

What is the relationship between the global structures of apo and holo proteins?

Michal Brylinski and Jeffrey Skolnick*

Center for the Study of Systems Biology, School of Biology, Georgia Institute of Technology, Atlanta, Georgia 30318

ABSTRACT

It is well known that ligand binding and release may induce a wide range of structural changes in a receptor protein, varying from small movements of loops or side chains in the binding pocket to large-scale domain hinge-bending and shear motions or even partial unfolding that facilitates the capture and release of a ligand. An interesting question is what in general are the conformational changes triggered by ligand binding? The aim of this work is to analyze the magnitude of structural changes in a protein resulting from ligand binding to assess if the state of ligand binding needs to be included in template-based protein structure prediction algorithms. To address this issue, a nonredundant dataset of 521 paired protein structures in the ligand-free and ligand-bound form was created and used to estimate the degree of both local and global structure similarity between the apo and holo forms. In most cases, the proteins undergo relatively small conformational rearrangements of their tertiary structure upon ligand binding/release (most root-mean-square-deviations from native, RMSD, are $<1 \text{ \AA}$). However, a clear difference was observed between single- and multiple-domain proteins. For the latter, RMSD changes greater than 1 \AA and sometimes larger were found for almost 1/3 of the cases; these are mainly associated with large-scale hinge-bending movements of entire domains. The changes in the mutual orientation of individual domains in multiple-domain proteins upon ligand binding were investigated using a mechanistic model based on mass-weighted principal axes as well as interface buried surface calculations. Some preferences toward the anticipated mechanism of protein domain movements are predictable based on the examination of just the ligand-free structural form. These results have applications to protein structure prediction, particularly in the context of protein domain assembly, if additional information concerning ligand binding is exploited.

Proteins 2008; 70:363–377.
© 2007 Wiley-Liss, Inc.

Key words: holoproteins; apoproteins; ligand binding; mass-weighted principal axes; pheromone-binding protein; HPPK; guanylate kinase; EPSP synthase; clamp loader gamma subunit; renin.

INTRODUCTION

It is well documented that ligand binding and release may induce a wide range of structural changes in a receptor protein, which are essential for biological activity.^{1–4} These changes in protein conformation mainly refer to small movements of loops^{5,6} or side chains in the binding pocket^{7,8} but also originate from large-scale domain motions.^{9,10} It has been shown that protein domains tend to move as rigid bodies in response to interactions with substrates and products of enzymatic reactions.^{11,12} Moreover, isolated domains often retain their main structural and functional features.^{13,14} The structural mechanisms, by which conformational rearrangements assist relative domain motions, have been rationalized in terms of two fundamental movements: hinge-bending and shear, depending on whether or not a sliding over a continuously maintained and tightly packed interface occurs.^{15,16} Furthermore, partial unfolding was also suggested as a possible mechanism to facilitate the capture and release of a ligand.^{17,18} On the other hand, it has been suggested that in most cases the changes in backbone structure are negligible and only side chain reorientations occur upon ligand binding.⁸ It remains somewhat unclear to what extent ligand binding can be considered as a purely local structural phenomenon.

The huge number of protein structures solved by experimental methods and available in the Protein Data Bank, PDB,¹⁹ provides an opportunity to analyze the distribution of various protein properties. It is now common that alternative experimentally solved structures of the same protein are available in the PDB. Exhaustive studies focused on analyzing and systematizing the instances of protein structures solved in multiple conformations resulted in comprehensive databases of macromolecular motions, intended to be of use in studying structure–function relationships.^{20–24} Many of the alternative solved structures correspond to ligand-free, apo-proteins and proteins forming different complexes or binding to different

Grant sponsor: Division of General Medical Sciences, National Institute of Health; Grant number: GM-48835.

*Correspondence to: Jeffrey Skolnick, Center for the Study of Systems Biology, Georgia Institute of Technology, 250, 14th Street NW, Atlanta, GA 30318.

E-mail: skolnick@gatech.edu

Received 30 October 2006; Revised 15 February 2007; Accepted 9 March 2007

Published online 6 August 2007 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.21510

ligands, the holo form. However, the great majority of the investigations of ligand-bound/ligand-free structural differences focused on side chain flexibility and local changes in a protein structure upon ligand binding to improve docking algorithms.^{7,8,25} The aim of this work is to analyze the changes in protein secondary and tertiary structure as well as in mutual orientation of protein domains resulting from ligand binding to assess whether the state of ligand binding needs to be considered if the goal is the prediction of the global fold of a protein. To address this issue, a representative dataset of ligand-bound/ligand-free protein pairs was prepared and used to estimate the degree of both local and global structural differences between these two forms.

It is of great importance to analyze whether significant structural rearrangements of a protein structure result from protein–ligand interactions in the context of its applicability to the improvement of protein structure prediction approaches, especially for multiple-domain proteins. Although there has been progress in the field of protein tertiary structure prediction over the past decade,^{26–28} a common weakness of many approaches is the relatively low success rate for predicting the mutual orientation of protein domains.^{29–31} Since predicted structures of individual domains frequently have the correct topology, the improvement of the ability to predict domain orientations may yield a substantial increase of the prediction accuracy of multiple-domain proteins. One concrete way to make progress on this problem would be to exploit the dependence of domain orientations on the ligand binding state of the protein.

MATERIALS AND METHODS

Ligand-bound/ligand-free dataset creation

The flowchart followed for the creation of the dataset is presented in Figure 1. Protein structures determined by X-ray crystallography to a resolution ≤ 2.5 Å and that have at least 50 residues were selected from the Protein Data Bank.¹⁹ In this study, we focused on the changes in a receptor structure triggered by the binding of small molecules and short peptides, therefore protein–DNA as well as protein–RNA complexes were excluded from the dataset. The structures were divided into two groups: ligand-bound and ligand-free forms based on the following criteria: The ligand-bound form must contain at least one ligand molecule. The minimum number of six ligand non-hydrogen atoms was assumed to exclude ions, water, and very small molecules. Short peptides composed of up to 15 standard or modified amino acids as well as single nucleotides were also considered as ligands. Moreover, to exclude nonspecific ligand interactions, a minimum number of six residues in contact with the ligand atoms were imposed. The interatomic contacts were calculated by LPC.^{32,33} Using this criterion, each ligand

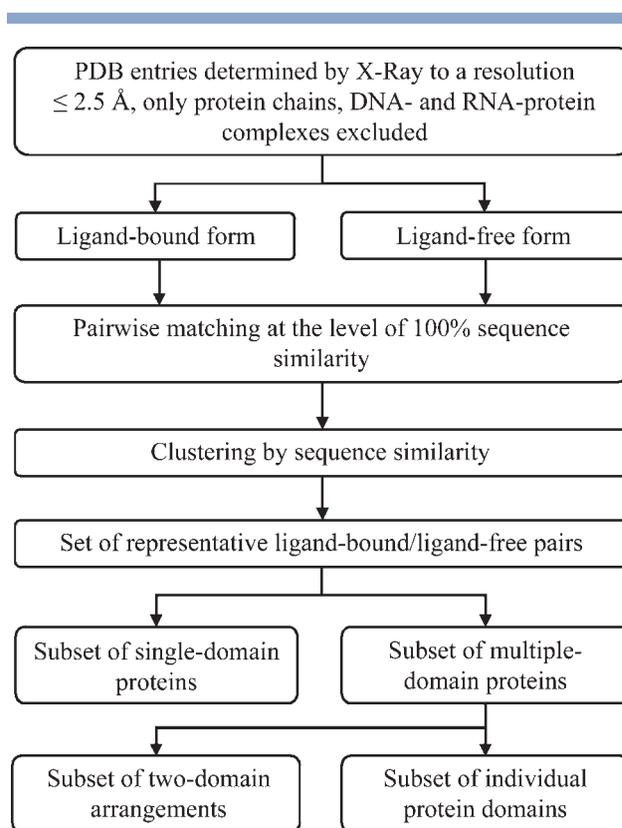


Figure 1

Flowchart of dataset creation. Details are given in the text.

non-hydrogen atom has six protein heavy atom neighbors (within a 4.5-Å cutoff) on average. A protein structure that did not meet the aforementioned criteria was considered as a ligand-free form. Those proteins with little or no secondary structure have been removed. Moreover, to avoid unresolved fragments of a structure and misnumbering of residues, the amino acid sequences used in this study were determined for the fragments of structures, for which at least backbone coordinates were available.

Subsequently, ligand-bound and ligand-free proteins were paired at the level of 100% sequence identity and all the possible pairs were subjected to a clustering procedure, using a cutoff of 35% sequence identity between clusters to remove redundancy. The most representative pair from each cluster (closest to the cluster centroid) was selected to create the nonredundant set of ligand-bound/ligand-free pairs used in this study. If more than one holo structure was found for a given sequence, the ligand-bound form was selected randomly. Furthermore, those proteins were separately clustered by sequence similarity, and structure comparisons were done for the cluster centroids to investigate changes in a protein structure across different ligand-bound forms. Subsequently, the

complete set was categorized into the subset of single- and multiple-domain proteins by the protein domain parser PDP.³⁴ The multidomain proteins were further split by PDP into the subset of individual protein domains and the subset of two-domain arrangements. A two-domain arrangement requires that two domains are in contact if their interface buried surface in both structural forms (ligand-bound and ligand-free) was larger than 200 Å². If a protein was composed of more than two domains, all possible two-domain combinations were investigated.

