hydrogen bonds are broken one by one, from weakest to strongest, and the constraint counting algorithm is run after each bond is broken. This simulates incremental thermal denaturation of the structure, as the calculated temperature rises and hydrogen bonds weaker than the current energy level are broken. The protein commonly appears as a single large rigid region when the simulated temperature is low and very weak hydrogen bonds and salt bridges are included. The structure then gradually breaks into two or more rigid regions (often corresponding to the known native state), before going through a cooperative phase transition to a completely flexible chain as the simulated temperature rises (Rader et al. 2002). Hydrophobic interactions are maintained throughout the process, as the strength of these interactions increases somewhat with modest increases in temperature (Tanford 1980).

Domain and Hinge Prediction Using StoneHingeP

      Selecting an energy level and domain size. There is no single energy level that corresponds to the native state of all proteins, in terms of rigid and flexible regions. This may be a result of the different conditions and forcefields under which protein structures are determined. Because our goal is to predict hinges, we wish to identify an energy level in which the protein contains at least two rigid domains of substantial size (20 or more residues). Rigid clusters of this size typically correspond to supra-secondary structures (e.g., two packed helices, rather than one long helix), while smaller rigid clusters do not. The appropriateness of this size threshold

was assessed using the training set (Table 1) and fixed before running *StoneHingeP* on the test set proteins. *StoneHingeP* identifies the energy level at which the second-largest rigid cluster is maximal in size, measured by the number of residues with mutually rigid backbones. This typically occurs when the protein has just relaxed from a single large rigid cluster into two clusters, each potentially representing a domain. If more than one energy level has a second-largest rigid cluster of the same size, the more rigid (lower-temperature) state of the protein is selected. The two largest rigid clusters are then checked against the 20-residue size criterion. If either cluster is smaller than this, then no hinge is predicted.

  Identifying hinges in StoneHingeP. Hinges are predicted as the residues intervening between two rigid clusters meeting the above criteria. Predicted hinges can be one or multiple stretches of residues, as rigid clusters do not necessarily consist of contiguous amino acids. For example, in the open form of adenylate kinase (Supplemental Figure 1), the largest rigid cluster (called cluster 1 or the first rigid domain) is predicted as residues 5 through 12, 17 through 113, 163 through 195, and 198 through 217. These residues are proximal in space, but not contiguous in sequence. The second-largest rigid cluster (cluster 2 or the second rigid domain) is predicted as residues 124 through 158. Thus, in primary structure, cluster 2 occurs between two regions of cluster 1 and the hinges are predicted as the two stretches of residues between these clusters (residues 113 through 124 and 158 through 163). This corresponds to one flexible region of the backbone passing from rigid cluster 1 to rigid cluster 2, then a second flexible region later in sequence passing from rigid cluster 2 back to cluster 1. If cluster 2 also contained discontinuous sequence, as in the case of cAMP dependent protein kinase, then more than two hinges could be predicted.

  Training and test sets. During development, *StoneHingeP* was frequently tested against

closed and open cyclic AMP dependent protein kinase; the open forms of Bence-Jones protein, lysine-arginine-ornithine binding protein, and adenylate kinase; and two proteins without hinges, the immunoglobulin domain of protein G and hydropterin pyrophosphokinase. The training results led to adjustments in the domain size threshold and the determination of appropriate energy levels for analysis. Thus, the *StoneHingeP* results of these training proteins may be favorably biased. *StoneHingeP* was also run on additional proteins once the algorithm was finalized. The results on this test set may be more representative of its general performance. The training and test set proteins are designated in Table 1.

Multi-chain proteins. At present, *StoneHinge* only predicts on single-chain proteins. The *StoneHingeP* methodology is applicable to oligomeric proteins, but interpretation of the results is not straightforward. For example, the two largest rigid sub-domains of the protein may fall on different chains. Assigning a hinge between these two domains would not make sense, as they are not covalently connected. It is also possible that one chain contains two domains with an intervening hinge, whereas another chain contains a single large domain not involved in the hinge motion. In this case, selecting the two largest domains of the complex would not necessarily select the domains bounding the hinge. Due to these difficulties, *StoneHinge* will generate a *StoneHingeP* hydrogen-bond dilution plot (e.g., Supplementary Materials Figure 1) for oligomeric proteins without making a prediction. The dilution plot can then be manually interpreted to locate potential hinges.

