BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 5683–5704 5683
DOI: 10.1021/bi100110x

Diversity of Function-Related Conformational Changes in Proteins: Coordinate Uncertainty, Fragment Rigidity, and Stability[†]

Alexander A. Rashin,*[,‡,∥] Abraham H. L. Rashin,[‡,§] and Robert L. Jernigan[∥]

[‡]*BioChemComp Inc., 543 Sagamore Avenue, Teaneck, New Jersey 07666,* [§]*Rutgers, The State University of New Jersey, 22371 BPO Way, Piscataway, New Jersey 08854-8123, and* [∥]*L. H. Baker Center for Bioinformatics and Department of Biochemistry, Biophysics and Molecular Biology, 112 Office and Lab Building, Iowa State University, Ames, Iowa 50011-3020*

Received January 25, 2010; Revised Manuscript Received May 14, 2010

ABSTRACT: It was found that the variety of function-related conformational changes ("movements") in proteins is beyond the earlier simple classifications. Here we offer biochemists a more comprehensive, transparent, and easy-to-use approach allowing a detailed and accurate interpretation of such conformational changes. It makes possible a more multifaceted characterization of protein flexibility via identification of rigidly and nonrigidly repositioned fragments, stable and nonstable fragments, and domain and nondomain repositioning. "Coordinate uncertainty thresholds" derived from computed differences between independently determined coordinates of the same molecules are used as the criteria for conformational identity. "Identical" rigid substructures are localized in the distance difference matrices (DDMs). A sequence of simple transformations determines whether a structural change occurs by rigid-body movements of fragments or largely through non-rigid-body deformations. We estimate the stability of protein fragments and compare stable and rigidly moving fragments. The motions computed with the coarse-grained elastic networks are also compared to those of their DDM analogues. We study and suggest a classification for 17 structural pairs, differing in their functional states. For five of the 17 proteins, conformational change cannot be accomplished by rigid-body transformations and requires significant non-rigid-body deformations. Stable fragments rarely coincide with rigidly moving fragments and often disagree with the CATH identifications of domains. Almost all monomeric apo chains, containing stable fragments and/or domains, indicate instability of the entire molecule, suggesting the importance of fragments and domains motions prior to stabilization by substrate binding or crystallization. Notably, kinases exhibit the greatest extent of nonrigidity among the proteins investigated.

The importance of flexibility for protein function is well-understood (*1*). The conformational changes in response to the binding of another molecule are of the utmost importance for protein function. These conformational changes are often termed "protein motions", while usually only a few snapshots can be gleaned from the experimental structures. Such a change can involve only a single side chain or may be as dramatic as long-range repositioning ("movement") of large fragments or domains and even partial folding and/or refolding. It was suggested earlier that whole folded domains never undergo large distortions at ordinary temperatures (*1*). However, this last suggestion seems to be open to question (see below).

Chothia, Lesk, and Gerstein pioneered studies of large-scale protein flexibility from comparisons of two or more structural states of the same protein (*2*−*4*). Their studies led to the creation of an important database of significant protein motions (*5*−*8*). They suggested a classification of the major types of large protein motions (*4*, *7*) characterized by three extents of magnitude (no motion, minor movers, and major movers), three sizes (fragment, domain, and subunit), and three mechanisms (hinge, shear, and other). More recently, they found that "The degree of movement in many of the structures that have been examined is striking, particularly in light of the variety of mechanisms involved" (*9*). This makes the original classification (*8*) seem oversimplified and justifies the need for a new scheme for classifying flexibility in proteins (*10*).

It has often been stated (*1*, *4*) that large conformational changes in proteins involve repositioning of rigid loops or preformed rigid domains. This appears to be based on limited analysis and unsettled definitions. There is still no consensus of what constitutes a domain (*11*, *12*), making many structural interpretations dubious. The earliest of the two major definitions considered structural domains to be parts of protein structures having more interactions inside individual domains than between them (*13*, *14*). A later definition considered a structural domain to be a segment (or several segments) of the polypeptide that forms a compact and stable structure, with a hydrophobic core, and which can fold by itself (*12*, *15*−*20*). Because some papers identified domains with structural "modules" that move relative to each other in protein functioning (*21*), it raises the question of whether such modules are stable and can fold independently.