Global and local structure comparisons

Global structure similarity was measured by the RMSD of the C α atoms. The RMSD values were calculated for the superimposed entire structures of ligand-bound and ligand-free form of a protein with the exception of RMSD reported for individual protein domains, which were superimposed independently. The local structure similarity expresses the fraction of residues having the same secondary structure in the ligand-bound and ligand-free forms as reported by DSSP.³⁵ All seven types of ordered local structure recognized by DSSP were taken into account to assess local structure similarity: H (α -helix), B (residue in isolated β -bridge), E (extended strand), G (3/10 helix), I (π -helix), T (hydrogen bonded turn), and S (bend). The remaining residues were considered to be in the random coil conformation.

Large-scale motions of protein domains

An intrinsic molecular property of a protein domain can be described by its principal axes from the inertia tensor matrix. To investigate large-scale domain movements in the subset of two-domain arrangements extracted from multiple-domain proteins, we used a similar approach to that previously used to characterize the quaternary and tertiary differences as observed in glycogen phosphorylase with different effectors bound.³⁶ In this method, the mass-weighted principal axes are created to provide a quantitative description of differences of domain mutual orientations within various protein structures. The essential details are given in the Appendix. The degree of bending was defined as the difference between angles formed by the corresponding principal axes of the individual domains in two conformers (ligand-free and ligand-bound). To simplify the comparisons across the great variety of proteins and domain arrangements, a single angular movement corresponding to the largest observed change was used to describe the degree of bending. Additionally, the displacement of equivalent domains is also reflected by the RMSD upon the global superposition of the two conformations of a protein. The comparison of the degree of bending and RMSD provides useful characteristics of the domain

motion. The underlying assumption is that the hinge-bending motion of protein domains results in a considerable degree of bending as well as high RMSD between two conformers, whereas shear movements can be characterized by a detectable RMSD change and simultaneously, by relatively small angular movement of the domains' principal axes.

The vast majority of individual protein domains tend to behave as rigid bodies with very small internal distortion; however, for calculating the degree of bending by the principal axes approach, those two-domain arrangements containing domains that undergo significant internal distortion were excluded from the analysis. Moreover, to reduce the impact of an intradomain distortion on the orientation of principal axes, corresponding domains in ligand-free and ligand-bound structural forms were averaged. Finally, we evaluated the ability of this model to classify the motions of protein domains into simple categories (shear, hinge, other) by comparison with the detailed description of selected proteins provided by the Database of the Macromolecular Movements^{20,21} (<http://molmovdb.org/MolMovDB>).

Interdomain surface area analysis

The analysis of interdomain surface area in two-domain arrangements was performed to investigate the amounts of interface surface that become buried upon ligand binding. Moreover, the potential predictability of large-scale domain movements and preferences toward the anticipated type of motion were evaluated using the interface buried surface area of a ligand-free structural form as a classifier. The surface calculations were done using the Jackal Protein Structure Modeling Package.³⁷

RESULTS

Ligand-bound/ligand-free dataset

The examination of protein structures deposited in the PDB (Oct 2006) resulted in 21,743 protein chains that met the criteria for a ligand-bound form given in "Materials and methods section". Simultaneously, 13,043 were classified as ligand-free structures. A pairwise matching procedure at the level of 100% sequence identity resulted in a redundant set of 25,344 ligand-bound/ligand-free pairs. For 806 nonredundant protein chains, more than one ligand-bound structure was detected. After clustering all the possible pairs with a 35% sequence identity cutoff, the resulting set of 521 representative ligand-bound/ligand-free pairs with lengths varying from 50 to 1392 residues was compiled. This results in 328 single-domain proteins, 193 multiple-domain proteins, 487 individual protein domains, and 311 two-domain arrangements. The list of proteins and their associated ligands may be

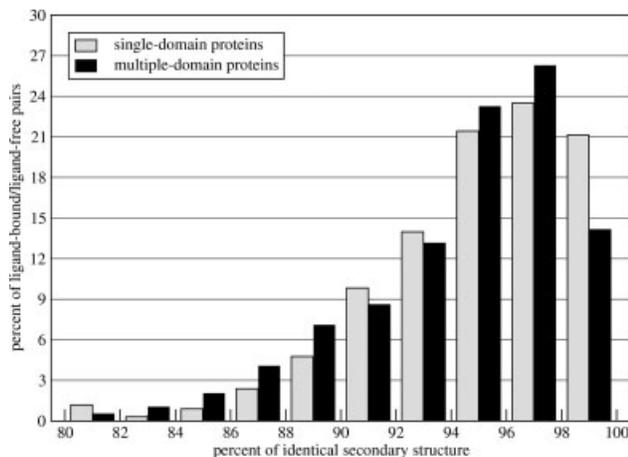


Figure 2

Distribution of the secondary structure similarity between the ligand-bound and ligand-free form for the subset of single- and multiple-domain proteins.

found at: <http://cssb.biology.gatech.edu/skolnick/files/LigandBinding/>.

Local structure similarity

The results of the analysis of secondary structure similarity between ligand-bound and ligand-free form are presented in Figure 2. Only minor differences in ordered local structure between these forms were detected. Moreover, the average secondary structure similarity between ligand free and ligand bound forms for single- and multiple-domain proteins was found to be comparable: 94.8 and 94.4%, respectively. The only perceptible difference between single- and multiple-domain proteins was detected in the highest as well as lowest similarity range. Apparently, in the case of single-domain proteins, over 21% of pairs have identical or almost identical (98–100%) secondary structure, whereas a significantly lower fraction of multiple-domain proteins (14%) was observed in this interval. The slightly larger alteration of secondary structure observed for multiple-domain proteins corresponds to structural changes induced by ligand binding/release in the hinge regions. On the other hand, the comparative analysis of protein structures solved in multiple ligand-bound conformations revealed that the average secondary structure similarity across the holo forms of single- and multiple-domain proteins is 95.9 and 95.8%, respectively. Most frequently, the differences in DSSP assignments were observed in the termini of secondary structure elements or in flexible fragments of the structure. Hence, it may be concluded that ligand binding/release induces rather small alteration of protein secondary structure, slightly larger in multiple- than single-domain proteins.

Global structure similarity

The RMSD after global rigid-body superposition is a widely used measure to express the global structure similarity between proteins. Figure 3 shows the distribution of RMSD between ligand-bound and ligand-free forms for the single- and multiple-domain proteins as well as the individual protein domains present in multiple-domain proteins. The average RMSD value was found to be 0.74, 1.08, and 0.49 Å for the subset of single-, multiple-domain proteins and individual domains that are part of multiple-domain proteins, respectively. It is clearly seen that a RMSD < 1 Å is most frequently observed, regardless of the subset used. However, a lower fraction of multiple-domain proteins are found in this range than for single-domain proteins as well as the individual protein domains participating in multiple-domain proteins. Consequently, more ligand-bound/ligand-free pairs are characterized by higher RMSD if the protein is composed of more than one domain. For these proteins, a RMSD > 1 Å and sometimes larger was found for almost 1/3 of the cases, which is clearly associated with large-scale movements of entire domains. Simultaneously, the vast majority of individual protein domains in multiple-domain proteins have a very low RMSD; thus the individual domains are more insensitive than single domain proteins to the ligand binding state.

The global similarity of protein structures solved in multiple ligand-bound conformations was also investigated. The results are shown in Figure 4. The average pairwise RMSD across the holo forms of single- and multiple-domain proteins was found to be 0.49 and 0.52 Å, respectively. The differences between the global structures of apo and holo proteins are more prominent than those observed across various ligand-bound conformations. The dataset can be found at <http://cssb.biology.gatech.edu/skolnick/files/LigandBinding/>.

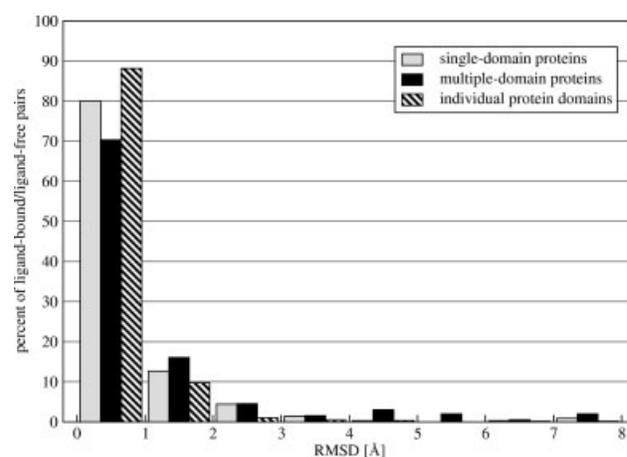


Figure 3

Distribution of the C α RMSD calculated for the subset of single-, multiple-domain proteins and individual protein domains in multiple-domain proteins.

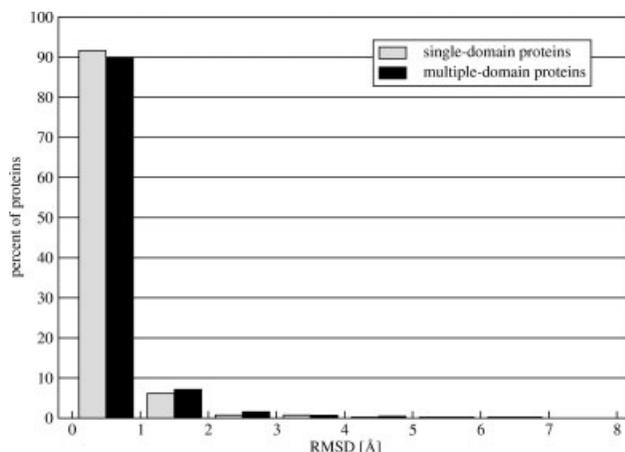


Figure 4

Distribution of the average pairwise RMSD across the holo forms of single- and multiple-domain proteins.