StoneHingeD Predictions

StoneHingeD predictions utilize *DomDecomp* (Kundu et al. 2004), which is freely available at http://stonehinge.molmovdb.org or by request to Prof. George Phillips

(phillips@biochem.wisc.edu). *StoneHingeD* makes predictions based on *DomDecomp* GNM normal mode analysis, which identifies domains as groups of residues undergoing anti-correlated motion at the lowest non-zero frequency. The last residue assigned to one domain is consecutive with the first residue of the next. Thus, hinges are identified as pairs of residues at anti-correlated domain boundaries.

StoneHinge Consensus Prediction

To combine *StoneHingeP and StoneHingeD* results, the residues in a *StoneHingeD* prediction are assigned as a hinge if they fall within five residues of a *StoneHingeP* prediction. (*N.B.* For cases presented, the same results are obtained if a more stringent criterion - falling within two residues - is used.) For proteins in which hinges are only predicted by one method, the predictions of that method are used alone, as in the cases of glutamine binding protein and the closed state of LAO binding protein. However, these predictions have lower confidence, as both component methods tend to overpredict the number of hinge residues.

FlexProt Predictions

StoneHinge results were compared with hinges identified by the *FlexProt* server of Shatsky, Nussinov, and Wolfson (http://bioinfo3d.cs.tau.ac.il/FlexProt/), which detects hinges by superimposing domains from open and closed structures (Shatsky et al. 2002; 2004). *FlexProt* requires both the open and closed conformations of the protein as input, and the hinges are identified as the pair of residues bordering adjacent rigid fragments. The hinge assignments were made using default settings: a maximum of 3.0 Å RMSD between matched fragments in the two structures and a minimum size of 15 amino acids for matched fragments. Hinges reported

here are based on the alignment with lowest RMSD between all matched fragments.

Assessing Prediction Sensitivity

There can be variability when assessing hinge motion from experimental or simulation data, as evidenced by the differences observed between the literature and visually-defined hinges. Therefore, we assessed the sensitivity of the predictors using the intersection of the hinges from the literature and visual hinge sets. Hinges from these sets were said to intersect, or represent consensus, if they included amino acids within five residues of each other. These hinges are referred to as the gold standard set.

StoneHinge Availability

StoneHinge is available online at http://stonehinge.molmovdb.org, as shown in Figure 4. Proteins can be submitted for prediction online, and the predicted hinge locations will be mapped onto the three-dimensional structure and can be rotated in *Jmol*. Hinge predictions are also given in tabular form for *StoneHingeP*, *StoneHingeD*, and the *StoneHinge* consensus. Additionally, the hydrogen-bond dilution plot used for *StoneHingeP* predictions can be displayed. For those preferring to run the software locally, the *StoneHinge* Perl code is also available for download under the GNU Public License at the above web site. It can be run locally on Linux or Unix systems and modified as desired.

Protein Structures

All protein structures were retrieved from the Protein Data Bank (Berman et al. 2000). Protein Data Bank entries are as follows: 1CTP (Karlsson et al. 1993), 1ATP (Zheng et al. 1993),

4BJL (Huang et al. 1996), 2LAO, 1LST (Oh et al. 1993), 2AK3 (Diederichs and Schulz 1991), 1AKE (Muller and Schulz 1992), 1GGG (Hsiao et al. 1996), 1WDN (Sun et al. 1998), 2BPG (Pelletier et al. 1994), 1BPD (Sawaya et al. 1994), 1CFD (Kuboniwa et al. 1995), 1CLL (Chattopadhyaya et al. 1992), 1K23, 1K20 (Ahn et al. 2001), 1URP (Bjorkman and Mowbray 1998), 2DRI (Bjorkman et al. 1994), 1PGD (Gallagher et al. 1994), 1HKA (Xiao et al. 1999), 1BCK (Kallen et al. 1998), 2APR (Suguna et al. 1987a), 3APR (Suguna et al. 1987b), 2CLA (Gibbs et al. 1990), 3CLA (Leslie 1990), 2SGA (Moult et al. 1985), and 5SGA (James et al. 1980).

COMPETING INTERESTS

None declared.

ACKNOWLEDGMENTS

REFERENCES

Ahmed, A., and Gohlke, H. 2006. Multiscale modeling of macromolecular conformational changes combining concepts from rigidity and elastic network theory. *Proteins* **63:** 1038-1051.

