PROTEINS: Structure, Function, and Genetics 36:425–435 (1999)

Structural Principles Governing Domain Motions in Proteins

Steven Hayward*
BIOSON Research Institute, Laboratory of Biophysical Chemistry, University of Groningen, Groningen, The Netherlands

ABSTRACT **With the use of a recently developed method, twenty-four proteins for which two or more X-ray conformers are known have been analyzed to reveal structural principles that govern domain motions in proteins. In all 24 cases, the domain motion is a rotation about a physical axis created through local interactions both covalent and noncovalent. In many cases, two or more mechanical hinges separated in space create a stable hinge axis for precise control of the domain closure. The terminal regions of α-helices and β-sheets have been found to act as mechanical hinges in a significant number of cases. In some cases, the two terminal regions of neighboring strands of a single β-sheet can create a hinge axis, as can the two termini of a single α-helix. These two structures have been termed the "double-hinged β-sheet" and "double-hinged α-helix," respectively. A flexible loop that attaches one domain to another and through which the effective hinge axis passes is another construct that is used to create a hinge. Noncovalent interactions between segments remote along the polypeptide chain can also form hinges. In addition α-helices that preserve their hydrogen bonding structure when bent have been found to behave as mechanical hinges. It is suggested that these α-helices act as a store of elastic energy that drives the closing of domains for rapid capture of the substrate. If the repertoire of possible interdomain structures is as limited as this study suggests, the dynamic behavior of proteins could soon be predicted using bioinformatics techniques. Proteins 1999;36:425–435.**

Key words: hinge bending; hinge axis; X-ray conformers; dynamic domains

INTRODUCTION

Domain motions play an important role in the function of many proteins. In enzymes, the closure of one domain onto another can isolate a substrate in an environment with specific interactions. This allows the reaction to take place in a controlled way, and prevents the escape of intermediates.[1] Experimentally, such domain motions are observed from X-ray crystallography when the open structure of an enzyme is found in the absence of a substrate, or substrate analogue, and the closed structure in the complexed state. In some cases, a domain motion is observed in crystals of different symmetry,[2] or as in the case of the M6I mutant of bacteriophage T4 lysozyme, within the asymmetric unit of a single crystal.[3]

Most of the analysis of domain motions on X-ray conformers is carried out by the crystallographers themselves. In their review, Gerstein et al.[4] have brought together this information and made their own analysis. They see two basic types of elemental motions, hinge and shear, combining to produce the various kinds of domain motions, and they have assigned the domain motions of particular proteins to these two classes.

Lacking in the field generally has been an objective method that can be implemented to determine domains, hinge axes, and the regions involved in the "hinge bending" automatically. Recently, Wriggers and Schulten[5] developed a procedure that is based on Lesk's method, which sieves away residues that are not part of a "static core."[6] In their extension, multiple domains and hinge axes can be identified automatically. The method of Hayward and Berendsen [7]approaches the problem differently, in that domains are identified on the basis of their differing rotational properties. The method has an overall consistency in that the same principles of rigid body kinematics are used in all three stages: domain determination, hinge axis determination, and uniquely, the determination of the residues involved in the hinge bending. This method has been fully implemented in the computer program, "DynDom."

If, as in the vast majority of cases, only one X-ray conformer is available, then simulations can be performed to get information on the dynamic behavior of the protein concerned. Often then, a molecular dynamics simulation is performed without any experimental evidence to support its findings. However, given that there exist some doubts regarding force fields and convergence, more comparisons with experimental data are desirable. The set of proteins for which more than one X-ray conformer is known provides a resource of information on protein motions against which simulations can be tested. The comparison itself is not trivial, because a molecular dynamics simulation

Grant sponsor: European Union; Grant number: CT96–0166, "PROMOD."
*Correspondence: Dr. Steven Hayward, Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. E-mail: steve@chem.rug.nl
Received 4 January 1999; Accepted 19 March 1999

generates a welter of conformations, whereas one usually has no more than two conformations from X-ray. Recently, Professor Berendsen and coworkers have become interested in not only performing simulations, but also in developing methods to analyze trajectories for "important" motions. One development is the essential dynamics technique.[8] This is a multivariate analysis of the trajectory, which when combined with a DynDom analysis depicts the important motions in an easy-to-understand way. Such a comparison has been made for the small protein, Bacteriophage T4 lysozyme,[9] where a combined essential dynamics and DynDom analysis of 38 different crystal structures was compared with the same analysis of molecular dynamics simulations. Those results suggested that molecular dynamics simulations can be used to successfully predict functional motions in proteins. However, for larger proteins, molecular dynamics is, at present, unable to simulate time-scales of biological relevance.

Ironically, this work suggests an alternative approach: the possibility of predicting functional motions in domain proteins through the recognition of key structural motifs that control interdomain motions. Such an approach would not depend on the size of the protein. The existing experimental data therefore provide not only a challenge to simulation, but also an opportunity to develop a different approach.

In this work, DynDom has been used to analyze 24 proteins for which at least two X-ray conformations are known. The aim is not to describe the domain motion of each of the 24 proteins individually, but rather to find from this set general principles that may govern domain motions in proteins. For those interested in the results on individual proteins, please refer to the DynDom home page (at URL: http://rugmd0.chem.rug.nl/~steve/dyndom.html) from which the DynDom source code is also available.

METHODS

For this analysis, 24 proteins have been selected for which at least two X-ray conformations are known. All but one of these has been selected from the "Database of Macromolecular Movements" website (at URL: http://bioinfo.mbb.yale.edu/MolMovDB/) as it was in September 1997. At this site, which has been created following the review by Gerstein and colleagues, [4] protein domain motions are classified into groups according to type. These types are: predominantly shear, predominantly hinge, unclassified, not hinge or shear, and involving partial refolding of the structure. Proteins for this study have been selected from all of these groups except the last one. Exceptions also include those for which the sequences of the two structures are not identical. In the case of glutamate dehydrogenase, the conformational change implied by the two structures from the asymmetric unit used in this study is not the same as that described by Stillman et al.[11] In the case of T4 lysozyme, a large number of structures are available.[2] Recently, a multivariate analysis of 38 of the structures has been made.[9] In this work, the two structures implied by the first essential, or principal, mode have been used. Because of the great interest in the chaperone protein GroEL, whose structure has been recently solved,[12] the two different subunit structures in the *cis* and *trans* rings have also been included. Table I lists all the proteins successfully analyzed for domain motions together with their PDB codes and chain identifiers. Missing in this list is annexin V,[13] which was found not to satisfy one of the criteria for domain motion described below.

The program DynDom determines dynamic domains, interdomain screw axes, and regions involved in the interdomain bending. It has been described in detail elsewhere,[7] but a short description is given below to make clear some of the parameters required for its use.

Dynamic domains are determined by clusters of rotation vectors of main-chain segments. These main-chain segments are generated by use of a sliding window. The length of this window is one of the parameters of the program. The longer the window, the better local intrasegment rotations, which may have nothing to do with the overall domain motion, are eliminated. However, shorter windows are preferable if one wants to accurately identify specific residues involved in the interdomain bending.

It is feasible that a cluster of rotation vectors could correspond to two or more well-separated regions in the protein, rotating similarly. If the heavy atoms belonging to a cluster are not connected by a network of distances of less than 3.5 Å, then the cluster will be split up into domains that satisfy this criterion.