Furthermore, it has been suggested that "if you know how it moves, you can infer how it works; the knowledge of structural flexibility offers a straight-line connection between structure and function" (*22*). In fact, one of the central issues in studies of protein flexibility is how much of which observed protein

[†]This work was supported by National Institutes of Health Grants R01GM072014 and R01GM081680 and National Science Foundation Grant CNS-0521568.

*To whom correspondence should be addressed. Telephone: (201) 836-7960. E-mail: Alexander_Rashin@hotmail.com. Fax: (515) 294-5256.

Published on Web 05/14/2010

conformational change is actually required for its function (*5*, *10*, *23*), especially since many changes can be induced by crystallization (*10*) or other factors not necessarily directly related to function.

Because a few snapshots of a protein's conformational change cannot fully reveal all details along the pathways of such change, pathway movies of the motions (morphs) have been constructed as interpolations between two crystal structures (*8*, *24*). "Straight Cartesian interpolation" and "interpolation with restraints", where energy minimization is applied to interpolated intermediates, can however yield intermediate structures with rather high energies. Newer methods (*25*, *26*) based on elastic models for creation of the motion pathway movies appear to be better in avoiding high energies and more informative about intermediate structures.

Our goal here is to illustrate and provide for biochemists methods (*10*, *16*, *19*, *25*, *26*), allowing a clear, detailed, and accurate interpretation of function-related conformational changes in proteins. To move toward this goal here, we (1) use distance difference matrices (DDM)[1] (*27*) utilizing "coordinate uncertainty thresholds" and a novel, simple, and fast algorithm to identify and characterize rigid-body movements in proteins (*10*, *28*); (2) verify whether a sequence of rigid-body transformations can bring two functionally different protein conformations into coincidence within an "uncertainty threshold" between their coordinates or whether non-rigid-body deformations significantly contribute to function-related motions; (3) determine whether domains and loops involved in function-related conformational changes "move" as rigid bodies or exhibit more complex behaviors; (4) estimate the stability of the protein fragments and/or domains (*16*−*19*) to study whether they behave as rigid bodies in protein functional motions and what other roles stability may play in such motions; (5) propose a more detailed classification of functional motions in proteins; and (6) generate pathways of protein functional motions using elastic models of conformational transitions (*25*, *26*) and compare them to DDMs of the intermediates of such pathways.

METHODS

Distance Difference Matrices (*DDM*). For a protein of N residues, the distance matrix (DM) is a square $N \times N$ matrix, in which element ij represents the distance between residues i and j (or their C^{α} atoms as used here). Because the distances i to j and j to i are equal, usually only half of the DM matrix is considered (*27*). For two different conformations of the same protein, a distance difference map (DDM) is constructed as an $N \times N$ matrix of differences (DDs) between the corresponding elements of the two DMs (*10*). We follow our previous work (*10*) representing DDMs in three shades (black, gray, and white) based on the ranges of the absolute DD values falling below 0.5 Å, between 0.5 and 1 Å, and above 1 Å. Each DDM is characterized by the rms of all its DDs, denoted as RMSDD, and with the percentage of DD values lying outside the range from −1 to 1 Å, denoted as Δ.

Contact Distance Difference Matrices (*CDDM*) *and an Estimation of the Extent of Shear*. Shear movements (*3*) can occur at contact interfaces. It has been found (*6*) that, because of packing restrictions, shearing segments move relative to one another by no more than 2 Å. To estimate the amount of restricted motion at contact interfaces, we have constructed contact distance difference matrices (CDDM). Contact distance matrices (CDM) for each structure (indexed by $k = 1$ or 2) contain only $|(C^{\alpha i})_k - (C^{\alpha j})_k|$ distances shorter than a "contact" cutoff, chosen here as 8 Å, based on tabulated distances between contacting helices and β-strands in proteins (*29*, *30*). For each ij element, marked as a contact on at least one of two CDMs, a distance difference is calculated and marked on the CDDM. All ij positions that were not marked as a contact on both CDMs are shown as blank spots on the CDDM. The percentage of contact distance changes within any range can be computed from the CDDM. We calculate such percentages only for $j > i + 4$, to avoid a dominance in our contact statistics of contacts within α-helices and turns. An example of a CDDM is shown in Figure S2b of the Supporting Information.