Large-scale domain movements

In multiple-domain proteins, the relative disposition of one domain to another can be described by the angular values, which describe the rotational motion vectors.

The motions of protein domains were investigated for the subset of two-domain arrangements extracted from multiple-domain proteins. Two parameters were used to describe the change in the mutual orientation of domains in ligand-free and ligand-bound forms: the degree of bending (defined as the largest difference in the angle formed by the principal axis of individual domains) and the domain shift (detected by the RMSD change upon global superposition). The results of domain motion analysis using those parameters are presented in Figure 5. It turned out that conformational changes triggered by ligand binding/release that involve relatively small angular movements ($<5^\circ$) correspond to a RMSD change $<1 \text{ \AA}$ in most cases. Nevertheless, large-scale domain movements are detected in a significant number of proteins (Table I). Most of the proteins that undergo large-scale domain motion have been analyzed and the detailed description of the motion is available. Based on the distinctive dependence of the scale of domain shift on the degree of bending, it can be inferred whether a particular conformational change involves shear (RMSD > 1 and $< 2 \text{ \AA}$) or hinge (RMSD $> 2 \text{ \AA}$, bending $> 10^\circ$) motion, or other movement. We found that shear-type mechanism is the most frequent type of domain motion in the dataset (53%); hinge-bending is common for 45%, and the remaining cases cannot be unambiguously classified.

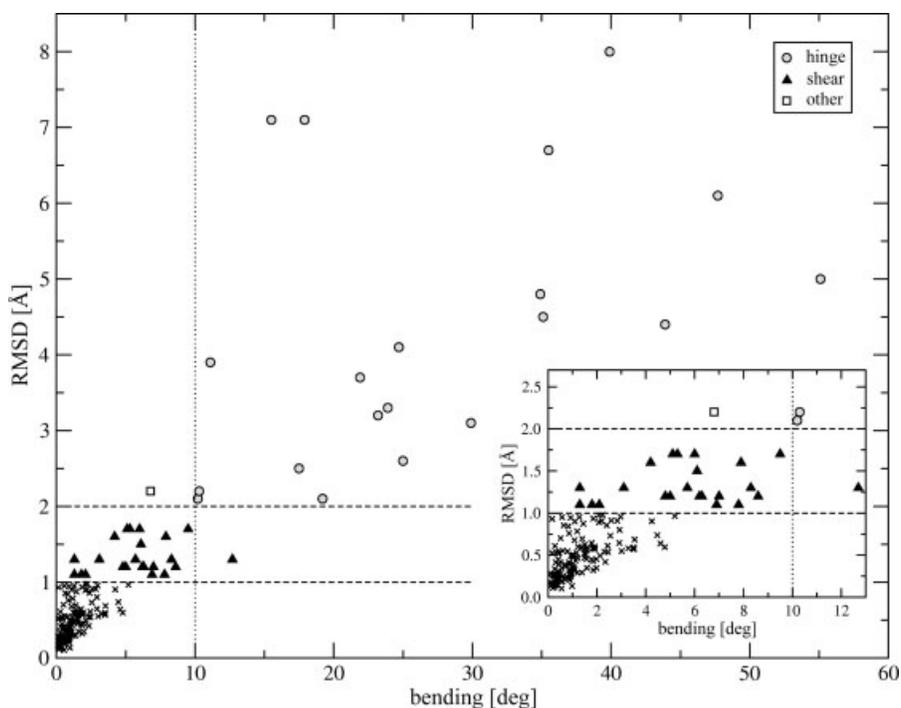


Figure 5

Domain motions upon ligand binding described by the degree of bending and the RMSD between ligand-free and ligand-bound structural forms.

Table 1
Statistics for Proteins Involved in Large-Scale Domain Motions upon Ligand Binding

Protein	PDB code, chain		Two-domain arrangement ^a	Interface buried surface area [Å ²]		Large-scale movement		
	Apo	Holo		Apo	Holo	Bending [deg]	RMSD [Å]	Type ^b
DAHPS ^c	1vr6A	1rzmA	D1 (1–64) D2 (65–338)	288.0	1024.7	39.9	8.0	[D-h-2]
L-leucine-binding protein	1usgA	1usiA	D1 (1–124, 247–333) D2 (125–246, 334–345)	1336.1	2291.9	17.9	7.1	[D-h-2]
Adenylate kinase ^{c,d}	4akeA	1akeA	D1 (1–121, 160–214) D2 (122–159)	733.3	1614.8	15.5	7.1	[D-h-2]
Lactoferrin ^{c,d}	1cb6A	1bka_	D2 (91–250, 329–341) D3 (342–434, 595–691)	1342.4	681.3	35.5	6.7	[D-h-2]
			D1 (4–90, 251–328) D2 (91–250, 329–341)	1417.3	2246.0	47.7	6.1	[D-h-2]
			D1 (4–90, 251–328) D3 (342–434, 595–691)	1313.4	1298.9	5.7	1.3	[D-s-2]
Osmoprotection protein (ProX) ^c	1sw5A	1sw2A	D1 (6–110, 21–275) D2 (111–210)	1687.5	2190.2	55.1	5.0	[D-h-2]
Alginate-binding protein (AlgQ1) ^c	1y3qA	1y3nA	D1 (1–135, 308–399) D2 (136–307, 400–490)	2084.6	2759.8	34.9	4.8	[D-h-2]
D-allose binding protein ^c	1gudA	1rpjA	D1 (1–112, 247–288) D2 (113–246)	1285.0	2132.2	35.1	4.5	[D-h-2]
Guanylate kinase ^{c,d}	1ex6B	1ex7A	D1 (1–32, 84–186) D2 (33–83)	880.5	1202.9	43.9	4.4	[D-h-2]
Uroporphyrinogen III synthase ^c	1wd7B	1wcvA	D1 (8–40, 166–261) D2 (41–165)	1408.3	945.2	24.7	4.1	[D-h-2]
Bovine mitochondrial F1-ATPase ^c	1w0jE	1e1rF	D1 (1–73) D2 (74–466)	1532.5	1667.9	11.1	3.9	[D-h-2]
EPSP synthase ^c	1rf5A	1rf4A	D1 (1–19, 230–427) D2 (20–229)	1336.1	2449.1	21.9	3.7	[D-h-2]
Sucrose phosphatase (SPP) ^{c,d}	1s2oA	1tj5A	D1 (1–83, 162–244) D2 (84–161)	1207.2	1522.5	23.9	3.3	[D-h-2]
Phosphate reductoisomerase (DXR) ^c	1k5hA	1q0qA	D1 (1–147) D2 (323–398)	895.6	943.1	23.2	3.2	[D-h-2]
β-Phosphoglucomutase ^c	1zolA	1o03A	D1 (1–16, 84–221) D2 (17–83)	1052.5	1708.1	29.9	3.1	[D-h-2]
Dephospho-CoA kinase ^c	1viyC	1vhlA	D1 (0–32, 96–207) D2 (33–95)	820.9	783.7	25.0	2.6	[D-h-2]
Diaminopimelate epimerase	1gqzA	2gkeA	D1 (1–116, 261–274) D2 (117–260)	1794.4	2062.5	17.5	2.5	[D-h-2]
Aspartate transcarbamylase (ATCase) ^c	1za1A	1q95A	D1 (1–136, 292–310) D2 (137–291)	2322.4	2530.2	10.3	2.2	[D-h-2]
Protein disulfide bond isomerase (DSBC) ^c	1tjdA	1eejB	D1 (1–66) D2 (67–216)	997.5	696.4	19.2	2.1	[D-h-2]
β-Glucosyltransferase ^{c,d}	1jejA	1jg6A	D1 (1–172, 335–351) D2 (173–334)	2079.8	2550.6	10.2	2.1	[D-h-2]
Adenylosuccinate synthetase ^{c,d}	1hooB	1cg0A	D1 (1–100, 202–431) D2 (101–201)	2131.2	2683.4	6.8	2.2	[D-?-2]
Saccharopine reductase ^c	1e5IA	1e5qA	D2 (130–258, 340–397, 437–450) D3 (259–339)	2271.8	2595.5	9.5	1.7	[D-s-2]
			D1 (2–129, 398–436) D2 (130–258, 340–397, 437–450)	2852.3	3197.6	7.9	1.6	[D-s-2]
Fatty acid responsive transcription factor ^c	1hw1B	1h9gA	D1 (5–82) D2 (83–227)	872.8	1043.8	6.0	1.7	[D-s-2]
Aspartyl-tRNA synthetase ^{c,d}	1l0wB	1g51A	D3 (142–240, 526–549) D4 (241–294, 414–525) D5 (295–413)	3528.7	3629.7	5.3	1.7	[D-s-2]
			D4 (241–294, 414–525) D2 (106–141, 550–562)	1339.9	1226.5	3.1	1.3	[D-s-2]
			D3 (142–240, 526–549) D1 (4–129, 235–282) D2 (130–234)	1531.9	1569.8	1.3	1.3	[D-s-2]
α-Ketoglutarate-dependent taurine dioxygenase ^c	1otjD	1gy9A	D1 (4–129, 235–282) D2 (130–234)	3068.5	3473.7	5.1	1.7	[D-s-2]