An additional parameter specifies the minimum value for the ratio of interdomain to intradomain displacement for any prospective domain pair. A domain division for which this ratio is less than the minimum value, for any connected domain pair, is rejected.

The final parameter specifies the minimum domain size. Any prospective domain having fewer residues than this minimum value is not considered for the next stage of analysis.

Table I lists window lengths that were chosen for each protein. Normally these were chosen to be as short as possible (i.e., to give a result), although sometimes a slightly longer window has been used to "clean up" the results. In all cases, the window lengths are very small compared with the size of the protein. In all cases except one, the parameter specifying the minimum ratio of interdomain to intradomain displacement was set to 1.0. Only in the case of TATA binding protein, where this ratio was found to be 0.92, was a value less than 1.0 permitted. The value for annexin V was 0.67 and this protein was therefore omitted. In all cases, the minimum domain size was chosen to be 10% of the total number of residues.

Interdomain screw axes can be classified into two extreme types: those parallel to the line joining the centers of mass of the two domains, and those perpendicular to this line. The former are called twist axes, and the latter are called closure axes. Any axis can be decomposed into components parallel or perpendicular to this line, and a percentage measure of its closure or twist motion can be defined from the square of the components.[14]

TABLE I. PDB Codes, Chain Identifiers, and Window Lengths

Protein	No. of residues	Conformer 1: PDB code, chain identifier	Conformer 2: PDB code, chain identifier	Window length
Adenylate kinase[23,24]	214	2eck, B	4ake, A	5
Alcohol dehydrogenase[25,26]	374	6adh, A	7adh	7
Aspartate aminotransferase[19,27]	410	1ama	9aat, A	5
cAMP-dependent protein kinase (catalytic domain)[28,29]	350	1atp, E	1ctp, E	5
Canine lymphoma immunoglobin[10]	658	1igt, A&B	1igt, C&D	3
Calmodulin[30,31]	146	1cll	1cdl, A	3
Citrate synthase[32]	437	5cts	5csc, A	5
Diptheria toxin[33,34]	535	1mdt, A	1ddt	3
Endothiapepsin[35,36]	326	5er2, E	4ape	9
Formate dehydrogenase[37]	374	2nad, A	2nac, A	5
Catabolite gene activator protein[18]	208	3gap, A	3gap, B	3
Glyceraldehyde-3-phosphate dehydrogenase[20,38]	333	1gd1, O	2gd1, O	5
Glutamate dehydrogenase[39]	449	1hrd, B	1hrd, C	5
GroEL[12]	525	1aon, A	1aon, H	5
Hexokinase[40,41]	458	2yhx	1hkg	9
Lactoferrin[42]	691	1lfh	1lfg	5
Lysine/arginine/ornithine (LAO) binding protein[43]	240	1lst	2lao	5
Maltodextrin binding protein[44,45]	370	1omp	1mbp	5
Phosphofructokinase[46]	319	1pfk, A	1pfk, B	5
T4 lysozyme[2,9]	164	1st essential mode from 38 structures		5
DNA polymerase β[47,48]	335	2bpg, A	1bpd	5
TATA-binding protein[49]	240	1tbp, A	1tbp, B	9
Tomato bushy stunt virus coat protein[50]	387	2tbv, C	2tbv, B	5
TRP repressor[51]	105	1wrp, R	2wrp, R	5

In this study, particular attention is paid to bending regions situated between the domains. DynDom determines regions of the *backbone* where a transition is seen between the rotational properties of the two dynamic domains.[7] These regions will be referred to as "bending regions" or "connecting regions" below. At bending regions, a rotational transition is observed. If the interdomain screw axis passes within 5.5 Å of an α-carbon atom of any of the residues of the bending regions, the axis will be called an "effective hinge axis," and the connecting region will be called a "mechanical hinge" because a true mechanical hinge is just where the hinge axis and the rotational transition coincide. Previously this cutoff was chosen to be 3 Å.[7] In the absence of any physical argument for this distance, and given that it is suspected that secondary structures are often involved in controlling domain motions, a distance that approximates the elemental distances found in α-helices and β-sheets would appear to be a reasonable choice. A distance of 5.5 Å is equivalent to the rise in one turn of an α-helix, about 0.5 Å shorter than the diameter of an α-helix, and about 0.8 Å longer than the distance between two neighboring strands of a β-sheet.

DynDom outputs all the details concerning the domain motion in a text file. In addition, an "arrow" file is created that can be appended to the PDB file of the first conformer, and a "RasMol" script file[15] for display of the protein together with the interdomain screw axis in the form of an arrow. Also, a "rotation vector" file is created that can also be viewed using RasMol for the analysis of the rotation of each backbone segment, or if a window length 1 is chosen, the rotation of each individual residue's N-CA-CB-C (PDB nomenclature) tetrahedron.

The program DSSP[16] was used to designate secondary structure type to residues, and the What If package[17] to determine contacts, salt-bridges, and side-chain hydrogen bonding and dihedrals.

RESULTS AND DISCUSSION

Effective Hinge Axes, Mechanical Hinges, Closure, Twist and Door Closing Model

Table II lists the number of domains found for each protein and the number of connections for each domain pair. In all, 29 domain pairs were found, giving 29 interdomain screw axes (i.e., five proteins had three domains). Table III shows the distribution of the number of connecting regions. It can be seen that two connecting regions per domain pair is the most common. The average is 2.7 connections per domain pair. Out of a total of 79 connecting regions, 57 are mechanical hinges. In 25 cases of the 29, the interdomain screw axis is an effective hinge axis. The fours cases that are not—endothiapepsin, the interlobe motion of lactoferrin, catabolite gene activator protein, and tomato bushy stunt virus coat protein—will be discussed below.

Of the 29 domain pairs, 21 have at least two connecting regions. Of these 21, 18 have at least two connections that are mechanical hinges. All of these 18 have predominantly closure motions (closure motion >50%; see Table II).

TABLE II. Domain Motion Data on Individual Proteins

Protein	No. of domains	No. of connecting regions	Angle of rotation (degrees)	Translation along axis (Å)	Percentage closure motion	Door-closing model	Type classification[a]
Adenylate kinase	3	2, 2	46, 53	1.3, 1.1	100, 96	Yes, Yes	—
Alcohol dehydrogenase	2	5	10	−0.1	87	Yes	DHB
Aspartate aminotransferase	2	3	15	−0.6	100	Yes	HL, EH
cAMP-dependent protein kinase (catalytic domain)	2	5	14	0.6	95	Yes	DHB, HL
Canine lymphoma immunoglobin	2	1	103	52.4	28	No	—
Calmodulin	2	1	154	−1.2	54	No	—
Citrate synthase	2	7	19	0.1	81	Yes	DHB, HL
Diptheria toxin	2	1	176	0.4	95	No	—
Endothiapepsin	2	2	4	0.1	25	No	—
Formate dehydrogenase	2	2	9	0.3	82	Yes	—
Catabolite gene activator protein	2	1	31	0.1	92	No	—
Glyceraldehyde-3-phosphate dehydrogenase	2	4	6	−0.2	78	Yes	EH
Glutamate dehydrogenase	2	7	6	−0.1	100	Yes	—
GroEL	3	2, 2	29, 90	0.7, 0.4	86, 34	Yes, No	DHB
Hexokinase	2	5	13	0.3	96	Yes	—
Lactoferrin	3	2, 2	55, 8	−0.8, −0.4	76, 14	Yes, No	DHB
Lysine/arginine/ornithine (LAO) binding protein	2	2	52	0.4	99	Yes	DHB
Maltodextrin binding protein	2	4	36	−0.5	98	Yes	DHA, DHB
Phosphofructokinase	2	6	5	−0.1	64	Yes	DHA
T4 lysozyme	2	2	47	−2.3	84	Yes	—
DNA polymerase β	3	3, 1	78, 22	−0.2, −1.0	95, 61	Yes, No	HL
TATA-binding protein	2	2	8	0.3	90	Yes	—
Tomato bushy stunt virus coat protein	2	1	20	−0.1	98	No	—
TRP repressor	3	1, 1	20, 29	0.4, 0.8	97, 82	No, No	DHA