We estimate the magnitude of shear in each functional "motion" by evaluating the percentage, δ, of contact distances in the corresponding protein's CDDM that change between 1 and 2 Å. A δ of $< 1.5\%$ includes pairs of structures with only coordinate uncertainty (see the section below). Here we take δ values of $< 1.5\%$ as an indication of no shear.

Estimation of Positional Uncertainties. We determined the range of coordinate uncertainties in our previous work (*10*) by calculating and analyzing the DDMs of 1014 pairs of structures (at approximately room temperature) of bovine ribonuclease A and of whale myoglobin, for which the authors did not report any significant conformational changes ("movements"). Each pair was characterized by its DDM, RMSDD, and Δ (see above). All of these 1014 pairs of structures had an RMSDD of ≤0.44 Å and a Δ of ≤4.05%.

Two DDMs (endothiapepsin, 4ape5er2, and thermolysin, 1l3f3tmn) corresponding to functionally significant conformational changes ("motions") (*5*, *7*) have the lowest RMSDs of 0.45 and 0.46 Å among more than 20 DDMs of functional "motions" that we have studied and Δ values of 5.2 and 3.4%, respectively.

The RMSDD and Δ values listed above led us to the criterion that a DDM indicates no significant "motion" but only coordinate uncertainty when the RMSDD is below 0.46 Å and its Δ is less than 5% (*10*). These criteria obviously classify DDMs 4ape5er2 and 1l3f3tmn as indicating significant conformational changes ("motions"), while classifying 1014 DDMs of ribonucleases and myoglobins as indicating the structural identity of the corresponding protein pairs within the coordinate uncertainty threshold. Further accumulation and analysis of data might change these criteria somewhat. For further details and discussion, see our previous paper (*10*).

Location of "Rigid" Fragments and Blocks. All DDs within a continuous fragment of the protein form a right triangle along the DDM's diagonal corresponding to this fragment. If this fragment's conformations in two molecules differ only within the coordinate uncertainty limits, then the right triangle corresponding to this fragment should be almost entirely black containing minimal white or gray areas. Such a fragment is likely to change its position in a conformational transition as a rigid body.

Two such rigid fragments (2−164 and 174−286) with clearly delineated nearly black triangles with straight white borders are exemplified in DDM 2tbvA2tbvC of chains A and C of Tobacco Bushy Stunt Virus capsid protein (Figure 1a). A largely white

[1]Abbreviations: PDB, Protein Data Bank; PQS, protein quaternary structure server; ASA, accessible surface area; DM, matrix of C^{α}−C^{α} distances in a protein conformation; DDM, matrix of differences between elements of two DMs; DD, element of a DDM; RMSDD, root-mean-square deviation of all elements of a DDM; CDDM, like a DDM but only "contact distances" considered; SM, matrix of stabilities of all folded protein fragments; ENM, elastic network model.