Ahn, S., Milner, A.J., Futterer, K., Konopka, M., Ilias, M., Young, T.W., and White, S.A. 2001. The "open" and "closed" structures of the type-C inorganic pyrophosphatases from Bacillus subtilis and Streptococcus gordonii. *J Mol Biol* **313:** 797-811.

Bahar, I., Atilgan, A.R., and Erman, B. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* **2:** 173-181.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* **28:** 235-242.

Bjorkman, A.J., Binnie, R.A., Zhang, H., Cole, L.B., Hermodson, M.A., and Mowbray, S.L. 1994. Probing protein-protein interactions. The ribose-binding protein in bacterial transport and chemotaxis. *J Biol Chem* **269:** 30206-30211.

Bjorkman, A.J., and Mowbray, S.L. 1998. Multiple open forms of ribose-binding protein trace the path of its conformational change. *J Mol Biol* **279:** 651-664.

Chattopadhyaya, R., Meador, W.E., Means, A.R., and Quiocho, F.A. 1992. Calmodulin structure refined at 1.7 A resolution. *J Mol Biol* **228:** 1177-1192.

Chun, H.M., Padilla, C.E., Chin, D.N., Watanabe, M., Karlov, V.I., Alper, H.E., Soosaar, K., Blair, K.B., Becker, O.M., Caves, L.S.D., et al. 2000. MBO(N)D: A multibody method for long-time molecular dynamics simulations. *Journal of Computational Chemistry* **21:** 159-184.

Cozzini, P., Kellogg, G.E., Spyrakis, F., Abraham, D.J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L.A., Morris, G.M., et al. 2008. Target flexibility: an emerging consideration in drug discovery and design. *J Med Chem* **51:** 6237-6255.

DeLano, W.L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA.

Diederichs, K., and Schulz, G.E. 1991. The refined structure of the complex between adenylate kinase from beef heart mitochondrial matrix and its substrate AMP at 1.85 A resolution. *J Mol Biol* **217:** 541-549.

Flores, S., Echols, N., Milburn, D., Hespenheide, B., Keating, K., Lu, J., Wells, S., Yu, E.Z., Thorpe, M., and Gerstein, M. 2006. The Database of Macromolecular Motions: new features added at the decade mark. *Nucleic Acids Res* **34:** D296-301.

Flores, S.C., and Gerstein, M.B. 2007. FlexOracle: predicting flexible hinges by identification of stable domains. *BMC Bioinformatics* **8:** 215.

Flores, S.C., Keating, K.S., Painter, J., Morcos, F.G., Nguyen, K., Merritt, E., Kuhn, L.A., and Gerstein, M. 2008. HingeMaster: Normal mode hinge prediction approach and integration of complementary predictors. *Proteins* **73:** 299-319.

Flores, S.C., Lu, L.J., Yang, J., Carriero, N., and Gerstein, M.B. 2007. Hinge Atlas: relating protein sequence to sites of structural flexibility. *BMC Bioinformatics* **8:** 167.

Gallagher, T., Alexander, P., Bryan, P., and Gilliland, G.L. 1994. Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **33:** 4721-4729.

Gerstein, M., Jansen, R., Johnson, T., Tsai, J., and Krebs, W. 1999. Studying Macromolecular Motions in a Database Framework: From Structure to Sequence. In *Rigidity Theory and Applications*. (eds. M. Thorpe, and P. Duxbury). Kluwer Academic/Plenum Publishers.

Gerstein, M., Lesk, A.M., and Chothia, C. 1994. Structural mechanisms for domain movements in proteins. *Biochemistry* **33:** 6739-6749.

Gerstein, M., Schulz, G., and Chothia, C. 1993. Domain closure in adenylate kinase. Joints on either side of two helices close like neighboring fingers. *J Mol Biol* **229:** 494-501.

Gibbs, M.R., Moody, P.C., and Leslie, A.G. 1990. Crystal structure of the aspartic acid-199 --> asparagine mutant of chloramphenicol acetyltransferase to 2.35-A resolution: structural consequences of disruption of a buried salt bridge. *Biochemistry* **29:** 11261-11265.

Gohlke, H., Kuhn, L.A., and Case, D.A. 2004. Change in protein flexibility upon complex formation: analysis of Ras-Raf using molecular dynamics and a molecular framework approach. *Proteins* **56:** 322-337.