(Continued)

Table I
Continued

Protein	PDB code, chain		Two-domain arrangement ^a	Interface buried surface area [Å ²]		Large-scale movement		
	Apo	Holo		Apo	Holo	Bending [deg]	RMSD [Å]	Type ^b
Threonyl-tRNA synthetase ^{c,d}	1evkA	1evlA	D1 (242–531) D2 (532–642)	1628.8	1735.0	4.2	1.6	[D-s-2]
Prion protein URE2 ^c	1g6wD	1k0bC	D2 (264–298) D3 (189–263, 299–341)	1058.5	1066.0	6.1	1.5	[D-s-2]
<i>E. coli</i> clamp loader γ subunit ^d	1njgB	1njfA	D1 (5–177) D2 (178–243)	1520.5	1470.3	12.7	1.3	[D-s-2]
Phenazine-biosynthesis protein (PhzF) ^c	1t6kA	1u1wA	D1 (1–117, 265–278) D2 (118–264)	1550.1	2028.1	8.3	1.3	[D-s-2]
Dihydrodipicolinate reductase (RV2773C) ^c	1yl5B	1yl7A	D1 (1–106, 211–245) D2 (107–210)	1279.1	1253.3	8.6	1.2	[D-s-2]
L-Rhamnulose kinase	2cgkB	2cgjA	D2 (76–237) D3 (238–300, 375–480)	2496.6	2761.4	7.8	1.1	[D-s-2]
<i>E. coli</i> ribokinase ^c	1rkaA	1gqtA	D1 (2–75) D3 (238–300, 375–480)	1557.9	1871.7	7.0	1.2	[D-s-2]
Shikimate 5-dehydrogenase (AroE) ^c	1wxdB	2cy0A	D1 (4–163, 242–308) D2 (164–241)	2278.7	2314.5	6.3	1.2	[D-s-2]
SARS protease ^{c,d}	2a5aA	1uk4A	D1 (1–103, 231–262) D2 (104–230)	1886.0	2118.3	6.2	1.2	[D-s-2]
Renin	2g26A	2fs4B	D1 (3–197) D2 (198–303)	1628.7	1601.5	5.0	1.2	[D-s-2]
DNA polymerase III, β subunit ^c	1mmiA	1ok7B	D1 (3–15, 142–333) D2 (16–141)	2757.4	2706.8	4.8	1.2	[D-s-2]
Tetanus toxin C fragment	1a8d_	1d0hA	D2 (122–247) D3 (248–366)	1405.8	1352.4	6.9	1.1	[D-s-2]
Dihydrofolate reductase	1pdbA	1kmsA	D1 (875–1110) D2 (1111–1315)	2763.4	2884.2	2.1	1.1	[D-s-2]
L-asparaginase	1hfkA	1hg1C	D1 (3–34, 115–186) D2 (35–114)	1951.8	1808.0	1.8	1.1	[D-s-2]
Cytosine deaminase ^c	1k6wA	1k70A	D1 (4–216) D2 (217–327)	2151.9	2140.4	1.8	1.1	[D-s-2]
			D1 (4–55, 367–410) D2 (56–366)	2210.9	2224.0	1.3	1.1	[D-s-2]

^aMutual orientation of two domains in a multiple-domain protein that has been investigated. The start and end residues of sequence segments constituting individual domains as reported by PDP are given in parenthesis.

^bClassification of macromolecular motions proposed by Gerstein and coworkers.^{15,21}

^cDomain motions analyzed by DynDom^{23,24} (<http://www.cmp.uea.ac.uk/dyndom/>).

^dProtein motions that have been classified in MolMovDB²⁰ (<http://molmovdb.mbb.yale.edu/molmovdb/>).

Change in the interdomain surface area upon ligand binding

The interdomain buried surface area was calculated for each two-domain arrangements found in multiple-domain proteins in the ligand-free as well as in ligand-bound state. The results are given in Table I and illustrated in Figure 6. Domains bury varying amounts of interface surface that usually increases upon ligand-induced hinge-bending motions, while the interface buried surface area remains more or less constant upon shear-type movements. The average ratio of the interdomain surface area observed in holo versus apo structural form was found to be 1.39 and 1.05 for proteins that undergo hinge and shear motions, respectively.

Interdomain surface area as a motion classifier

Using the subset of two-domain arrangements, we have evaluated the usability of the interface buried surface area analysis to predict the likelihood of large-scale domain motions. We note that in this analysis only the ligand-free structural forms are taken into consideration. The results are presented in Figure 7. Most of the large-scale movements were observed for the interface buried surface area spanning the range from 500 to 3000 Å². The largest average bending and average RMSD was found for the relatively small interface surface (between 1000 and 1500 Å²). Subsequently, for moderate interdomain surface area (1500–2500 Å²) both main types of

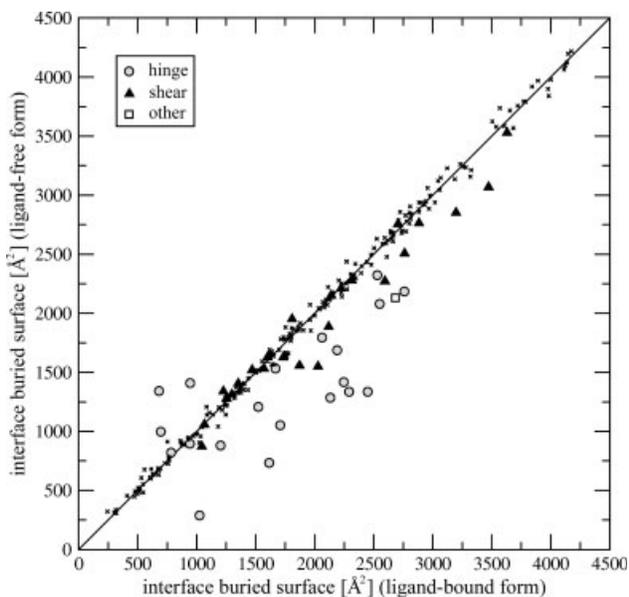


Figure 6

The change in the interdomain buried surface of two-domain arrangements upon ligand binding.

motions were observed in significant number of the cases. More extensive interactions between domains (2500–3500 Å²) account for the total preclusion of hinge-bending movements; however shear movements still occur. Finally, no large-scale movements were observed if the interdomain surface area was greater than 3550 Å². Those results suggest that some preferences toward the anticipate type of motion are deducible by the examination of a ligand-free structural form alone.

Examples

To illustrate significant conformational changes in single- as well as in multiple-domain proteins triggered by ligand binding, selected cases are discussed below.

Pheromone-binding protein

Pheromone-mediated intercommunication plays an important role in insects' sexual attraction.³⁸ Females of the silkworm moth (*Bombyx mori*) release a volatile 16-carbon alcohol, bombykol, as a sex pheromone that is recognized at long distances by the male mates.³⁹ The transport of hydrophobic pheromones through the sensillary lymph surrounding dendritic membrane receptors is assisted by pheromone-binding proteins.⁴⁰ The pheromone-binding protein from the silkworm moth (BmorPBP) is a single-domain, highly soluble protein, which consists of a tightly packed arrangement of six α -helices tightly held together by three disulfide bridges.

The X-ray structure of BmorPBP bound to bombykol (PDB ID: 1dqe)⁴¹ revealed that the ligand binds in a completely enclosed hydrophobic cavity formed by four antiparallel helices. BmorPBP was found to exhibit a significant pH-dependent conformational rearrangement of its tertiary structure upon bombykol binding/release. The N-terminal α -helix observed in the holo form is replaced by an extended conformation in the apo form of BmorPBP [Fig. 8(A)]. Moreover, in the bombykol-bound form, the C-terminus of the protein appears as a disordered tail, whereas in the unbound structure (PDB ID: 2fjy)⁴² its C-terminal region forms a helix that fills the binding pocket of the protein [Fig. 8(B)]. The C-terminal region is composed of mostly non-polar amino acids as would expected for it to be housed in a cavity designed to bind hydrophobic pheromones. Detailed studies on the C-terminal coil→helix transformation suggested that an occupied binding pocket is likely more energetically favorable than an open, empty cavity, making it more favorable for BmorPBP to bind either a ligand or the

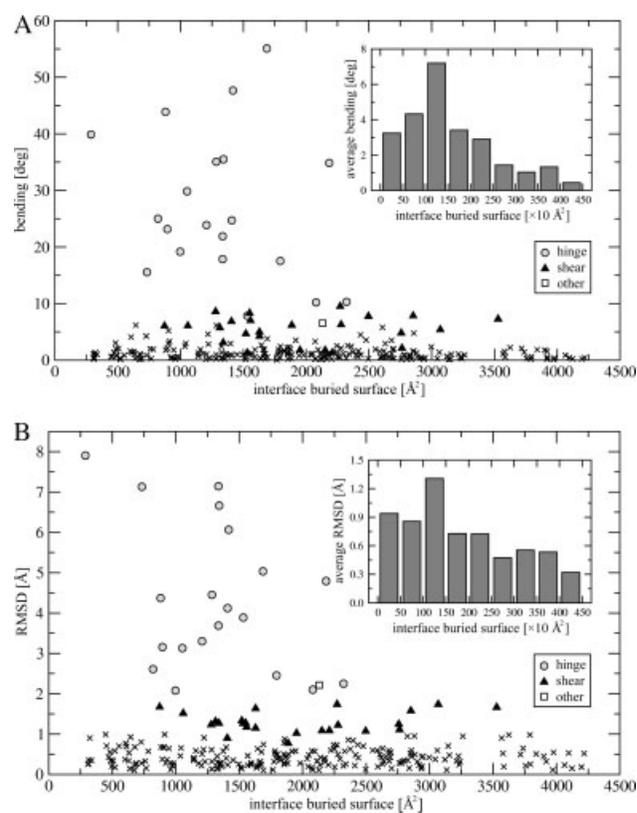


Figure 7

Domain motions associated with the interface buried surface area calculated for ligand-free forms of multiple-domain proteins. The degree of bending (A) as well as the RMSD (B) calculated for two-domain arrangements upon ligand binding are plotted versus the interdomain surface area. The average bending and RMSD observed in particular bins of the interface buried surface are depicted in inset plots.