[a]DHB: double-hinged β-sheet; DHA: double-hinged α-helix; HL: hinged loop; EH: α-helix as elastic hinge.

TABLE III. Distribution of Connecting Regions

No. of connecting regions	No. of occurrences
1	8
2	11
3	2
4	2
5	3
6	1
7	2

Figure 1a illustrates these cases schematically. The remaining three domain pairs (endothiapepsin, the interlobe motion of lactoferrin, one of the two domain pairs of GroEL) for which at most one of their two or more connecting regions are mechanical hinges, all have predominantly twist motions (closure motion <50%). These cases are illustrated in Figure 1b.

This result can be understood as follows. A twist axis is likely to lie parallel to any connecting region. If there are two well-separated connecting regions, at least one of them must be distant from the twist axis. For a closure motion, however, both separated connecting regions can be close to the axis, and indeed create this axis by acting as mechanical hinges, just as the two hinges of a door create the hinge axis about which the door opens and closes. It is perhaps for this reason that the most common number of connecting regions is two, the minimum number required to create a stable hinge axis that controls the domain closure. These examples are said to conform to the "door-closing model" of protein domain closure.

α-Helix and β-Sheet Termini Act as Hinges

By inspecting the 29 domain pairs visually, it appears that the termini of α-helices and β-sheets form connecting or bending regions in a large number of cases. To assess this, secondary structures were assigned using the program DSSP,[16] applied to conformers 1, given in Table I.

Table IV shows the distribution of the lengths of the bending regions. The average length is 5.3 residues, with a standard deviation of 3.9. The shortest bending regions comprised only 2 residues (the minimum number that can be assigned by DynDom), the longest 18. The most common length is 2, with 23 cases. Of these 23, 14 contained, or neighbored the terminal residue of either an α-helix or a β-sheet. To test the significance of this result, the probability was estimated of finding an α-helix or β-sheet terminal residue within, or neighboring, a segment of length 2, randomly selected from the whole set of 24 proteins. With

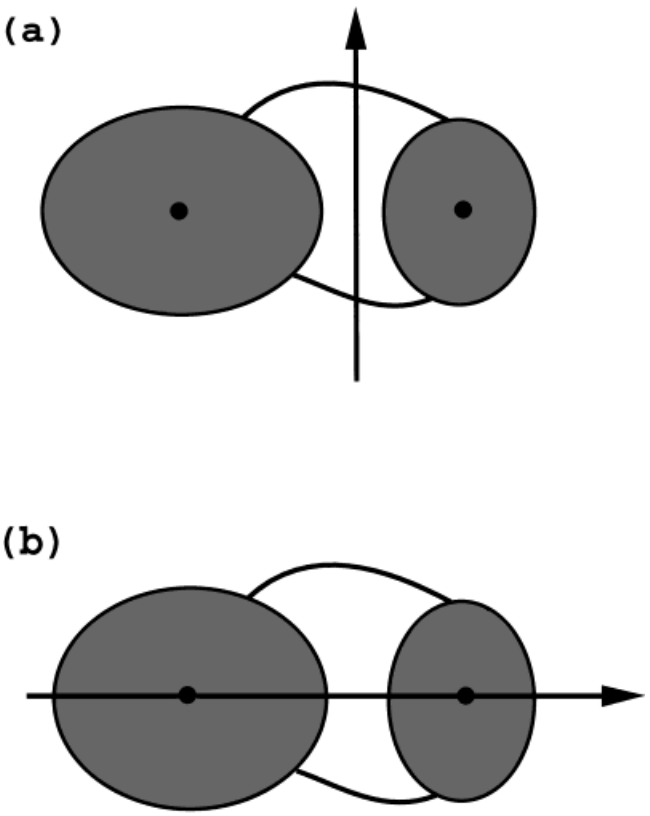

Fig. 1. **a**: Schematic representation of the ''door-closing model.'' In 18 of the 21 cases of domain pairs that have two or more connecting regions, at least two of the connecting regions are mechanical hinges. All 18 have predominantly closure motions. **b:** Schematic representation of the remaining three cases. Here, at most one of the connecting regions are mechanical hinges. All three cases have predominantly twist axes.

TABLE IV. Distribution of Connecting Region Lengths

Length of connecting region	No. of occurrences
2	23
3	12
4	5
5	12
6	5
7	8
8	1
9	2
10	2
11	2
12	2
13	0
14	0
15	3
16	0
17	0
18	2

this probability, the significance of this result was calculated using the binomial distribution. For segments with a length of two residues, it was found that to find 14 or more such cases of 23, by chance, is 4%. For longer-length regions, the results were not found to be significant. However, the smaller number of examples of segments of longer length, combined with the fact that it is more probable to find a terminus of a secondary structure element within or neighboring longer regions, indicates that a larger number of examples is required for this test to be properly made. However, the result for the connecting regions of two-residue length stands, indicating that at least in those cases where the bending region is short, the termini of the two main secondary structure elements are found. Of the 79 connecting regions, 59 contained or neighbored the terminal residue of either an α-helix or a β-sheet, and all 24 proteins apart from canine lymphoma immunoglobin had at least one connection that contained or neighbored the terminal residue of an α-helix or β-sheet.

No significant difference could be found between the frequency of occurrence of terminal residues of α-helix or β-sheet in connecting regions that passed near the interdomain screw axis (i.e., those that behave as mechanical hinges), and those more distant from the interdomain screw axis.

For a mechanical hinge in a protein, one would expect to see the polypeptide chain in a region relatively free of interactions, but in either direction enter a network of interactions that form the dynamic domains. A segment of the polypeptide chain forming a single strand of a β-sheet as part of one domain, and which then extends beyond the neighboring strand with which it is hydrogen bonded to eventually join another network of interactions forming another domain, fulfills this expectation. In the region where the chain is no longer able to form main-chain hydrogen bonds with the neighboring strand, dihedral angles will have more freedom, and this region with its greater flexibility will be able to act as the hinge. This explains why regions neighboring the termini of β-sheets are often found at interdomain regions. For an α-helix the situation is somewhat different. An α-helix is connected to a domain through interactions with its side chains. The emergence of an α-helix from this environment as it crosses between two domains is likely to be destabilizing, causing a termination of the α-helix and a consequent disruption of the ordered main-chain hydrogen bonding pattern. In this terminal region, one would again expect more flexibility.