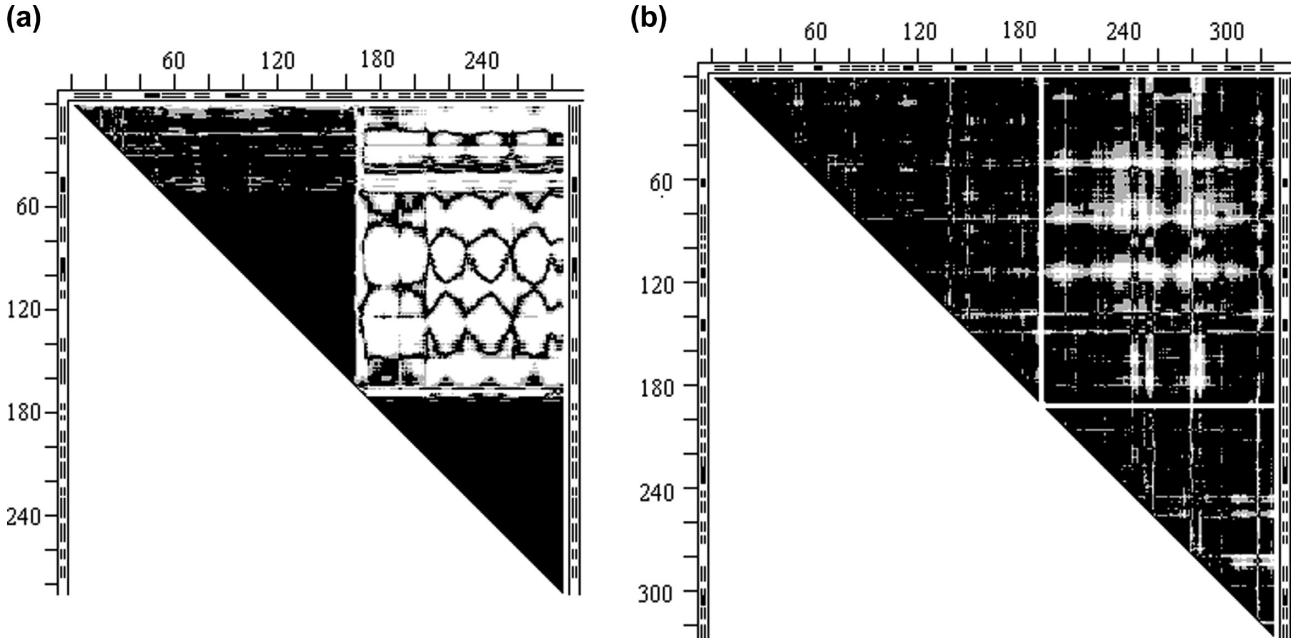

FIGURE 1: Distance difference matrices (DDMs). Short thick bars or segments of thin double lines along the tops and sides of the triangular matrices denote positions of α-helices or β-strands, respectively (taken from the PDB). Tick marks on the top and left are at intervals of 20 residues. Mostly black triangular areas mean that DDs in them are below 0.5 Å, and these areas represent likely rigid blocks in the conformational transitions between the two compared molecules; the gray areas mean that DDs are between 0.5 and 1 Å, and white spaces mean that the absolute values of the distance differences (DDs) between the corresponding pair of C^{α} atoms in the two structures (e.g., PDB entries) is more than 1 Å. (a) 2tbvAC for molecules A and C in the unit cell of the 2tbv virus capsid protein; the mostly white rectangle at the DDM's top right, providing straight white borders of the "rigid-body triangles" of this DDM, indicates that most DDs between the two rigid bodies are larger than 1 Å, meaning that most $C^{\alpha}-C^{\alpha}$ distances between the rigid bodies differ in two molecules by more than 1 Å. (b) 4ape5er2 of the apo−holo structures of endothiapepsin. This DDM has only some white spots and no straight white borders of the black rigid triangles; such borders can be provided by drawing white lines encompassing almost all white spots in a rectangle and thus delineating the mostly black rigid triangles (see also the text on thermolysin below).

rectangle indicates that most $C^{\alpha}-C^{\alpha}$ distances between the two rigid fragments changed by more than 1 Å.

No such clearly delineated black triangles with straight white borders are seen in DDM 4ape5er2 of the apo−holo pair of endothiapepsin structures, which shows only a set of white spots in its top right part. However, we are interested not in straight white borders, but in almost solidly black triangles that likely represent rigid fragments. Two such mostly black triangles can readily be separated (by drawing straight white lines as in Figure 1b) from the area containing almost all white spots. These white spots do not represent "distributed local distortions" but rather a systematic small amplitude difference between the positions of the rigid fragments in two structures, which can be reduced below the uncertainty threshold by a rigid-body transformation (see Results).