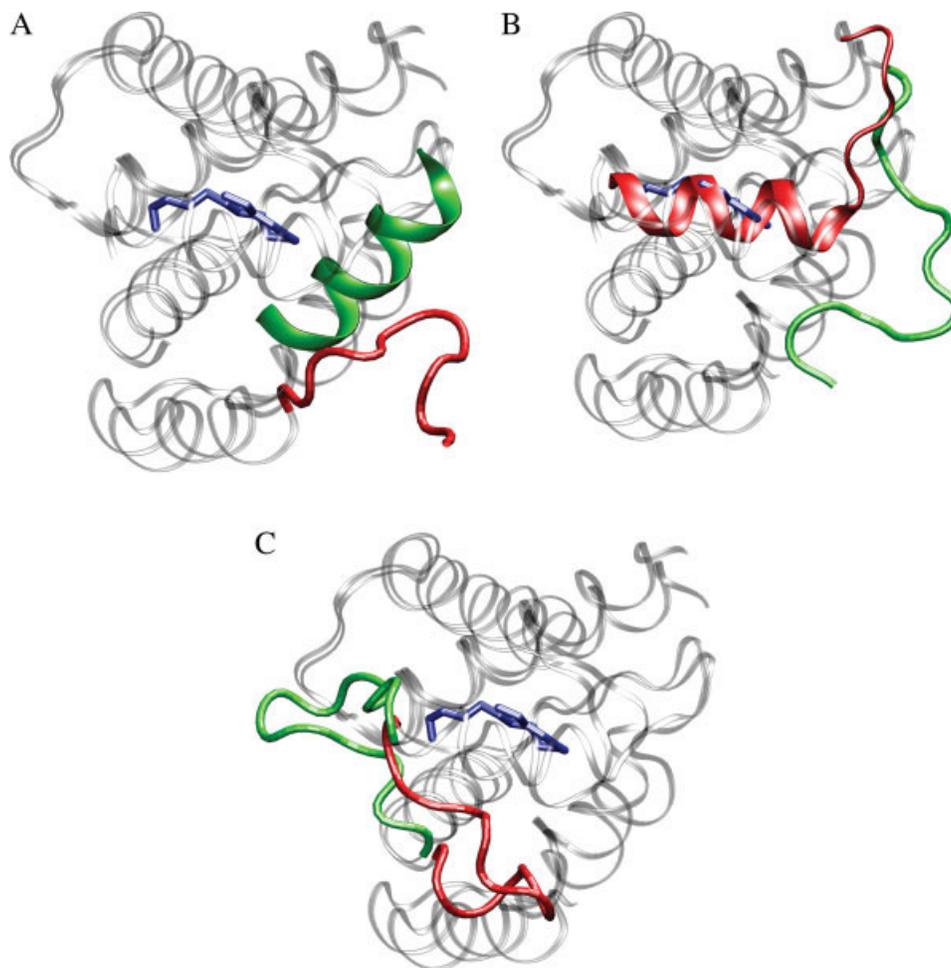


Figure 8

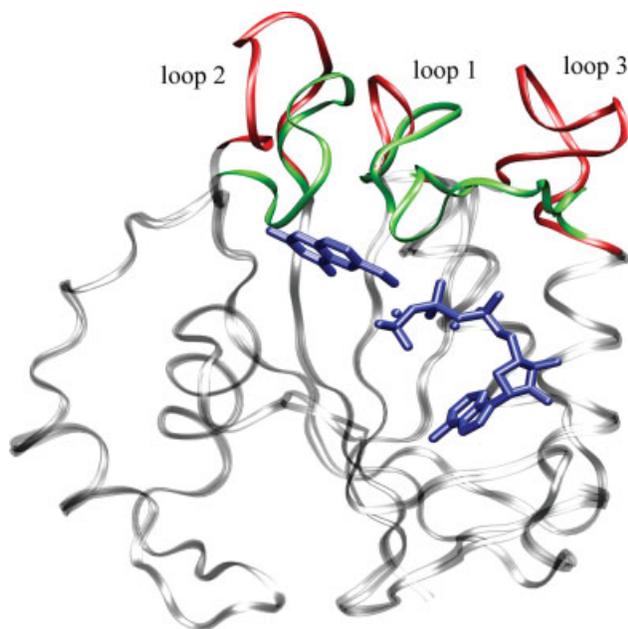
Comparison of bombykol-bound (green) and unbound (red) structure of pheromone-binding protein from the silkworm moth. The corresponding segments of the rigid helical scaffold that undergo a small rearrangement upon bombykol binding/release are presented as gray ribbons. Bound bombykol, enclosed in the core of the protein, is shown in blue. The extensive conformational changes induced by alcohol release include unwinding of the N-terminal helix (A), the insertion of a helix formed by the C-terminal into the space occupied by the ligand in the complex (B), and the displacement of a loop covering the binding site (C). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

C-terminal α -helix in its pocket.⁴² Additionally, significant changes are observed in the loop region covering the binding site [Fig. 8(C)]. This loop may move aside to allow entry of the ligand or the protein may partially unfold to allow the ligand to occupy the binding pocket.⁴¹ The structure of the apo BmorPBP is different from the structure of the alcohol-bound with an RMSD value of 7.0 Å. In the corresponding segments of the rigid helical scaffold, which is significantly less alterable, the backbone atoms can be superimposed with a RMSD < 2.0 Å.

HPPK

The first step in the folate biosynthesis pathway is the reaction of pyrophosphate transfer from ATP to 6-hy-

droxymethyl-7,8-dihydropterin (HP), catalyzed by the enzyme 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK).⁴³ The enzymes of this pathway are attractive targets for antimicrobial drug design, as the folate pathway is absent in mammals that can acquire folate cofactors from diet, but is essential for microorganisms that must synthesize folates themselves.⁴⁴ HPPK is a small, 158 residue monomeric protein for which experimentally determined structures in various liganded states have been reported. We compared the apo-form (PDB ID: 1g4c)⁴⁵ to HPPK complexed with bisubstrate (PDB ID: 1f9h).⁴⁵ Structural and biochemical studies revealed that three flexible surface loops (Fig. 9) are responsible for substrate binding and the catalytic action.^{46,47} The systematic analysis of different HPPK conformers has shown the dramatic ligand-induced conformational

**Figure 9**

Induced fit in HPPK. The structures were superimposed with respect to all but surface loop residues. Three flexible surface loops involved in ligand binding and catalysis are shown in red and green, for apo and holo form, respectively. The remaining parts of HPPK are gray. Both substrates and two Mg^{2+} ions required for catalysis (blue) occupy the catalytic center that assembles upon binding. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

changes of these catalytic loops, particularly loop 3.^{48,49} The global RMSD calculated for apo versus holo form was found to be 3.0 Å, while the truncated structure that has all three loops removed can be superimposed within a 0.5 Å C α RMSD.

Guanylate kinase

Guanylate kinase (GK) catalyzes the phosphorylation of GMP to GDP and is an important enzyme in nucleotide metabolic pathways.⁵⁰ It provides a particularly clear example of the general features of hinge-bending domain motion. The open conformation (ligand-free, PDB ID: 1ex6)⁵¹ and the closed conformation (ligand-bound, PDB ID: 1ex7)⁵¹ are presented in Figure 10(A). Like other NMP kinases, GK consists of three dynamic domains that encompass the group of concurrently moving residues: CORE, LID and NMP-binding domain. The CORE domain is composed of five-stranded, parallel β -sheet and six α -helices, the NMP-binding domain contains a four-stranded β -sheet and one α -helix and the LID domain consists of a single, seven residue long loop that links two CORE α -helices. We note that the PDP algorithm partitioned GK into two structural domains: CORE+LID and NMP-binding domain [Fig. 10(A)],

because the LID domain is too small to be recognized by automated domain parsing.⁵² Structural studies revealed that GMP binding induces a dramatic rigid-body domain displacement of the whole NMP-binding domain and the GMP-binding site was found to be completely assembled only upon binding of GMP.⁵¹ The shift of the LID domain is much smaller. The change in the relative orientation of CORE+LID and NMP-binding domains as described by the displacement of ellipsoids representing individual structural domains is presented in Figure 10(A) (right panel). Systematic analysis of the domain movements using the DynDom program (data available from <http://www.cmp.uea.ac.uk/dyndom/>) indicated that the large domain movements in GK upon binding of ligands involve the rotation around an effective hinge axis, so GK opens and closes its domains with a rotation angle of 47.0°. The degree of bending was calculated to be 43.9° using the principal axes analysis; the relative rotation of the domains results in the increase of the interdomain surface area by a factor of 1.37. Simultaneously, the RMSD between “open” and “closed” form of GK is 4.4 Å.