Hayward, S., and Berendsen, H.J. 1998. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. *Proteins* **30:** 144-154.

Hayward, S., Kitao, A., and Berendsen, H.J. 1997. Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins* **27:** 425-437.

Hespenheide, B.M., Rader, A.J., Thorpe, M.F., and Kuhn, L.A. 2002. Identifying protein folding cores from the evolution of flexible regions during unfolding. *J Mol Graph Model* **21:** 195-207.

Hinsen, K. 1998. Analysis of domain motions by approximate normal mode calculations. *Proteins* **33:** 417-429.

Holm, L., and Sander, C. 1994. Parser for protein folding units. *Proteins* **19:** 256-268.

Holm, L., and Sander, C. 1998. Dictionary of recurrent domains in protein structures. *Proteins* **33:** 88-96.

Hsiao, C.D., Sun, Y.J., Rose, J., and Wang, B.C. 1996. The crystal structure of glutamine-binding protein from Escherichia coli. *J Mol Biol* **262:** 225-242.

Huang, D.B., Ainsworth, C.F., Stevens, F.J., and Schiffer, M. 1996. Three quaternary structures for a single protein. *Proc Natl Acad Sci U S A* **93:** 7017-7021.

Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B., and Bax, A. 1992. Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* **256:** 632-638.

Islam, S.A., Luo, J., and Sternberg, M.J. 1995. Identification and analysis of domains in proteins. *Protein Eng* **8:** 513-525.

Jacobs, D.J., Rader, A.J., Kuhn, L.A., and Thorpe, M.F. 2001. Protein flexibility predictions using graph theory. *Proteins* **44:** 150-165.

James, M.N., Sielecki, A.R., Brayer, G.D., Delbaere, L.T., and Bauer, C.A. 1980. Structures of product and inhibitor complexes of Streptomyces griseus protease A at 1.8 A resolution. A model for serine protease catalysis. *J Mol Biol* **144:** 43-88.

Kallen, J., Mikol, V., Taylor, P., and Walkinshaw, M.D. 1998. X-ray structures and analysis of 11 cyclosporin derivatives complexed with cyclophilin A. *J Mol Biol* **283:** 435-449.

Karlsson, R., Zheng, J., Xuong, N., Taylor, S.S., and Sowadski, J.M. 1993. Structure of the mammalian catalytic subunit of cAMP-dependent protein kinase and an inhibitor peptide displays an open conformation. *Acta Crystallogr D Biol Crystallogr* **49:** 381-388.

Krebs, W. 2001. The database of macromolecular motions: A standardized system for analyzing and visualizing macromolecular motions in a database framework. In *Molecular Biophysics and Biochemistry*, pp. 288. Yale University, New Haven, CT 06511.

Krebs, W.G., and Gerstein, M. 2000. The morph server: a standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res* **28:** 1665-1675.

Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C.B., and Bax, A. 1995. Solution structure of calcium-free calmodulin. *Nat Struct Biol* **2:** 768-776.

Kumar, A., Widen, S.G., Williams, K.R., Kedar, P., Karpel, R.L., and Wilson, S.H. 1990. Studies of the domain structure of mammalian DNA polymerase beta. Identification of a discrete template binding domain. *J Biol Chem* **265:** 2124-2131.

Kundu, S., Sorensen, D.C., and Phillips, G.N., Jr. 2004. Automatic domain decomposition of proteins by a Gaussian Network Model. *Proteins* **57:** 725-733.

Lei, M., Zavodszky, M.I., Kuhn, L.A., and Thorpe, M.F. 2004. Sampling protein conformations and pathways. *J Comput Chem* **25:** 1133-1148.

Leslie, A.G. 1990. Refined crystal structure of type III chloramphenicol acetyltransferase at 1.75 A resolution. *J Mol Biol* **213:** 167-186.

Lindahl, E., Hess, B., and van der Spoel, D. 2001. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **7:** 306-317.

Maiorov, V., and Abagyan, R. 1997. A new method for modeling large-scale rearrangements of protein domains. *Proteins* **27:** 410-424.

Mayer, K.L., Earley, M.R., Gupta, S., Pichumani, K., Regan, L., and Stone, M.J. 2003. Covariation of backbone motion throughout a small protein domain. *Nat Struct Biol* **10:** 962-965.