In many cases, it is not only the post C-terminal or pre N-terminal region of the α-helix or β-sheet that is involved in the interdomain motion, but also a significant portion of the α-helix or β-sheet itself may be undergoing a conformational change. In these cases, the natural motions of the secondary structures themselves are playing a role in the interdomain motion. For example, it has been suggested[4] that hydrogen bonds between the two strands of a β-sheet can create a hinge axis. Although no such example of a β-sheet was found, in a later section two cases involving the elastic bending of an α-helix that does not involve any terminal region will be described.

Below, two examples are given of easily identifiable constructs that use the termini of β-sheets and α-helices to create axes.

TABLE V. Double-Hinged β-Sheets

Protein	Interdomain residue numbers	Strand residue numbers of double- hinged β-sheet
Alcohol dehydrogenase	292–293; 315–317	288–291; 314–316
cAMP protein kinase	104–105; 121–123	106–111; 115–120
Citrate synthase	56–57; 64–68	57–59; 63–65
LAO binding protein	89–93; 188–194	82–83; 198–199
GroEL	191–192; 373–377	186–190; 376–381
Lactoferrin	89–92; 249–253	91–98; 248–251
Maltodextrin binding protein	109–113; 257–261	106–111; 260–265

TABLE VI. Double-Hinged α-Helices

Protein	Interdomain residue numbers	Residue numbers of double-hinged α-helix
TRP repressor	65–66; 77–78	68–75
Phosphofructokinase	69–80; 93–94	79–91
Maltodextrin binding protein	312–314; 327–333	315–326

The Double-Hinged β-Sheet and α-Helix

Given that the termini of β-sheets can form mechanical hinges, it is perhaps not surprising that the two neighboring strands of a β-sheet can both form mechanical hinges that help create an effective hinge axis of specific location and orientation. This "double-hinged β-sheet" construction is seen particularly clearly in alcohol dehydrogenase, cAMP-dependent protein kinase, citrate synthase, LAO binding protein, maltodextrin binding protein, GroEL, and lactoferrin (Table V and Fig. 2). In LAO, lactoferrin, and GroEL, the double-hinged β-sheet provides the only connecting regions between the domains. In all cases, the interdomain screw axis passes within 7 Å of both of the bending regions, implying that they are acting more or less as mechanical hinges. In *all* cases, at least one of these connections extends from an outermost strand of the β-sheet.

In analogy to this result for β-sheets, the terminal regions at either end of a single α-helix could both form connecting regions, in a "double-hinged α-helix" construct. This construct has been found in maltodextrin binding protein, TRP repressor, and phosphofructokinase (Table VI). In maltodextrin binding protein and TRP repressor, both of the bending regions at either end of the helix form mechanical hinges. In phosphofructokinase, both termini are slightly more distant from the interdomain screw axis.

The Hinged Loop

This structure is created by a short loop (here we take loop to have its usual meaning in English, and do not imply that it is unstructured) that is connected covalently through the polypeptide backbone to one domain, but also has a region that has noncovalent interactions with the other domain. These noncovalent interactions cause this part of the loop to move with the other domain by way of two flexible regions of the backbone that form mechanical hinges. The hinge axis passes near both sides of the loop. The most beautiful example of this is seen in citrate synthase where the loop is a two-stranded antiparallel β-sheet, the β-hairpin region of which forms a hydrogen bond and salt bridges with the small domain.[7] This β-sheet is a double-hinged β-sheet at the end where the two strands join the large domain. The interdomain screw axis passes directly between these two flexible regions.

Although this particular construction was seen only in citrate synthase, other types of hinged loops have been found in aspartate aminotransferase, cAMP-dependent protein kinase, and DNA-polymerase β. In the case of aspartate aminotransferase, a short loop (residues 357–361) extending from one domain is connected through main-chain hydrogen bonds (Ile357– Gly197 in both 1AMA and 9AAT, and Gly358–Phe228 in 1AMA, and Gly358–Ala229 in 9AAT) to the other domain (Fig. 3). The loop is situated between a short α-helix and a short β-sheet. Most of the interdomain rotation is occurring in the ψ-angles of residues Ile357 and Cys361. The interdomain screw axis passes through the middle of this loop.

In cAMP-dependent protein kinase, the flexible part of the loop (residues 90–105) is also formed from the terminal regions of a β-sheet and α-helix . In fact, the loop connects one strand of the β-sheet to the α-helix, which lays over the β-sheet. The neighboring strand of the β-sheet also forms a connecting region. This β-sheet is the double-hinged β-sheet described above. The conformational flexibility is situated mainly in residues 90–98 and 104–105. Main-chain hydrogen bonds are found in both conformations between Val182 and Phe102 and Val104 and Val182. In addition, hydrophobic interactions appear to be maintained between residues 100–102 and Tyr306. Again the interdomain screw axis passes near both sides of the loop.

A further example is seen in DNA-polymerase β (residues 179–188). The tip of the loop, Arg182, is salt-bridged in both conformations to Glu316, causing it to rotate more with the domain of Glu316 than with the domain to which it is covalently linked. In addition, a main-chain hydrogen bond is maintained between Gly179 and Phe272 in both conformations. Most of the rotation occurs at the ψ-angle of Gly179 and at the ψ, ϕ angles of residues 182–184 on the other side of the tip. The axis passes though both sides of the loop.

Noncovalent Interactions Between Two Short Segments Can Form Hinge Axes

As mentioned above, for proteins endothiapepsin, lactoferrin (interlobe motion), catabolite gene activator protein, and tomato bushy stunt virus coat protein, did not have any mechanical hinges. The first two have two connecting regions, the latter two, just one. These cases are of interest because it is to be expected that in order to control their