Fitting of Rigid Fragments. For the fitting of rigid or nearly rigid fragments, we use the fast and accurate algorithm described in detail in our previous paper (*10*), which uses a superposition of three noncollinear points of a rigid body to superimpose all points of this body, together with quaternion rotations (*28*, *31*−*34*). Such a superposition can be represented as an initial superposition of the centers of mass of the fragments (represented by C^{α} atoms) followed by two rotations around the axes passing through the center of mass and "nearly" (*10*) superimposing two C^{α} atoms from each fragment. A few residues at proteins' termini or next to crystallographically unresolved fragments are sometimes excluded from the fitting. Fragments of no fewer than three residues were used.

Sequence of Fitting Steps and Evaluation of the Result. In the first step, the largest rigid fragment of the second molecule is fit to its coordinates in the first (reference) molecule and the resultant transformation is applied to coordinates of the entire second molecule (this does not change the RMSDD; however, parameters of all subsequent rigid-body movements depend on the choice of the first transformation). Coordinates corresponding to atoms of the bound substrate (or cofactor) are not included in these calculations.

In the following fitting steps, the structures of all fragments of the entire sequence of the second molecule should be fit to the structure of the corresponding fragments of the first molecule to verify whether the "functional movement" is a result of a series of rigid-body movements of protein fragments. The particular order of fitting steps is arbitrary. For details, see our previous paper (*10*). An example of the transformation sequence and its outputs is provided in the Supporting Information.

If the RMSDD and Δ after rigid-body fittings of the entire second structure to the first are outside the uncertainty limits, it means that the functional "movement" involves significant nonrigid deformation of the main chain. It may be useful to look at a DDM to obtain a fast estimate of whether one structure of a protein can be transformed into another by no more than 12 rigid-body movements. Here we offer two such simple cues. More might emerge as we gain experience.

An idealized L-shaped white band is shown in Figure 2a. Instead of a half-matrix above the diagonal, which we use throughout the paper, it is convenient here to draw the entire matrix with triangular halves related by a diagonal reflection. Here we consider (in a highly simplified manner) the relationship of the width of the band, measured by the number of residues, and Δ.

Let N be the number of residues in a protein in Figure 2a, w the width of the band given as the number of residues, S the area of

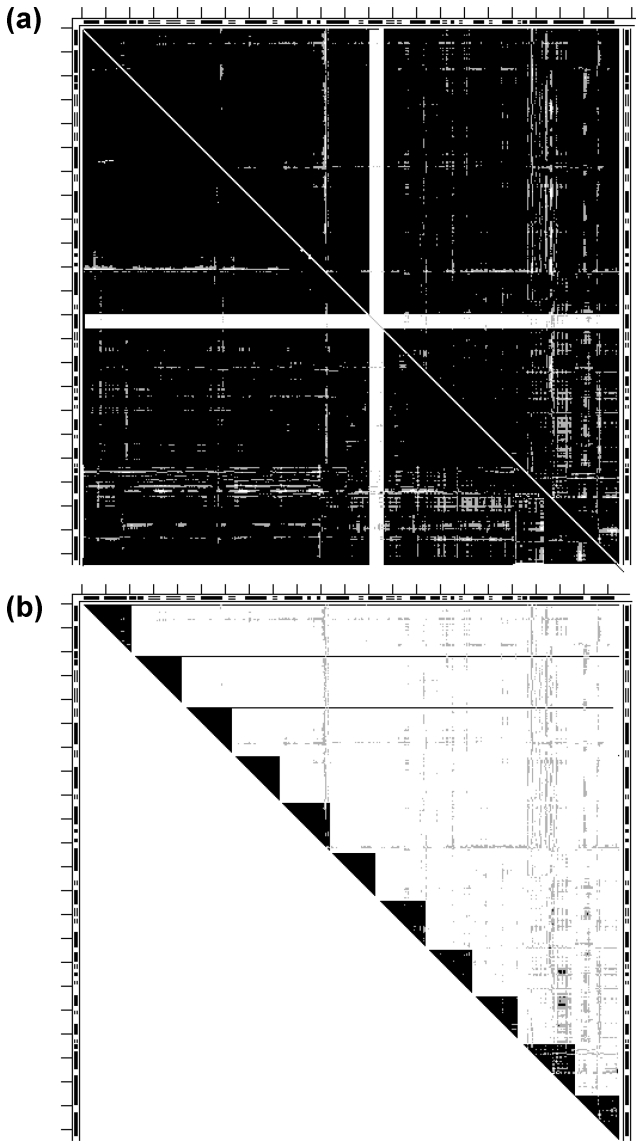

FIGURE 2: Simple guides to estimate the flexibilities of structures. (a) Contribution of L-shaped white bands (see the text). (b) Role of the size of the rigid fragments (see the text).