Meador, W.E., Means, A.R., and Quiocho, F.A. 1992. Target enzyme recognition by calmodulin: 2.4 A structure of a calmodulin-peptide complex. *Science* **257:** 1251-1255.

Moult, J., Sussman, F., and James, M.N. 1985. Electron density calculations as an extension of protein structure refinement. Streptomyces griseus protease A at 1.5 A resolution. *J Mol Biol* **182:** 555-566.

Muller, C.W., and Schulz, G.E. 1992. Structure of the complex between adenylate kinase from Escherichia coli and the inhibitor Ap5A refined at 1.9 A resolution. A model for a catalytic transition state. *J Mol Biol* **224:** 159-177.

Oh, B.H., Pandit, J., Kang, C.H., Nikaido, K., Gokcen, S., Ames, G.F., and Kim, S.H. 1993. Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. *J Biol Chem* **268:** 11348-11355.

Painter, J., and Merritt, E.A. 2006. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallographica* **D62:** 439-450.

Pelletier, H., Sawaya, M.R., Kumar, A., Wilson, S.H., and Kraut, J. 1994. Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science* **264:** 1891-1903.

Rader, A.J., Hespenheide, B.M., Kuhn, L.A., and Thorpe, M.F. 2002. Protein unfolding: rigidity lost. *Proc Natl Acad Sci U S A* **99:** 3540-3545.

Raymer, M.L., Sanschagrin, P.C., Punch, W.F., Venkataraman, S., Goodman, E.D., and Kuhn, L.A. 1997. Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm. *J Mol Biol* **265:** 445-464.

Sawaya, M.R., Pelletier, H., Kumar, A., Wilson, S.H., and Kraut, J. 1994. Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. *Science* **264:** 1930-1935.

Scheibel, T., and Lindquist, S.L. 2001. The role of conformational flexibility in prion propagation and maintenance for Sup35p. *Nat Struct Biol* **8:** 958-962.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R., and Wolfson, H.J. 2005. Geometry-based flexible and symmetric protein docking. *Proteins* **60:** 224-231.

Shatsky, M., Nussinov, R., and Wolfson, H.J. 2002. Flexible protein alignment and hinge detection. *Proteins* **48:** 242-256.

Shatsky, M., Nussinov, R., and Wolfson, H.J. 2004. FlexProt: alignment of flexible protein structures without a predefinition of hinge regions. *J Comput Biol* **11:** 83-106.

Siddiqui, A.S., and Barton, G.J. 1995. Continuous and discontinuous domains: an algorithm for the automatic generation of reliable protein domain definitions. *Protein Sci* **4:** 872-884.

Suguna, K., Bott, R.R., Padlan, E.A., Subramanian, E., Sheriff, S., Cohen, G.H., and Davies, D.R. 1987a. Structure and refinement at 1.8 A resolution of the aspartic proteinase from Rhizopus chinensis. *J Mol Biol* **196:** 877-900.

Suguna, K., Padlan, E.A., Smith, C.W., Carlson, W.D., and Davies, D.R. 1987b. Binding of a reduced peptide inhibitor to the aspartic proteinase from Rhizopus chinensis: implications for a mechanism of action. *Proc Natl Acad Sci U S A* **84:** 7009-7013.

Sun, Y.J., Rose, J., Wang, B.C., and Hsiao, C.D. 1998. The structure of glutamine-binding protein complexed with glutamine at 1.94 A resolution: comparisons with other amino acid binding proteins. *J Mol Biol* **278:** 219-229.

Swindells, M.B. 1995a. A procedure for detecting structural domains in proteins. *Protein Sci* **4:** 103-112.

Swindells, M.B. 1995b. A procedure for the automatic determination of hydrophobic cores in protein structures. *Protein Sci* **4:** 93-102.

Tanford, C. 1980. *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, 2 ed. John Wiley & Sons, New York.

Tsigelny, I., Greenberg, J.P., Cox, S., Nichols, W.L., Taylor, S.S., and Ten Eyck, L.F. 1999. 600 ps molecular dynamics reveals stable substructures and flexible hinge points in cAMP dependent protein kinase. *Biopolymers* **50:** 513-524.

van der Spoel, D., de Groot, B.L., Hayward, S., Berendsen, H.J., and Vogel, H.J. 1996. Bending of the calmodulin central helix: a theoretical study. *Protein Sci* **5:** 2044-2053.