EPSP synthase

5-Enolpyruvylshikimate-3-phosphate (EPSP) synthase catalyzes a reversible transfer of an enolpyruvyl group from phospho(enol)pyruvate (PEP) to shikimate 3-phosphate (S3P) to form 5-enolpyruvylshikimate-3-phosphate, an intermediate in the shikimate pathway leading to aromatic amino acid biosynthesis.⁵³ The enzymes of this pathway are desirable targets for the development of new antibiotics, antiparasitics, and herbicides, since the shikimate pathway occurs in higher plants, fungi, algae, bacteria and apicomplexan parasites, but not in mammals. EPSP synthase has been crystallized in the unliganded state (PDB ID: 1rf5) as well as in two different ligand-bound states: complexed with S3P/GLP (PDB ID: 1rf6) and in the tetrahedral intermediate state with S3P and a pseudosubstrate F-PEP (PDB ID: 1rf4).⁵⁴ EPSP synthase has two distinct globular domains of similar molecular weight connected by two short linkage fragments [Fig. 10(B), left panel]. The domain motion observed in EPSP synthase is brought about by the rotation of one domain over the other as depicted by the relative rotation of ellipsoids representing individual domains in ligand-bound and ligand-free form in Figure 10(B) (right panel). As the result of this motion, the ligand molecule is captured in the active site. The movement of the domains was described by DynDom as rotation around an effective hinge axis of 26.7°.⁵⁴ The largest change in the angles of domains principal axes, was calculated to be 21.9°, resulting in the global RMSD of 3.7 Å. Interdomain interactions in the ligand-bound conformation are distinctly stronger when compared with that observed in the unliganded state, accounting for the large increase of

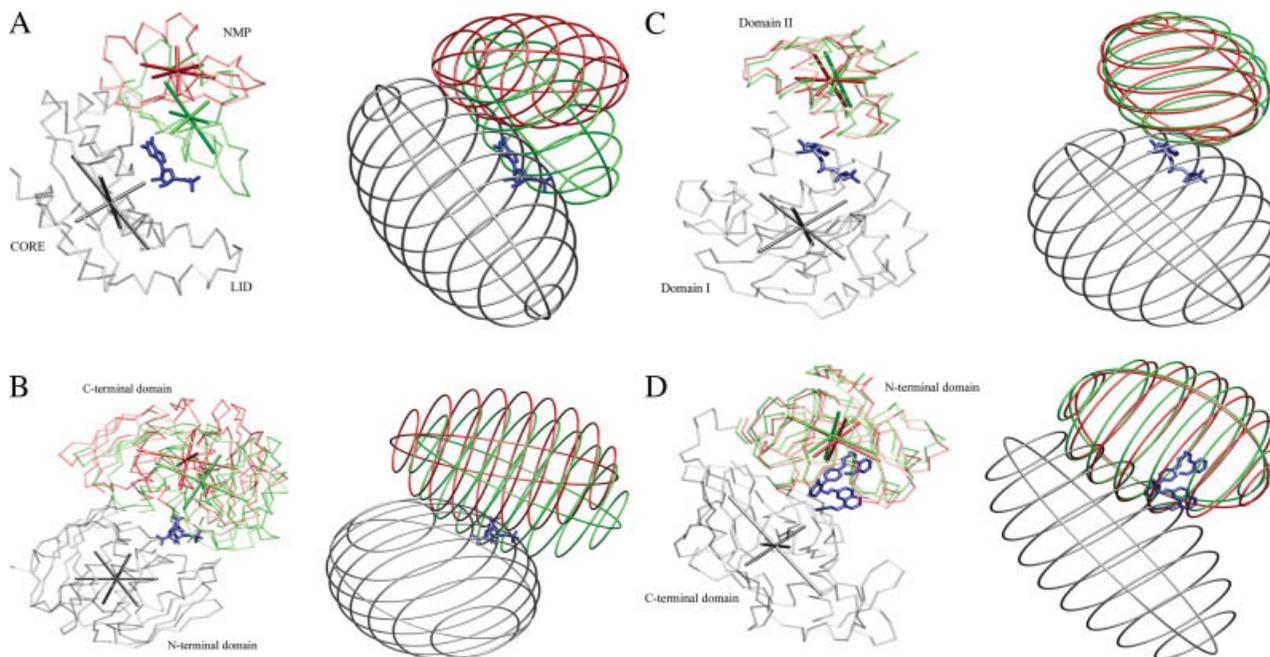


Figure 10

Rigid-body domain movements upon ligand binding described by a mechanistic model based on mass-weighted principal axes. Left panel: C α trace of one domain in ligand-free (red) and ligand-bound (green, blue) form upon superposition of the other domain (grey). The corresponding domains in apo and holo structures are averaged and the calculated mass-weighted principal axes are shown. Right panel: domain displacement depicted by the change in the relative position of ellipsoids representing the shape of each domain in the apo and holo structural form (using the same orientation and color notation as in left panel). Proteins that undergo hinge-bending domain motion: (A) guanylate kinase; the structures were superimposed using CORE+LID structural domain (gray) and (B) EPSP synthase; the superposition is on the N-terminal domain (gray). Proteins that undergo shear-type domain movements: (C) *E. coli* clamp loader γ subunit; the superposition is on domain I (gray) and (D) renin; the structures were superimposed using C-terminal domain (gray).

the buried interface surface upon ligand binding from 1336.1 to 2449.1 Å². We note that binding of different ligands results in the global RMSD between two holo forms of EPSP synthase of 0.18 Å.

***E. coli* clamp loader γ subunit**

Sliding clamps and their associated clamp loader (CL) provide essential assistance to DNA polymerase activity, accounting for the remarkable speed of DNA replication.^{55,56} CLs are DNA-dependent ATPase complexes that consist of five subunits. The γ subunit binds ATP and powers the action cycle of the CL. Each subunit is composed of three domains (I–III), where domains I and II constitute the nucleotide-binding site as shown in Figure 10(C). The crystal structure of a truncated form (domains I + II) of the isolated γ -ATPase subunit of the *E. coli* CL was solved in both nucleotide-free (PDB ID: 1njg) and nucleotide bound (PDB ID: 1njf) form.⁵⁷ Structural studies revealed that nucleotide binding to the CL γ subunit results in a rotation of domain I with respect to domain II by 10°. ⁵⁷ The degree of bending calculated using principal axes approach was found to be 12.7°. The domain shift observed in CL γ subunit

[Fig. 10(C), right panel] corresponds to a global RMSD of 1.3 Å and results in a slight decrease of the interface buried surface from 1520.5 to 1470.3 Å². The results available from the Database of Macromolecular Movements²⁰ (<http://molmovdb.mbb.yale.edu/molmovdb/>) show the predominantly shear-type motion of the two domains of CL γ subunit upon the nucleotide binding.

Renin

The renin-angiotensin endocrine system is involved in regulating cardiovascular and renal function and in maintaining the fluid electrolyte balance of the body.⁵⁸ Renin is a member of the aspartic proteinase family that catalyses the cleavage of angiotensinogen to release the decapeptide angiotensin I. The prominent specificity of renin against angiotensinogen accounts for its high attractiveness as a target for antihypertensive drugs. In this study, we used a renin structure bound to a ketopiperazine-based inhibitor (PDB ID: 2fs4) as well as in the unliganded state (PDB ID: 2g26).⁵⁹ Two domains were identified based on the PDP assignment [Fig. 10(D)]. The RMSD after global superposition of both ligand-bound and ligand-free structural forms is 1.2 Å. The

principal axes analysis revealed that the predominantly shear-type motion of the two domains of renin occurs upon the binding of a ligand and involves the shift of one domain relatively to the other, with relatively small changes in the angles that define the relative orientation of the principal axes of individual domains as depicted in Figure 10(D), right panel (the largest change corresponds to a rotation of 4.8 deg). These findings are additionally supported by the interdomain surface calculations, which indicate that the motion involves sliding over a tightly packed interface of $\sim 2700 \text{ \AA}^2$. This simple analysis gives a clear indication that the mechanism of domain motion in this case can be considered as shear-like.

DISCUSSION

This work focused on the local and global backbone structure similarity between ligand-bound and ligand-free forms of proteins. The secondary structure similarity between ligand-bound and ligand-free form was measured for a broad spectrum of local structure types. In general, the similarity of local ordered structure of ligand-bound versus ligand-free form is relatively high. Nevertheless, an alteration of secondary structure upon ligand binding/release was also detected in several cases, for example, lactoferrin, adenylate kinase, or EPSP synthase, and is mainly associated with hinge motions in multiple-domain proteins.