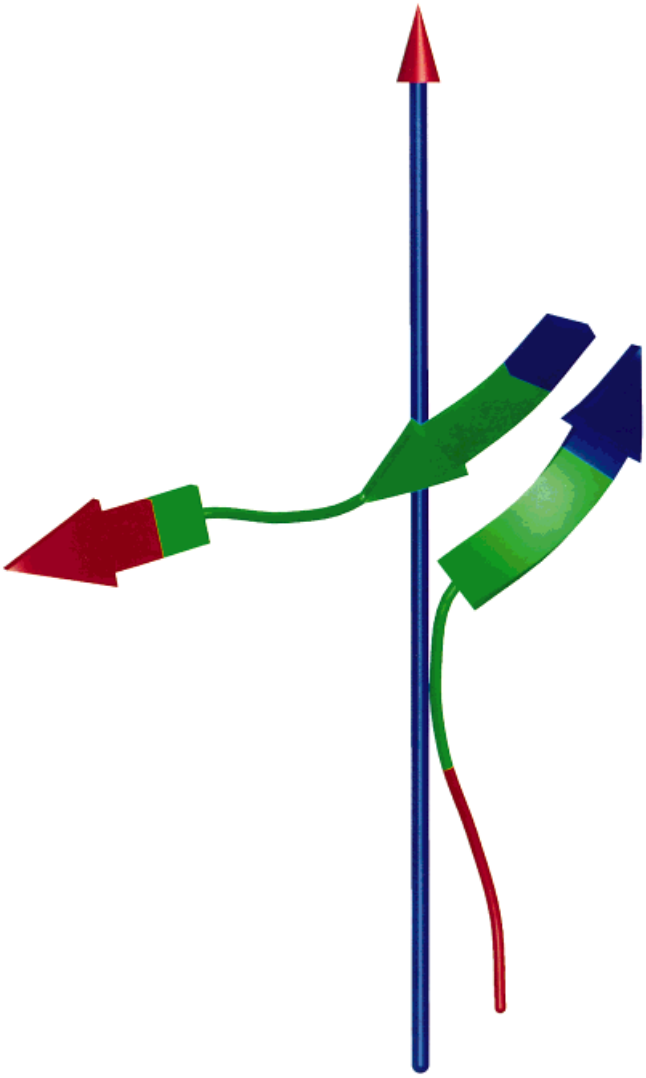

Fig. 2. Example of a double-hinged β-sheet. Shown are residues 86–94 and 248–255 from lactoferrin (PDB file: 1LFH). The blue region is part of one dynamic domain, the red, part of the other, and the green indicates the bending region. The arrow coloring indicates the direction of the rotation of the red domain relative to the blue domain by the right-hand rule, i.e., point the thumb of your right hand along the direction of the arrow and the natural curl of your fingers shows the direction of rotation of the red domain relative to the blue. The double-hinged β-sheet is shown in blue and green. For details see Table V. This kind of construct is also found in citrate synthase, alcohol dehydrogenase, cAMP-dependent protein kinase, LAO binding protein, maltodextrin binding protein, and GroEL. (This figure was created using "Bobscript"[21] and "Raster3D" [22]).

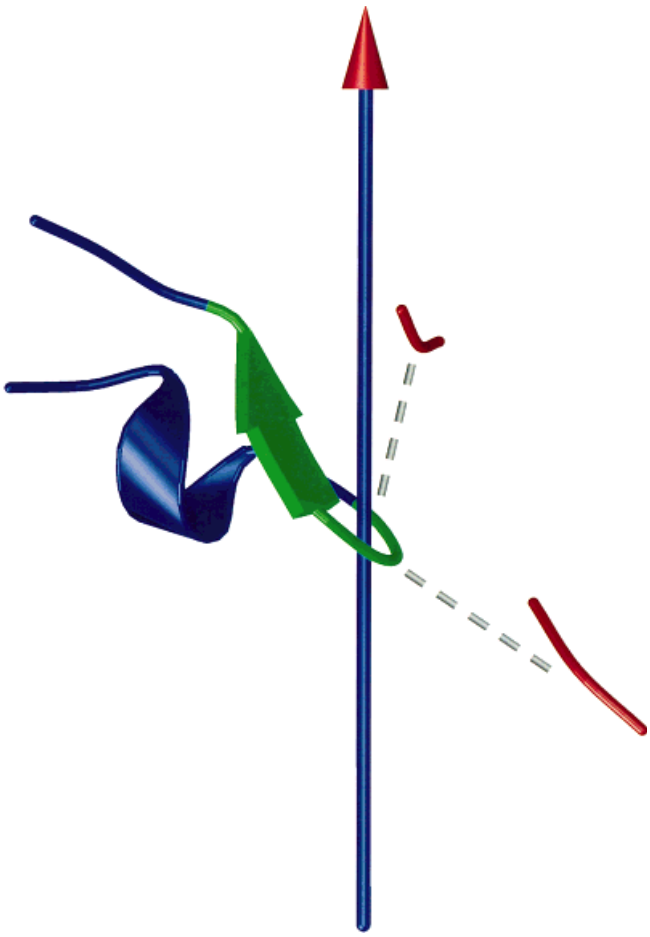

Fig. 3. An example of a hinged loop from aspartate-aminotransferase (PDB file: 1AMA). For interpretation of the colors, see the legend to figure 2. Shown as broken lines are the main-chain hydrogen bonds between Ile357 and Gly197, and Gly358 and Phe228. The effective hinge axis is seen to pass through the loop. Similar constructs are seen in citrate synthase, cAMP-dependent protein kinase, and DNA-polymerase β. (This figure was created using "Bobscript"[21] and "Raster3D"[22]).

interdomain motions, some kind of mechanical hinge exists.

In catabolite gene activator protein, DynDom has determined that the residues 137–138, located at the terminus of a long α-helix, are where the main bending between the two domains takes place. Residues 137–138 are more than 10 Å from the interdomain screw axis, and therefore this region is not a mechanical hinge according to the definition. Weber and Steitz[18] have found that the "hinge" region comprises residues 130–138. However, by looking at the rotation vectors it is clear that, although, residues 130–135 undergo conformational change because of the interdomain motion, they precede the region where the rotational transition takes place. In fact, most of the rotation between the domains takes place at the ψ-dihedral of Leu137, the ϕ- and ψ-dihedrals of Asp138, and the ϕ-dihedral of Val139. Physically, it is impossible for a single linkage to cause a body to rotate about an axis situated outside of the body itself. Therefore, another connecting region must act as the mechanical hinge. The interdomain screw axis passes between the β-strand of residues 59–65 in the large domain, and the α-helix of residues 169–175 in the small domain. Contacts between the residues 58–60 and Gln174, and Met59 and Ile175 are preserved in both conformers. The preserved hydrogen bond between the side chains of Glu58 and Gln174 appears to be particularly important.[18] Significantly, the axis is located directly between these two interacting segments, passing through the side chains of Met59 and Gln174 (Fig. 4). An analysis of the rotations