the entire DDM as the number of residues squared, and S_{Lb} the area of the L-band. Then $S = N^2$, and $2S_{Lb} = 2Nw$. If we ignore double counting where straight bands intersect, then $2 \times S_{Lb}/S = 2w/N$ is the ratio of DDs larger than 1 to the total number of DDs in the DDM [$\Delta = 100 \times 2w/N$; $w = N \times \Delta/(200)$]; if the threshold is $\Delta = 5(\%)$, then $w_{thr} = 0.025N$. This means that for a protein of 400 residues the critical width of the L-band (w_{thr}) would be 10 residues. For two equal-sized L-bands, it would be five residues each. A total width of L-bands in a DDM larger than this indicates that finding rigid transformations is unlikely. Note that this estimate holds only for L-bands with no black triangles in their corners. Such black triangles can be moved as rigid bodies eliminating the white L-band in the DDM.

Our use of a cutoff of 12 rigid-body transformations to label a functional "motion" as nonrigid, highly "flexible", or "glove" is somewhat arbitrary. We might instead choose to scale the cutoff number of rigid transformations by the protein length.

Let us assume, as shown in Figure 2b, that all rigid blocks (black triangles in the DDM) have an equal width, w, and also assume that m is the allowed number of rigid-body transformations (equal to the number of black triangles). Then $w = N/m$; if $N = 204$ and $m = 12$, $w_{200} = 17$ (residues). This means that if the white area comes mostly closer than w from the diagonal, one might fail to rigidly transform one structure into another within the required cutoff number, m, of rigid-body transformations.

Estimations of the Stability of Protein Fragments and Blocks. We previously suggested that stable fragments with nativelike conformations within proteins can be located on the basis of calculations of the buried surface area (*15*−*19*). Theoretically predicted stable fragments (*18*) were successfully found experimentally in thermolysin (*35*). While this method has its limitations, it nonetheless remains a useful tool for domain identification (*11*) and can serve as a tested guide in the experimental isolation of stable protein fragments. We use it here to estimate the stability of all protein fragments, including rigidly moving blocks. Results are also compared to CATH (*36*) or SCOP (*37*) domains. Only polypeptide parts of proteins were included in the stability calculations. When PDB (*38*) entries contain alternative side chain conformations, the "A" conformer is always used [a selection of alternative conformers was found (*17*) to have only a minor effect on the results].

An estimate of free energy $\Delta G_D{}^N$ corresponding to the stability of a protein fragment of molecular weight M can be obtained from the expression below (*17*)

$$-\Delta G_D{}^N = \Delta G_B - T\Delta S_{conf} + T\Delta S_{S-S} \quad (1)$$

where $\Delta G_D{}^N$ is protein stability, ΔG_B is assumed to be proportional to the surface area, B, buried upon folding of the fragment, $T\Delta S_{conf}$ is the loss of conformational entropy of the chain without disulfide cross-links upon folding and is a function of its molecular weight, and $T\Delta S_{S-S}$ is the decrease in the absolute value of this entropy caused by disulfide cross-links. For a protein of N residues, an $N \times N$ table of stabilities of all of its continuous fragments forms a "stability matrix" (SM). Note that eq 1 should be considered as an empirical formula allowing a reasonable degree of success but leaving aside controversies regarding the mechanism of protein folding (*19*, *39*, *40*). Parameters used here for the stability calculations and some general estimates of the method's accuracy are provided in section 1 of the Supporting Information.

Here we show a fragment of SM for the X-ray structure of residues 1−180 of 2tbvA (Figure 3a) and its ribbon representation (Figure 3b).