The analysis of global structural similarity between the ligand-bound and ligand-free forms of proteins showed that most of the proteins undergo relatively small conformational rearrangements of tertiary structure upon ligand binding/release. This observation was particularly evident for the individual domains of multiple-domain proteins where the vast majority is found to be insensitive to the state of ligand binding. It had been already demonstrated that the structural domains of the proteins present in the PDB are far more robust than originally thought, since ligand binding does not lead to a significant change of domain stability in the majority of proteins.⁶⁰ On the other hand, ligand-bound/ligand-free protein pairs characterized by relatively high RMSD were also observed, particularly if a protein was composed of more than one domain. RMSDs from the apo form $>1 \text{ \AA}$ were found for a significant fraction of multiple-domain proteins. This effect obviously corresponds to large-scale movements of entire domains required for signal transduction, including allostery^{61,62} or to facilitate capture or release of a ligand.^{63,64}

Many algorithms were developed to infer the mechanisms of the conformational changes based on the detailed comparisons of the different structures adopted by the same protein.^{15,22,23,65–67} A wide range of such experimentally observed molecular motions have been al-

ready classified and systematized in databases of macromolecular motions, such as MolMovDB^{20,21} or DynDom,^{23,24} that are freely available for structural biology community. Protein motions are usually classified into various categories first on the basis of size (fragment, domain, and subunit motions) and then on the basis of packing (shear, hinge, other) depending on whether or not they involve a parallel-plane sliding of one domain relative to the other.^{15,22} It is noteworthy that domain motions provide the most common examples of protein flexibility (in MolMovDB, the majority of motions were categorized as domain motions²¹). Here, we used a simple mechanistic model based on the mass-weighted principal axes calculation³⁶ to investigate the differences in the relative orientation of individual domains in ligand-free and ligand-bound forms of multiple-domain proteins. The results indicate that, in general, the conformational changes triggered by ligand binding/release involve relatively small hinge-bending movements as well as the shift of individual domains. On the other hand, large-scale domain movements were also observed. Among those, the shear-type mechanism was found to be the most common in the dataset, and is slightly more frequent than hinge-bending.

Several surveys have been carried out to examine the nature of domain–domain interactions.^{68–70} The results presented in this article confirmed that in most of the hinge-bending domain motions upon ligand binding, more interdomain interface surface becomes buried as a consequence of the closure of one domain onto the other, while a shear-type mechanism of domain motion maintains a tightly packed interdomain interface. Moreover, the analysis of the interface buried surface area revealed that some preferences toward the anticipated mechanism of protein domain movements are predictable by the examination of a ligand-free structural form alone. It has been already shown that the conservation of sequence and volume at domain interfaces can be used as a potential predictor for domain interactions.⁷⁰ In addition, the possibility of predicting functional motions in multiple-domain proteins through the recognition of key structural motifs that control interdomain motions has been suggested.⁶⁸

The analysis presented here also suggests that the knowledge of ligand binding state may provide assistance to protein structure prediction. It has been already shown that in some cases it may be impossible to correctly generate models without knowing the details about the exact biological function of a protein and its functional state.⁷¹ An estimated one-third of multiple-domain targets will require additional information concerning the state of ligand binding so that the mutual orientation of the domains can be appropriately described. However, even if the global structure of a ligand-free form does not differ significantly from the ligand-bound counterpart, many prediction algorithms penalize large

cavities and therefore will result in structurally distorted molecules. This suggests that the implementation of a protocol to maintain ligand binding cavities is required.

REFERENCES

- Butler BC, Hanchett RH, Rafailov H, MacDonald G. Investigating structural changes induced by nucleotide binding to RecA using difference FTIR. *Biophys J* 2002;82:2198–2210.
- Khan MY, Jaikaria NS, Frenz DA, Villanueva G, Newman SA. Structural changes in the NH2-terminal domain of fibronectin upon interaction with heparin. Relationship to matrix-driven translocation. *J Biol Chem* 1988;263:11314–11318.
- Toney JH, Wu L, Summerfield AE, Sanyal G, Forman BM, Zhu J, Samuels HH. Conformational changes in chicken thyroid hormone receptor $\alpha 1$ induced by binding to ligand or to DNA. *Biochemistry* 1993;32:2–6.
- Li J, Zagotta WN, Lester HA. Cyclic nucleotide-gated channels: structural basis of ligand efficacy and allosteric modulation. *Q Rev Biophys* 1997;30:177–193.
- Karthikeyan S, Zhou Q, Osterman AL, Zhang H. Ligand binding-induced conformational changes in riboflavin kinase: structural basis for the ordered mechanism. *Biochemistry* 2003;42:12532–12538.
- Malkowski MG, Martin PD, Guzik JC, Edwards BF. The co-crystal structure of unliganded bovine α -thrombin and prethrombin-2: movement of the Tyr-Pro-Pro-Trp segment and active site residues upon ligand binding. *Protein Sci* 1997;6:1438–1448.
- Zavodszky MI, Kuhn LA. Side-chain flexibility in protein-ligand binding: the minimal rotation hypothesis. *Protein Sci* 2005;14:1104–1114.
- Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. *Proteins* 2000;39:261–268.
- Doring K, Surrey T, Nollert P, Jahnig F. Effects of ligand binding on the internal dynamics of maltose-binding protein. *Eur J Biochem* 1999;266:477–483.
- Kraft L, Sprenger GA, Lindqvist Y. Conformational changes during the catalytic cycle of gluconate kinase as revealed by X-ray crystallography. *J Mol Biol* 2002;318:1057–1069.
- Arnold GE, Manchester JJ, Townsend BD, Ornstein RL. Investigation of domain motions in bacteriophage T4 lysozyme. *J Biomol Struct Dyn* 1994;12:457–474.
- Mizutani K, Mikami B, Hirose M. Domain closure mechanism in transferrins: new viewpoints about the hinge structure and motion as deduced from high resolution crystal structures of ovotransferrin N-lobe. *J Mol Biol* 2001;309:937–947.
- Minard P, Hall L, Betton JM, Missiakas D, Yon JM. Efficient expression and characterization of isolated structural domains of yeast phosphoglycerate kinase generated by site-directed mutagenesis. *Protein Eng* 1989;3:55–60.
- Sherman MA, Chen Y, Mas MT. An engineered amino-terminal domain of yeast phosphoglycerate kinase with native-like structure. *Protein Sci* 1997;6:882–891.
- Gerstein M, Lesk AM, Chothia C. Structural mechanisms for domain movements in proteins. *Biochemistry* 1994;33:6739–6749.
- Lesk AM, Chothia C. Mechanisms of domain closure in proteins. *J Mol Biol* 1984;174:175–191.
- Gee AC, Katzenellenbogen JA. Probing conformational changes in the estrogen receptor: evidence for a partially unfolded intermediate facilitating ligand binding and release. *Mol Endocrinol* 2001;15:421–428.
- Miyashita O, Onuchic JN, Wolynes PG. Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci USA* 2003;100:12570–12575.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–242.
- Echols N, Milburn D, Gerstein M. MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res* 2003;31:478–482.
- Gerstein M, Krebs W. A database of macromolecular motions. *Nucleic Acids Res* 1998;26:4280–4290.
- Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M. Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic. *Proteins* 2002;48:682–695.
- Lee RA, Razaz M, Hayward S. The DynDom database of protein domain motions. *Bioinformatics* 2003;19:1290–1291.
- Qi G, Lee R, Hayward S. A comprehensive and non-redundant database of protein domain movements. *Bioinformatics* 2005;21:2832–2838.
- Frimurer TM, Peters GH, Iversen LF, Andersen HS, Moller NP, Olsen OH. Ligand-induced conformational changes: improved predictions of ligand binding conformations and affinities. *Biophys J* 2003;84:2273–2281.
- Sippl MJ, Lackner P, Domingues FS, Koppensteiner WA. An attempt to analyse progress in fold recognition from CASP1 to CASP3. *Proteins* 1999;Suppl 3:226–230.
- Venclovas C, Zemla A, Fidelis K, Moutl J. Assessment of progress over the CASP experiments. *Proteins* 2003;53 (Suppl 6):585–595.
- Zemla A, Venclovas C, Moutl J, Fidelis K. Processing and evaluation of predictions in CASP4. *Proteins* 2001;Suppl 5:13–21.
- Kosinski J, Cymerman IA, Feder M, Kurowski MA, Sasin JM, Bujnicki JM. A "Frankenstein's monster" approach to comparative modeling: merging the finest fragments of Fold-Recognition models and iterative model refinement aided by 3D structure evaluation. *Proteins* 2003;53 (Suppl 6):369–379.
- Zhang Y, Skolnick J. Tertiary structure predictions on a comprehensive benchmark of medium to large size proteins. *Biophys J* 2004;87:2647–2655.
- Bradley P, Malmstrom L, Qian B, Schonbrun J, Chivian D, Kim DE, Meiler J, Misura KM, Baker D. Free modeling with Rosetta in CASP6. *Proteins* 2005;61 (Suppl 7):128–134.
- Sobolev V, Edelman M. Modeling the quinone-B binding site of the photosystem-II reaction center using notions of complementarity and contact-surface between atoms. *Proteins* 1995;21:214–225.
- Sobolev V, Sorokine A, Prilusky J, Abola EE, Edelman M. Automated analysis of interatomic contacts in proteins. *Bioinformatics* 1999;15:327–332.
- Alexandrov N, Shindyalov I. PDP: protein domain parser. *Bioinformatics* 2003;19:429–430.
- Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 1983;22:2577–2637.
- Browner MF, Fauman EB, Fletterick RJ. Tracking conformational states in allosteric transitions of phosphorylase. *Biochemistry* 1992;31:11297–11304.
- Xiang Z, Honig B. Extending the accuracy limits of prediction for side-chain conformations. *J Mol Biol* 2001;311:421–430.
- Roelofs WL. Chemistry of sex attraction. *Proc Natl Acad Sci USA* 1995;92:44–49.
- Hansson BS. Olfaction in Lepidoptera. *Experientia* 1995;51:1003–1027.
- Vogt RG, Riddiford LM. Pheromone binding and inactivation by moth antennae. *Nature* 1981;293:161–163.
- Sandler BH, Nikonova L, Leal WS, Clardy J. Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. *Chem Biol* 2000;7:143–151.
- Lautenschlager C, Leal WS, Clardy J. Coil-to-helix transition and ligand release of *Bombyx mori* pheromone-binding protein. *Biochem Biophys Res Commun* 2005;335:1044–1050.
- Talarico TL, Dev IK, Dallas WS, Ferone R, Ray PH. Purification and partial characterization of 7,8-dihydro-6-hydroxymethylpterinpyrophosphokinase and 7,8-dihydropteroate synthase from *Escherichia coli* MC4100. *J Bacteriol* 1991;173:7029–7032.