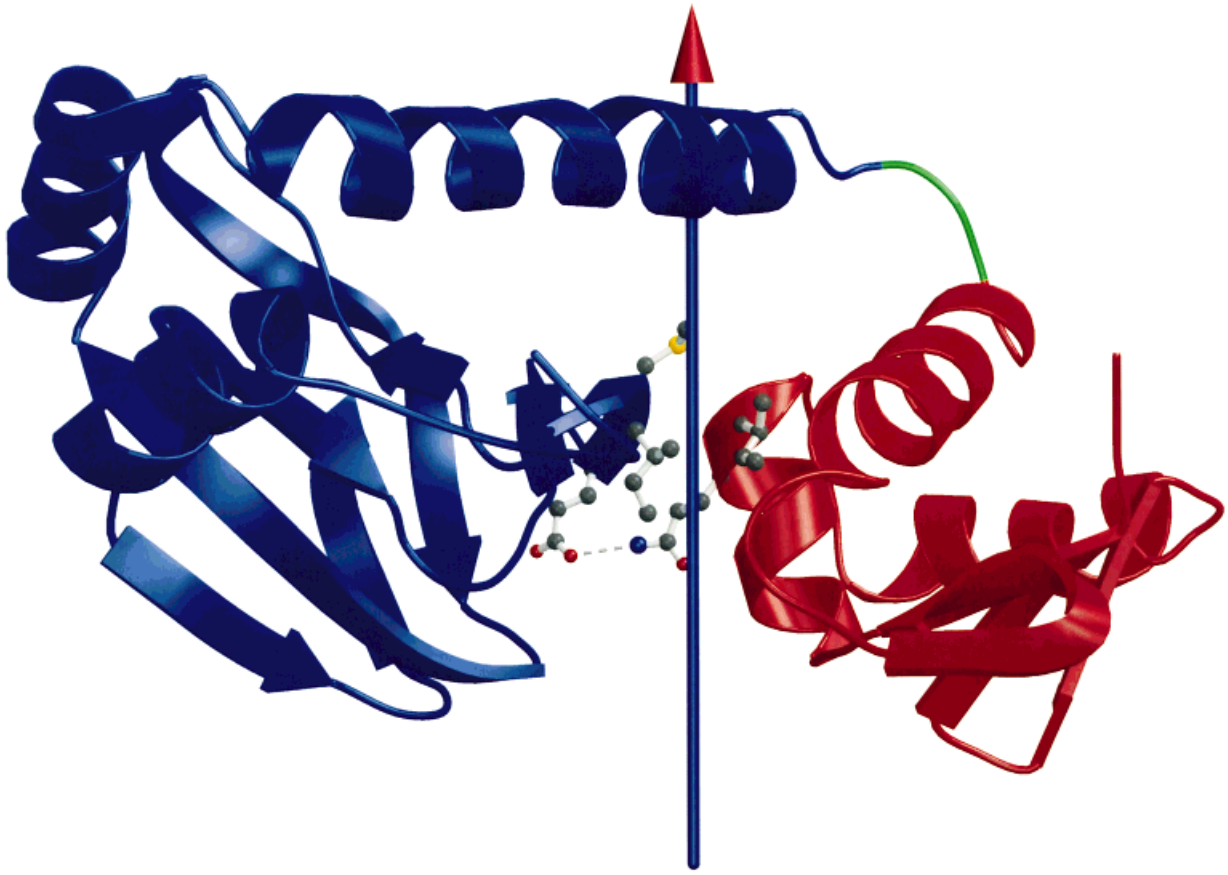

Fig. 4. An example of a hinge axis that is not determined by the connecting region through the backbone, colored green, but by the interactions between two segments remote along the chain. For interpretation of the colors, see the legend to figure 2. Shown is catabolite gene activator protein (PDB file: 3GAP, chain A). The side-chain atoms of residues 58–60 and 174–175 from the two segments are shown in ball and stick model. Contacts between 58–60 and 174, and 59 and 175, are preserved in both conformations. The preserved hydrogen bond between the side chain of Glu58 and Gln174, indicated by the broken line in the figure, appears to be particularly important. An analogous example is found in tomato bushy stunt virus coat protein. (This figure was created using "Bobscript"[21] and "Raster3D"[22]).

vectors of the backbone tetrahedra of these residues reveals that residues Gln174 and Ile175 clearly belong to the small domain, whereas residues Met59 and Ile60 clearly belong to the large domain. Glu58 is situated in the boundary region. This means that, unlike previous examples where the mechanical hinge is created at certain flexible main-chain dihedrals, in this case a hinge is created through the flexibility of noncovalent interactions between these residues, and the natural flexibility in their side-chain χ-dihedrals, which are seen to undergo some large rotations.

In tomato bushy stunt virus coat protein, one finds an analogous construction. There also, the bending region is more than 10 Å from the interdomain screw axis. It appears that interactions between residues Gly139, Thr140 in one domain, and residues Gly291, Thr292, and Arg307 in the other domain, together with an interaction between residue Ser150, and the interdomain residue Leu272, determine the location of the interdomain screw axis. The interdomain screw axis is seen to pass near the region created by these residues, again suggesting that a physical hinge is created through noncovalent interactions and side-chain dihedral flexibility.

Endothiapepsin and the interlobe motion of lactoferrin are the other two proteins that were found not to have any mechanical hinges. Both of these proteins have twist motions with their axes passing through regions that remain in contact during their small interdomain rotations. Therefore, it seems likely that these regions also form a physical hinge.

As pointed out previously,[7,14] the interdomain screw axis can be located anywhere in space, depending on the nature of the interdomain motion. If a real mechanical hinge, about which the rotation of one domain relative to the other occurs, is itself displaced, then the resulting interdomain screw axis will not pass through the mechanical hinge. Here, this means that the region of the backbone where the rotational transition occurs will not coincide with the location of the interdomain screw axis (Fig. 5c). However, if the interdomain screw axis passes through those regions where the rotational transition occurs, then the interdomain screw axis will coincide with the physical

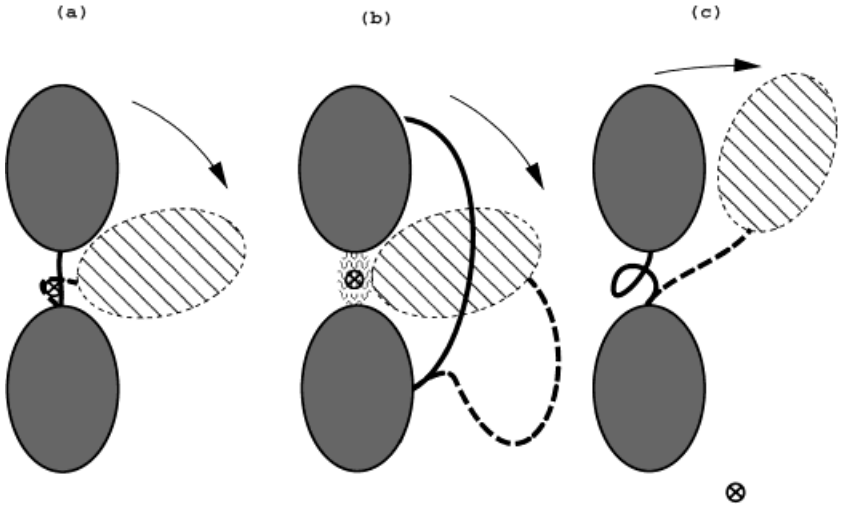

Fig. 5. **a**: Case where the interdomain screw axis coincides with a region of the backbone where the rotational transition occurs. This was found to be the case in 25 of the 29 domain pairs found. **b:** Case where the interdomain screw axis is distant from the region of the backbone where the rotational transition is found. Instead, the interdomain screw axis passes through a region where interactions between residues remote along the sequence are preserved. The interacting regions are indicated by the patterned region in the figure. This was found in four cases, namely, catabolite gene activator protein, tomato bushy stunt virus coat protein, endothiapepsin, and the interlobe motion of lactoferrin. **c:** Case where the interdomain screw axis is remote from any connecting region between the two dynamic domains where the rotational transition takes place. No case like this was found.

axis created by the real hinge. This was the motivation for the term "effective hinge axis" and "mechanical hinge." In 25 of 29 cases, the interdomain screw axis is located near at least one region of the backbone where the rotational transition takes place (Fig. 5a). In the four cases in which the interdomain screw axis is distant from the backbone region where the rotational transition takes place, it was found that the rotational transition takes place through segments from the two domains that preserve noncovalent interactions in both conformations, with the interdomain screw axis passing near these segments (Fig. 5b). No example of the case depicted in Figure 5c was found. Thus, in all 29 cases, the interdomain screw axis is a real physical axis. As a consequence, the definition of a "mechanical hinge" in the Methods section was too limited. It appears that mechanical hinges are not only created by flexible regions of the backbone, but also by noncovalent interactions between residues that are remote from each other along the backbone, and whose backbone dynamics assigns them to different dynamic domains.

It is concluded that in all the cases studied here, the domain motion is a rotation about a physical axis created through local interactions both covalent and noncovalent.