Williams, M.A., Goodfellow, J.M., and Thornton, J.M. 1994. Buried waters and internal cavities in monomeric proteins. *Protein Sci* **3:** 1224-1235.

Xiao, B., Shi, G., Chen, X., Yan, H., and Ji, X. 1999. Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents. *Structure* **7:** 489-496.

Zavodszky, M.I., Lei, M., Thorpe, M.F., Day, A.R., and Kuhn, L.A. 2004. Modeling correlated main-chain motions in proteins for flexible molecular recognition. *Proteins* **57:** 243-261.

Zavodszky, M.I., Sanschagrin, P.C., Korde, R.S., and Kuhn, L.A. 2002. Distilling the essential features of a protein surface for improving protein-ligand docking, scoring, and virtual screening. *J Comput Aided Mol Des* **16:** 883-902.

Zehfus, M.H. 1994. Binary discontinuous compact protein domains. *Protein Eng* **7:** 335-340.

Zhang, X.J., Wozniak, J.A., and Matthews, B.W. 1995. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. *J Mol Biol* **250:** 527-552.

Zheng, J., Trafny, E.A., Knighton, D.R., Xuong, N.H., Taylor, S.S., Ten Eyck, L.F., and Sowadski, J.M. 1993. 2.2 A refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. *Acta Crystallogr D Biol Crystallogr* **49:** 362-365.

Table 1. *Proteins Analyzed*

Protein	Form	Set	PDB	Chain	Size of Protein[a]	StoneHingeP Size of Largest Rigid Cluster	StoneHingeP Size of Second Rigid Cluster	StoneHingeD Number of Domains
Hinged proteins:								
CAPK	open	training	1CTP	E	350	111	67	2
	closed	training	1ATP	E	350	141	42	2
Bence-Jones Protein	open	training	4BJL	B	216	104	96	2
	closed	test	4BJL	A	216	98	95	2
LAO Binding Protein	open	training	2LAO		238	112	21	2
	closed	test	1LST		239	171	12	2
Adenylate Kinase	open	training	2AK3	A	226	158	35	2
	closed	test	1AKE	A	214	121	24	2
Glutamine Binding Protein	open	test	1GGG	A	226	189	13	2
	closed	test	1WDN	A	226	102	12	1
DNA Polymerase β	open	test	2BPG	A	335	235	39	4
	closed	test	1BPD		335	206	70	3
Calmodulin	open	test	1CFD		148	64	54	2
	closed	test	1CLL		148	30	28	2
Inorganic Pyrophosphatase	open	test	1K23	A	307	181	112	2
	closed	test	1K20	A	311	148	96	2
Ribose Binding Protein	open	test	1URP	C	271	92	59	3
	closed	test	2DRI		271	108	84	3
Control Proteins:								
Ig domain of Protein G		training	1PGB		56	41	0	NP[b]
Hydropterin Pyrophosphokinase		training	1HKA		158	97	16	1
Cyclophilin A		test	1BCK	A	165	81	8	1
Rhizopuspepsin	apo	test	2APR		325	296	11	1
	inhibitor removed	test	3APR	E	325	296	11	1
Chloramphenicol Acetyltransferase	apo	test	2CLA		213	129	14	1
	chloramphenicol removed	test	3CLA		213	102	16	1
Proteinase A	apo	test	2SGA		181	174	4	2
	tetrapeptide removed	test	5SGA	E	181	175	5	2

Table 1 Caption:

All hinged and non-hinged (control) proteins analyzed by *StoneHinge*. The training set of proteins was tested frequently during the development of *StoneHinge*, and thus their results may be favorably biased. The test set contains proteins not analyzed by *StoneHinge* until the algorithm had been finalized. Additionally, *StoneHingeP* and *StoneHingeD* were analyzed on control proteins with no known hinge bending motion. Correct prediction of the absence of hinges in these proteins is highlighted in gray. For *StoneHingeP*, the presence of fewer than 20 residues in the second largest rigid cluster leads to a no-hinge prediction, with the closed state of LAO binding protein and both states of glutamine binding protein being misclassified. For *StoneHingeD*, the prediction of a single domain leads to a no-hinge prediction, with the closed state of glutamine binding protein and both states of proteinase A being misclassified. All sizes are given as the number of amino acid residues.