44. Hitchings GH, Burchall JJ. Inhibition of folate biosynthesis and function as a basis for chemotherapy. *Adv Enzymol Relat Areas Mol Biol* 1965;27:417–468.
45. Blaszczyk J, Li Y, Shi G, Yan H, Ji X. Dynamic roles of arginine residues 82 and 92 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: crystallographic studies. *Biochemistry* 2003;42:1573–1580.
46. Blaszczyk J, Li Y, Wu Y, Shi G, Ji X, Yan H. Essential roles of a dynamic loop in the catalysis of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase. *Biochemistry* 2004;43:1469–1477.
47. Li Y, Wu Y, Blaszczyk J, Ji X, Yan H. Catalytic roles of arginine residues 82 and 92 of *Escherichia coli* 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: site-directed mutagenesis and biochemical studies. *Biochemistry* 2003;42:1581–1588.
48. Xiao B, Shi G, Gao J, Blaszczyk J, Liu Q, Ji X, Yan H. Unusual conformational changes in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by X-ray crystallography and NMR. *J Biol Chem* 2001;276:40274–40281.
49. Blaszczyk J, Shi G, Yan H, Ji X. Catalytic center assembly of HPPK as revealed by the crystal structure of a ternary complex at 1.25 Å resolution. *Structure* 2000;8:1049–1058.
50. Oeschger MP, Bessman MJ. Purification and properties of guanylate kinase from *Escherichia coli*. *J Biol Chem* 1966;241:5452–5460.
51. Blaszczyk J, Li Y, Yan H, Ji X. Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes. *J Mol Biol* 2001;307:247–257.
52. Holland TA, Veretnik S, Shindyalov IN, Bourne PE. Partitioning protein structures into domains: why is it so difficult? *J Mol Biol* 2006;361:562–590.
53. Pittard J, Wallace BJ. Distribution and function of genes concerned with aromatic biosynthesis in *Escherichia coli*. *J Bacteriol* 1966;91:1494–1508.
54. Park H, Hilsenbeck JL, Kim HJ, Shuttleworth WA, Park YH, Evans JN, Kang C. Structural studies of *Streptococcus pneumoniae* EPSP synthase in unliganded state, tetrahedral intermediate-bound state and S3P-GLP-bound state. *Mol Microbiol* 2004;51:963–971.
55. Ellison V, Stillman B. Opening of the clamp: an intimate view of an ATP-driven biological machine. *Cell* 2001;106:655–660.
56. Jeruzalmi D, O'Donnell M, Kuriyan J. Clamp loaders and sliding clamps. *Curr Opin Struct Biol* 2002;12:217–224.
57. Podobnik M, Weitze TF, O'Donnell M, Kuriyan J. Nucleotide-induced conformational changes in an isolated *Escherichia coli* DNA polymerase III clamp loader subunit. *Structure* 2003;11:253–263.
58. Reid IA, Morris BJ, Ganong WF. The renin-angiotensin system. *Annu Rev Physiol* 1978;40:377–410.
59. Holsworth DD, Cai C, Cheng XM, Cody WL, Downing DM, Erasga N, Lee C, Powell NA, Edmunds JJ, Stier M, Jalaie M, Zhang E, McConnell P, Ryan MJ, Bryant J, Li T, Kasani A, Hall E, Subedi R, Rahim M, Maiti S. Ketopiperazine-based renin inhibitors: optimization of the "C" ring. *Bioorg Med Chem Lett* 2006;16:2500–2504.
60. Yesylevskyy SO, Kharkyanen VN, Demchenko AP. The change of protein intradomain mobility on ligand binding: is it a commonly observed phenomenon? *Biophys J* 2006;91:3002–3013.
61. He X, Chow D, Martick MM, Garcia KC. Allosteric activation of a spring-loaded natriuretic peptide receptor dimer by hormone. *Science* 2001;293:1657–1662.
62. Marvin JS, Corcoran EE, Hattangadi NA, Zhang JV, Gere SA, Hellinga HW. The rational design of allosteric interactions in a monomeric protein and its applications to the construction of biosensors. *Proc Natl Acad Sci USA* 1997;94:4366–4371.
63. Al-Rabee R, Zhang Y, Grant GA. The mechanism of velocity modulated allosteric regulation in D-3-phosphoglycerate dehydrogenase. Site-directed mutagenesis of effector binding site residues. *J Biol Chem* 1996;271:23235–23238.
64. Flachner B, Varga A, Szabo J, Barna L, Hajdu I, Gyimesi G, Zavodszky P, Vas M. Substrate-assisted movement of the catalytic Lys 215 during domain closure: site-directed mutagenesis studies of human 3-phosphoglycerate kinase. *Biochemistry* 2005;44:16853–16865.
65. Hayward S, Lee RA. Improvements in the analysis of domain motions in proteins from conformational change: DynDom version 1.50. *J Mol Graph Model* 2002;21:181–183.
66. 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.
67. Alexandrov V, Lehnert U, Echols N, Milburn D, Engelman D, Gerstein M. Normal modes for predicting protein motions: a comprehensive database assessment and associated Web tool. *Protein Sci* 2005;14:633–643.
68. Hayward S. Structural principles governing domain motions in proteins. *Proteins* 1999;36:425–435.
69. Hubbard SJ, Argos P. A functional role for protein cavities in domain: domain motions. *J Mol Biol* 1996;261:289–300.
70. Littler SJ, Hubbard SJ. Conservation of orientation and sequence in protein domain–domain interactions. *J Mol Biol* 2005;345:1265–1279.
71. Venclovas C. Comparative modeling in CASP5: progress is evident, but alignment errors remain a significant hindrance. *Proteins* 2003;53 (Suppl 6):380–388.
72. Goldstein H, Poole CP, Safko J. *Classical mechanics*. Addison-Wesley: Reading, MA; 2002.

APPENDIX: PRINCIPAL AXES CALCULATION AND SIMPLE VISUALIZATION OF PROTEIN DOMAINS

In mechanics, a rigid body is an idealization of a solid body of finite size, in which deformation is neglected.⁷² In other words, there is no migration of mass within a rigid body regardless of external forces exerted on it. Any rigid body has three principal axes, which are mutually orthogonal, or can be chosen to be so. A principal axis may be simply defined as one about which no torque is needed to maintain rotation at a constant angular velocity. For a rigid object of N point masses m_i , the moment of inertia tensor is given by

$$I = \begin{bmatrix} I_{xx} & I_{xy} & I_{xz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix}$$

where the components can be calculated from the Cartesian coordinates (x_i, y_i, z_i) :

$$I_{xx} = \sum_{i=1}^N m_i (y_i^2 + z_i^2) \quad I_{yy} = \sum_{i=1}^N m_i (x_i^2 + z_i^2) \\ I_{zz} = \sum_{i=1}^N m_i (x_i^2 + y_i^2)$$

$$I_{xy} = I_{yx} = - \sum_{i=1}^N m_i x_i y_i \quad I_{xz} = I_{zx} = - \sum_{i=1}^N m_i x_i z_i \\ I_{yz} = I_{zy} = - \sum_{i=1}^N m_i y_i z_i$$

assuming the origin as the center of mass.

By appropriate choice of the orientation of the body-fixed coordinate system, the inertia tensor can be reduced to the diagonal form:

$$I = \begin{bmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{bmatrix}$$

where the coordinate axes are the principal axes (the set of eigenvectors that diagonalize the inertia tensor) and the root constants I_1 , I_2 , and I_3 correspond to the moment of inertia about each principal axis. In general, the principal axes depend upon both the shape of the body and the choice of origin.

In a multiple-domain protein, the principal axes can describe the relative orientation of individual domains, assuming the latter as rigid bodies. We used the mass-weighted principal axes, which have been shown to provide an intuitive means of characterizing domain relationships within a protein, as well as the disposition of domains in different protein conformers.³⁶ In this

approach, unit vectors along the principal axes are weighted by the mass distribution projected along the corresponding axis. The shape of a protein domain is accurately described by the set of shape-weighted axes, denoted as L , M , and S , where $L \leq M \leq S$. The mutual orientation of individual domains in a multiple-domain protein can be described by the angles between the sets of principal axes of each domain. For a protein composed of two domains the angle between their two L vectors (L_1 and L_2) is given by $\theta_L = \arccos(L_1 \cdot L_2 / |L_1||L_2|)$. The values of θ_M and θ_S can be calculated analytically.

To visualize the differences in the relative orientation of individual domains in the two conformers of a multiple-domain protein, the domains can be presented as ellipsoids. A full-size ellipsoid, which represents a protein domain, is created by scaling its mass-weighted axes by a factor determined from geometric constraints: $\sqrt{5} \left(x^2 \frac{L}{|L|} + y^2 \frac{M}{|M|} + z^2 \frac{S}{|S|} \right) = 1$. Such ellipsoids reflect the shape of the protein domain and provide an easy means of visualizing domain movements.³⁶