α-Helices as Elastic Hinges

In aspartate aminotransferase, the effective hinge axis passes near to a long interdomain α-helix, which undergoes some bending.[19] This α-helix runs from residues 313–343. The bending region was found to be at residues 323–329, i.e., the bending takes place mainly at these residues. The DSSP assignments of residues 316–340 show that they have perfect i to $i+4$ and i to $i-4$ main-chain hydrogen bonds in *both* conformers. In other words, despite the bending of the α-helix the main-chain hydrogen bonding pattern is preserved. In this sense, the α-helix is perfect in both of these conformations. Considering just the hydrogen bonding interaction, this deformation is elastic, because the higher energy conformation will, after the removal of the applied stress, relax to the lowest energy conformation. Looking at this α-helix in isolation (Fig. 6), the two conformations reveal a 12.5° rotation and −0.8 Å translation of the 330–340 portion relative to the 316–322 portion. In fact, the helix is straighter in the closed structure, suggesting that the bent α-helix is in a higher energy conformation in the open structure. Using DSSP, it is found that the main-chain hydrogen bonding energy of the bent α-helix is a substantial 14.4 kcal/mol higher than the straight α-helix. It is tempting to think, therefore, that binding of the substrate triggers the release of the stress that keeps this α-helix in a bent conformation, and the relaxation to the straight conformation drives the closing of the domains. In this model, the α-helix is acting just as a bent spring, storing elastic energy.

A similar case is found in glyceraldehyde-3-phosphate dehydrogenase.[20] Here, the C-terminal α-helix from residues 316–330 is also an interdomain α-helix. The bending region was found to be at residues Arg320 and Val321, and as for aspartate aminotransferase, the axis passes near this α-helix. The main-chain hydrogen bonding between the two conformations is preserved in residues 316–327. Both are almost perfect α-helices with i to $i+4$ and i to $i-4$ main-chain hydrogen bonding apart from the amide hydrogen of residue 317, which bonds to the carbonyl oxygen of residue 312 rather than 313 in both. Again looking at this α-helix in isolation, the two conformations reveal a 4° rotation and −0.02 Å translation of the 322–327 portion relative to the 316–319 portion. The main-chain hydrogen bonding energy for this α-helix in the closed conformation is 4 kcal/mol lower than for the open conformation.

Although no example was found, it is to be expected that β-sheets could also create elastic hinges.

CONCLUSIONS

An analysis has been presented of 24 domain proteins for which at least two X-ray conformers are known.

In all cases, the implied motion is controlled by at least one physical hinge created by covalent and nonconvalent interactions. No case was found where the interdomain screw axis is distant from all regions that preserve their

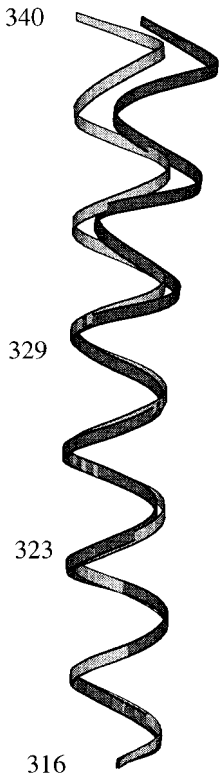

Fig. 6. The two conformations of the α-helix from residues 316–340 in aspartate aminotransferase (PDB files: 1AMA and 9AAT, chain A). The straight helix is from the closed structure, the bent from the open structure. All *i* to *i* +4 and *i* to *i* −4 main-chain hydrogen bonds are intact in both structures according to DSSP.[16] The 330–340 portion of the α-helix rotates 12° relative to the 316–322 portion, with most of the bending occurring in residues 323–329. According to DSSP, the hydrogen bonding energy of the bent helix is 14.4 kcal/mol higher than the straight structure. It is thought that this α-helix is acting as an elastic hinge. An analogous α-helix is found in glyceraldehyde-3-phosphate dehydrogenase. (This figure was created using "Bobscript"[21]).

interactions during the conformational change. In most cases, these bending regions involve the backbone, but in some cases, interactions between segments remote along the chain can create a hinge axis.

In a large majority of cases where at least two backbone connecting regions exist, at least two of these connecting regions are mechanical hinges. In all of these cases, the motion is a closure motion. It is suggested that a stable hinge axis is created by at least two separated mechanical hinges that provide precise control over the domain closure.

An analysis of the secondary structures of the bending regions revealed that at least for the shortest bending regions, the termini of the two main secondary structure elements are often found. This led to the identification of the "double-hinged β-sheet" and the "double-hinged α-helix" structures that appear to be used by domain proteins to create more than one bending region from a single structure. The former, which is the more frequent, arises when the two neighboring strands from a β-sheet situated in one domain diverge apart as they join the other domain. The two resulting termini then form the bending regions between the two domains. The double-hinged α-helix has two bending regions, one at either terminus. It appears to be rarer than the double-hinged β-sheet.

Short loops that connect one domain to another, and through which the effective hinge axis passes, are another identifiable construct that forms mechanical hinges. The noncovalent interactions that bind the loops have been identified in a number of cases.

Finally, two cases were found where the connecting region is the center of an interdomain helix that preserves its backbone hydrogen bonding pattern in both conformations. It is suggested that the bent conformations act as a store of elastic energy, which is used to drive the closing of the domains for rapid capture of the substrate.

Given that for the overwhelming number proteins whose structure has been solved by X-ray crystallography, only one conformer is available, other methods are required to determine the dynamic behavior of this majority. One method is clearly molecular dynamics simulation, and in the case of a small protein, there is some evidence of the success of this ab initio approach.[9] This study suggests the possibility of an alternative, bioinformatics approach: prediction, through the recognition of key structural motifs (i.e., the double-hinged β-sheet) that control the internal motions in proteins.

ACKNOWLEDGMENTS

I thank Professor Berendsen for supporting this work, his insightful comments, and indeed encouraging me to do this analysis in the first place. Also thanks to Bert de Groot and Frans van Hoesel for helpful discussions.

REFERENCES

1. Anderson CM, Zucker FH, Steitz T. Space-filling models of kinase clefts and conformational change. Science 1979;204:375–380.
2. Zhang X, Wozniak JA, Matthews BW. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. J Mol Biol 1995;250:527–552.
3. Faber HR, Matthews BW. A mutant T4 lysozyme displays five different crystal conformations. Nature 1990;348:263–266.
4. Gerstein M, Lesk AM, Chothia C. Structural mechanisms for domain movements in proteins. Biochemistry 1994;33:6739–6749.
5. Wriggers W, Schulten K. Protein domain movements: detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. Proteins 1997;29:1–14.
6. Lesk AM. Protein architecture. New York: Oxford University Press; 1991.
7. Hayward S, Berendsen HJC. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins 1998;30:144–154.
8. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins 1993;17:412–425.
9. de Groot BL, Hayward S, van Aalten DMF, Berendsen HJC. Domain motions in bacteriophage T4 lysozyme: a comparison between molecular dynamics and crystallographic data. Proteins 1998;31:116–127.
10. Harris LJ, Larson SB, Hasel KW, Day J, Greenwood A, McPherson A. The three-dimensional structure of an intact monoclonal antibody for canine lymphoma. Nature 1992;360:369–372.
11. Stillman TJ, Baker BJ, Britton KL, Rice DW. Conformational flexibility in glutamate dehydrogenase: role of water in substrate recognition and catalysis. J Mol Biol 1993;234:1131–1139.
12. Xu Z, Horwich AL, Sigler PB. The crystal structure of the