The SM and its cross sections (*15*, *18*, *19*) closely resemble "free energy landscapes" introduced later. Figure 3a shows that an unfolded chain has to cross a free energy barrier (white area) to reach a valley with the largest calculated stability of −5 kcal/mol. The longest fragment with this stability includes 162 residues starting from the second residue (fragment 2−163). Except for minor variations in the position of its ends, this tightly folded fragment can be easily seen in Figure 3b. It ends at the tip of the rightmost arrow on a β-strand, and it is obvious that the following coil and β-strand region do not interact with this stable domain. Even in this rather simple case, the SM has a large size that can be analyzed on a computer but is inconvenient to print. Therefore, in Results, we use three-color bitmap SMs with white for fragments with ΔG values above a chosen threshold, T, of instability (we used $T = 5$ or 10 kcal/mol), gray for ΔG values from 1 to T, and black for 0 and all negative ΔG values. The three-color bitmap SM of the entire 2tbvA is shown in Figure 3c, and the locations of the most stable fragments are listed in Table 3.

In addition to the general estimates of the method's accuracy (see the Supporting Information), we also obtained a rough

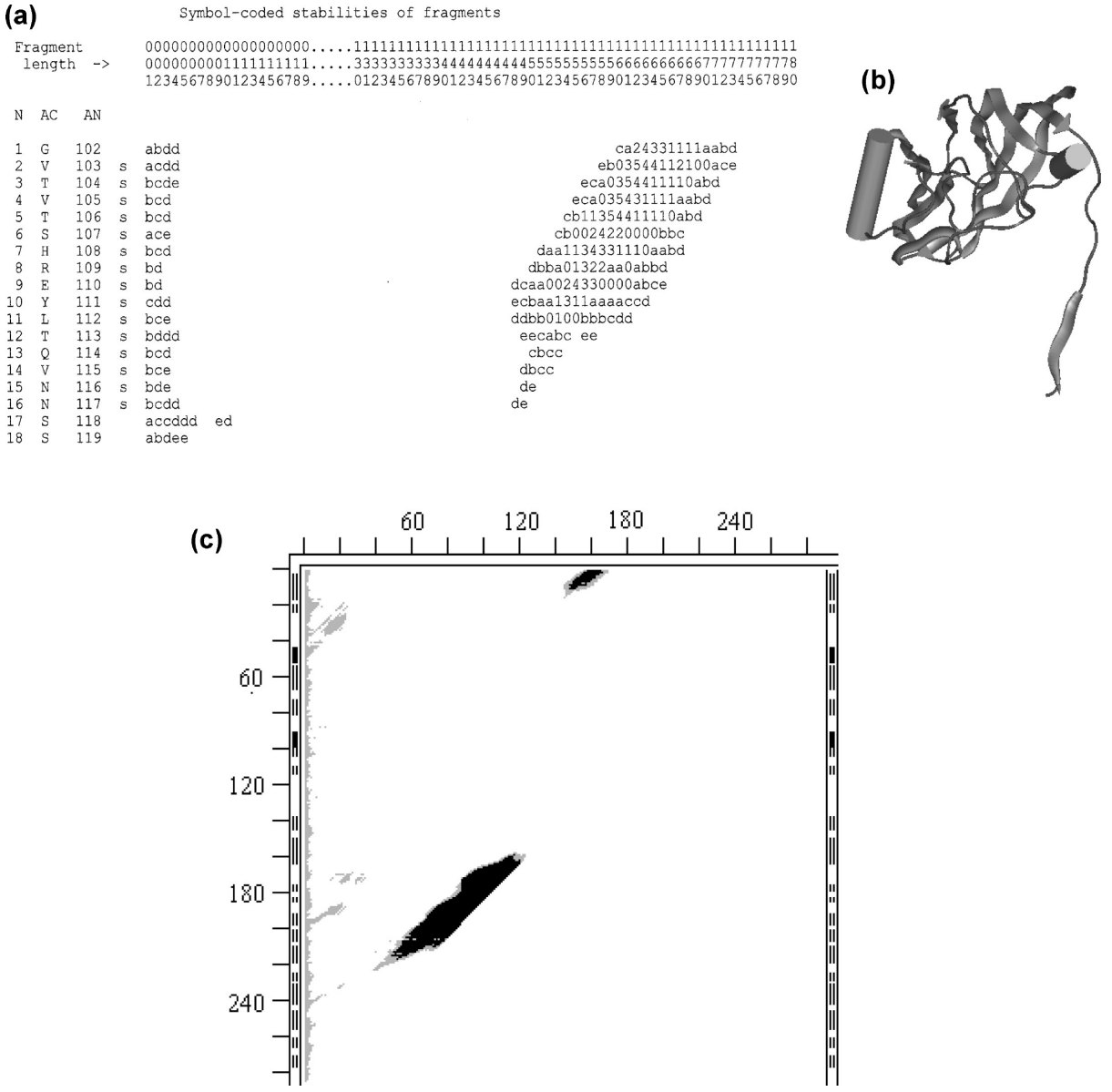

FIGURE 3: SMs (stability map/matrix) and secondary structure representation for 2tbv (Tobacco Bushy Stunt Virus capsid protein). (a) Stability map (SM) for 180 N-terminal residues of subunit A of 2tbv (large parts of SM that do not contain stable fragments have been omitted for the sake of compactness). Locations of continuous fragments in the chain are uniquely defined by two perpendicular axes: fragment's length (L) (horizontal) and position (N) of the fragment's first residue relative to the N-terminus of the entire chain or molecule (vertical on extreme left). The matrix elements of SM are fragments' stabilities (rounded to the nearest integer) represented by one-character symbols. For example, 1, 2, ..., 9 "−" denote "stable" ΔG values of −1, −2, ..., −9, and −10 kcal/mol, respectively, Uppercase letters A, B, C, D, etc., denote stable ΔG values of −11, −12, −13, −14, etc., kcal/mol, respectively; 0 denotes a ΔG of 0 kcal/mol (the fragment spends half