[a]There are several proteins for which the open and closed states differ slightly in size. This is due to differences in the form of the protein crystallized and does not affect the residue numbering used here.
[b]*StoneHingeD* did not complete successfully when running on the Ig domain of Protein G, so no prediction was made.

Table 2. *Hinge Predictions*

Protein	Literature Hinges	Visually Defined Hinges	StoneHinge Open	StoneHinge Closed	StoneHingeP Open	StoneHingeP Closed	StoneHingeD Open	StoneHingeD Closed	FlexProt Hinges
CAPK		31-32				32-43	25-26		
	98-99				98-99	81-97			
	106-107				104-107	104-105			
	120-126	119-120	124-125	127-128	121-126	121-126	124-125	127-128	
								170-171	
								189-190	
		319-320					317, 327[d]	315-316	330-331
Bence-Jones Protein	108-116	110-111[a]	112-113	114-115	110-114	110-114	112-113	114-115	110-111
LAO Binding Protein	89-93	90-91	89-90	89-90	89-120	NP[c]	89-90	89-90	83-84
	182-194	192-193	191-192	191-192	191-194		191-192	191-192	
Adenylate Kinase							32-33		
							57-58		
	118		113-114	110-111	113-124	111-123	113-114	110-111	
	124	125-126							120-121
	157-163	162-163			158-163	155-160			162-163
	177						173-174	170-171	
Glutamine Binding Protein	85-89	89-90	86-87	NP[b]	NP[c]	NP[c]	86-87	NP[e]	87-88
	181-185	178-179	182-183				182-183		180-181
DNA polymerase β	82-86	88-89	98-99	87-88	79-107	49-91	98-99	87-88	88-89
							151-152	141-142	
							214-215		
		263-264					256-257	260-261	
Calmodulin									24-25
	73-82	80-81	77-78	78-79	77-80	77-89	77-78	78-79	65-73
									101-102
									129-131
Inorganic Pyrophosphatase									120-121
	188-190	183-184	188-189	187-188	188-191	189-205	188-189	187-188	182-183
Ribose Binding Protein	103-104	103-104	103-104	103-104	88-105	104-106	103-104	103-104	100-101
							157-158	156-157	
	235-236	235-236	235-236	235-236	235-238	236-239	235-236	235-236	
	264-265				254-268	265-269			

Table 2 Caption:

The hinges predicted by *StoneHinge*, *StoneHingeP*, *StoneHingeD*, and *FlexProt*. Results from *FlexProt* are presented here for comparison and are not part of the *StoneHinge* consensus predictions. Each row of the table corresponds to a hinge, and each known or predicted hinge is represented as a range of residues. Visually defined hinge ranges are from Flores et al. 2008. Prediction sensitivity was assessed using the gold standard hinges, which are the intersection of the literature and visually defined hinges, shown in gray. A graphical representation of these results appears in Figure 1.

[a]Flores et al. 2008 does not include Bence-Jones protein. However, these hinge residues were selected using the same criteria and methodology.

[b]NP indicates no hinge prediction was made by either *StoneHingeP* or *StoneHingeD* for the closed form of glutamine binding protein, and thus no hinges were predicted by *StoneHinge*.

[c]NP indicates no prediction was made because the closed conformation of LAO binding protein and both conformations of glutamine binding protein were predicted to have fewer than twenty residues in their second largest rigid cluster.

[d]Residues 318-326 are disordered in the open state of CAPK. Thus, this hinge includes two non-consecutive residues.

[e]NP indicates no prediction was made because *DomDecomp* predicts a single domain for the closed form of glutamine binding protein.

FIGURE LEGENDS

Figure 1

A visual representation of the hinges predicted by *StoneHinge* compared to the visually defined and literature hinges as well as those detected by *FlexProt*. Each line represents the protein sequence from N-terminal (left) to C-terminal (right), with residue numbers appearing below and secondary structure denoted above. Colored diamonds represent hinge residues reported by each method.

Figure 2

Ribose binding protein hinge predictions highlighted in color on the backbone, represented as a ribbon diagram. Shown are the *StoneHingeP* predictions on the (A) open and (B) closed conformations, *StoneHingeD* predictions on the (C) open and (D) closed conformations, and *StoneHinge* consensus predictions on the (E) open and (F) closed conformations. (G) The gold standard hinges (consensus of literature and visually assigned hinges) are highlighted in blue, with an additional literature-defined hinge shown in cyan. (H) Hinges detected by *FlexProt* superposition between open and closed conformations are displayed. This figure was generated using PyMol (DeLano 2002).