asymmetric GroEL-GroES-$(ADP)_7$ chaperonin complex. Nature 1997;388:741–750.
13. Huber R, Berendes R, Burger A, et al. Crystal and molecular structure of human annexin V after refinement: implications for structure, membrane binding and ion channel formation of the annexin family of proteins. J Mol Biol 1992;223:683–704.
14. Hayward S, Kitao A, Berendsen HJC. Model free methods to analyze domain motions in proteins from simulation. A comparison of a normal mode analysis and a molecular dynamics simulation of lysozyme. Proteins 1997;27:425–437.
15. Sayle R, Milner-White EJ. Rasmol: biomolecular graphics for all. Trends Biochem Sci 1995;20:374–375.
16. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
17. Vriend G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52–56.
18. Weber IT, Steitz TA. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 resolution. J Mol Biol 1987;198:311–326.
19. McPhalen CA, Vincent MG, Jansonius JN. X-ray structure refinement and comparison of three forms of mitochondrial aspartate-aminotransferase. J Mol Biol 1992;225:495–517.
20. Skarzynski T, Wonacott AJ. Coenzyme-induced conformational changes in glyceraldehyde-3-phosphate dehydrogenase from bacillus stearothermophilus. J Mol Biol 1988;203:1097–1118.
21. Esnoug RM. An extensively modified versiton of MOLSCRIPT that includes greatly enhanced coloring capabilities. J Mol Graph 1997;15:132–134.
22. Merrit EA, Murphy MEP. Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta Crystallogr D 1994;50:869–873.
23. Berry MB, Meador B, Bilderback T, Liang P, Glaser M, Phillips JGN. The closed conformation of a highly flexible protein: the structure of E. coli adenylate kinase with bound AMP and AMPNP. Proteins 1994;19:183–198.
24. Mueller CW, Schlauderer GJ, Reinstein J, Schulz GE. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. Structure 1996;4:147–156.
25. Eklund J, Samaha JP, Wallen L, Branden CI, Akeson A, Jones TA. Structure of a triclinic ternary complex of horse liver alcohol dehydrogenase at 2.9 A resolution. J Mol Biol 1981;146:561–587.
26. Plapp BV, Eklund H, Jones TA, Branden C. Three- dimensional structure of isonicotinimidylated liver alcohol dehydrogenase. J Biol Chem 1983;258:5537–5547.
27. McPhalen CA, Vincent MG, Picot D, Jansonius JN, Lesk AM, Chothia C. Domain closure in mitochondrial aspartate aminotransferase. J Mol Biol 1992;227:197–213.
28. Zheng J, Trafney EA, Knighton DR, Nguyen HX, Taylor SS, Ten Eyck LF, Sowadski JM. 2.2Å refined crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MnATP and a peptide inhibitor. Acta Crystallogr D 1993;49: 362–364.
29. Karlsson R, Zheng JH, Xuong NH, Taylor SS, Sowadski JM. Structure of the mammalian catalytic subunit of cAMP-dependent protein kinase and inhibitor peptide displays an open conformation. Acta Crystallogr D 1993;49:381–388.
30. Chattopadhyaya R, Meador WE, Means AR, Quiocho FA. Calmodulin structure refined at 1.7Å resolution. J Mol Biol 1992;226:1177–1192.
31. Meador WE, Means AR, Quiocho FA. Target enzyme recognition by calmodulin: 2.4Å structure of a calmodulin-peptide complex. Science 1992;257:1251–1255.
32. Remington S, Wiegand G, Huber R. Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 angstroms resolution. J Mol Biol 1982;158:111–152.
33. Bennett MR, Choe S, Eisenberg D. Refined structure of dimeric diptheria toxin at 2.0Å resolution. Prot Sci 1994;3:1444–1463.
34. Bennett MR, Eisenberg D. Refined structure of monomeric diptheria toxin at 2.3Å resolution. Prot Sci 1994;3:1464–1475.
35. Sali A, Veerapandian B, Cooper JB, Foundling SI, Hoover DJ, Blundell TL. High-resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: analysis of the inhibitor binding and description of the rigid-body shift in the enzyme. EMBO J 1989;8:2179–2188.
36. Pearl L, Blundell T. The active site of aspartic proteinases. FEBS Lett 1984;174:96–101.
37. Lamzin VS, Dauter Z, Popov VO, Harutyunyan EH, Wilson KS. High resolution structures of holo and apo formate dehydrogenase. J Mol Biol 1994;236:759–785.
38. Skarzynski T, Moody PCE, Wonacott AJ. Structure of hologlyceraldehyde-3-phosphate dehydrogenase from bacillus stearothermophilus at 1.8Å resolution. J Mol Biol 1987;193:171–187.
39. Yip KSP, Stillman TJ, Britton KL, et al. The structure of pyrococcus furiosus glutamate dehydrogenase reveals a key role for ion pair networks in maintaining enzyme stability at extreme temperatures. Structure 1995;3:1147–1158.
40. Anderson CM, Stenkamp RE, Steitz TA. Sequencing a protein by X-ray crystallography. II. Refinement of yeast hexokinase B coordinates and sequence at 2.1Å resolution. J Mol Biol 1978;123: 15–33.
41. Benett WS Jr, Steitz TA. Glucose induced conformational change in yeast hexokinase. Proc Natl Acad Sci USA 1978;75:4848–4852.
42. Anderson BF, Baker HM, Norris GE, Rumball SV, Baker EN. Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins. Nature 1990;344:784–787.
43. Oh B-H, Pandit J, Kang C-H, Nikaido K, Gokcen S, Ames GF-L, Kim S-H. Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. J Biol Chem 1993;268:11348–11355.
44. Sharff AJ, Rodseth LE, Spurlino JC, Quiocho FA. Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. Biochemistry 1992;31:10657–10663.
45. Spurlino JC, Lu G-Y, Quiocho FA. The 23Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. J Biol Chem 1991;266: 5202–5219.
46. Shirakihara Y, Evans PR. Crystal structure of the complex of phosphofructokinase from *Escherichia coli* with its reaction products. J Mol Biol 1988;204:973–994.
47. Pelletier H, Sawaya MR, Kumar A, Wilson SH, Kraut J. Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer, and ddCTP. Science 1994;264:1891–1903.
48. Sawaya MR, Pelletier H, Kumar A, Wilson SH, Kraut J. Crystal structure of rat DNA polymerase beta: evidence for a common polymerase mechanism. Science 1994;264:1930–1935.
49. Chasman DI, Flaherty KM, Sharp PA, Kornberg RD. Crystal structure of yeast TATA-binding protein and model for interaction with DNA. Proc Natl Acad Sci USA 1993;90:8174–8178.
50. Olson AJ, Bricogne G, Harrison SC. Structure of tomato bushy stunt virus: the virus particle at 2.9 A resolution. J Mol Biol 1983;171:61–93.
51. Lawson CL, Zhang R, Schevitz RW, Otwinowski Z, Joachimiak A, Sigler PB. Flexibility of the DNA-binding domains of the trp repressor. Proteins 1988;3:18–31.