Figure 3

A flow chart of the *StoneHinge* algorithm. All steps except the optional inclusion of buried water molecules are performed automatically by *StoneHinge*. For all steps that make use of an external program, the program name is noted in the lower half of the box. Details are given in the Methods.

Figure 4

Proteins can be submitted online for a *StoneHinge* analysis at

http://stonehinge.molmovdb.org. This view of the user interface shows predictions made on

adenylate kinase in the open form. The *StoneHinge* hinges are highlighted in the Jmol viewer,

which can also display the *StoneHingeP* and *StoneHingeD* hinges and domains. Predicted hinge

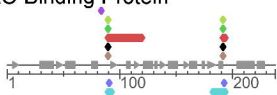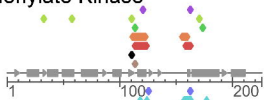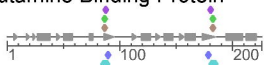residues are listed in the table below the molecular graphics.
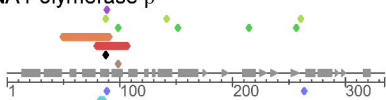
CAPK
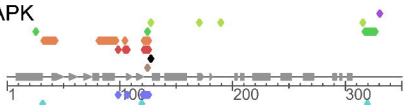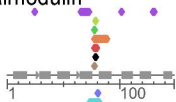
Bence-Jones Protein
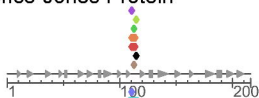
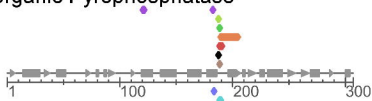LAO Binding Protein
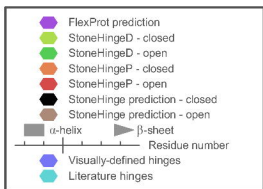
Adenylate Kinase

Glutamine Binding Protein

DNA Polymerase β

Calmodulin

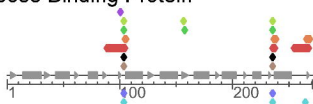Inorganic Pyrophosphatase

Ribose Binding Protein

Legend:
- FlexProt prediction
- StoneHingeD - closed
- StoneHingeD - open
- StoneHingeP - closed
- StoneHingeP - open
- StoneHinge prediction - closed
- StoneHinge prediction - open
- α-helix
- β-sheet
- Residue number
- Visually-defined hinges
- Literature hinges

Input protein structure

StoneHingeP

StoneHingeD

Include buried waters (optional)
PRO_ACT

Remove ligand if present

Add hydrogens
GROMACS

Calculate H-bond dilution profile
ProFlex

Determine appropriate energy level

Select the two largest rigid clusters

Calculate GNM domain predictions
DomDecomp

Determine the number of predicted domains

1 domain

≥ 2 domains

No StoneHingeD hinges detected.

Define StoneHingeD hinges as the residues between the predicted domains

Determine size of these rigid clusters

< 20 residues

≥ 20 residues

Rigid cluster is too small to be a domain. No StoneHingeP hinges detected.

Define StoneHingeP hinges as the residues between the two rigid clusters

If neither method detect a hinge, then the protein is predicted to have no hinge motion.

If one method detects a hinge and the other does not, report the predicted hinges but warn that the protein may not possess hinge motion.

If both methods detect hinges, define StoneHinge hinges as any StoneHingeD hinge within 5 residues of a StoneHingeP hinge.

Adenylate Kinase ([2AK3](#)) (Chain A)

StoneHinge ID: s583574-6543
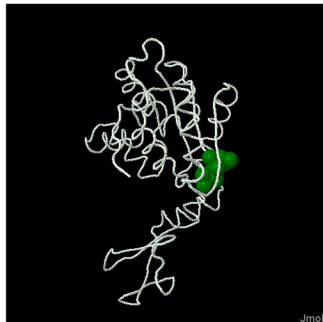
Jmol

Hinge Residues:

StoneHinge	113-114
StoneHingeP	113-124, 158-163
StoneHingeD	113-114, 173-174

[View detailed output](#)   [View H-bond dilution plot](#)