its time in its X-ray fold and half in an unfolded state). Lowercase letters a, b, c, d, e, etc., denote positive ΔG values of 1, 2, 3, 4, 5, etc., kcal/mol, respectively. White spaces denote instabilities above the cutoff value (here, above 5 kcal/mol); horizontal coordinate axis "fragment length" (L) is marked by columns of three numbers (to be read vertically to save on SM sizes). Dots in the L-axis markings indicate that a vertical strip of the full SM is omitted (there are no stable fragments of omitted lengths); three columns to the right of the N-axis markings are AC (one-letter residue code), AN (residue numbering of the chain accepted in the PDB entry), and the last one shows whether the residue in the X-ray structure is in β-strand (s), helix (H), or coil (white space). (b) Ribbon diagram of the X-ray structure of the same 180 residues [*MOLE* and manual available from G. P. Privalov (gpriv@axonx.com); wire molecular images in this paper were also made with *MOLE*]. (c) Bitmap SM for all continuous fragments in 2tbvA. The ticks (distance between them is 20 residues) on the left of the SM are the same as in Figure 1 and mark the position of the fragment's beginning (N-terminus) in the entire sequence; ticks on the top of the SM measure the length of the fragment from its N-terminus. Black areas in the SM correspond to negative ΔG values (stable); gray areas denote marginal stabilities (0−5 kcal/mol), and white areas indicate instabilities with a cutoff ΔG of > 5 kcal/mol (larger cutoffs are sometimes used for more reliable estimates in the text and Table 3).

estimate of the error in the calculated stabilities due to coordinate uncertainties of all atoms in a set of independently determined structures of the same molecule. We selected from the set of more than 30 ribonuclease A structures 7, which were classified as "monomeric" by PQS (*41*), to exclude possible side chain distortions caused by extensive contacts in oligomers. The calorimetrically measured stability of ribonuclease A is 10 kcal/mol (*42*). Calculated stability values for the seven selected structures varied from −6 to −12 kcal/mol, with an average of −9.7 kcal/mol and a variance of 2.8 kcal/mol. Calculated differences between the stability of the entire protein and the protein without the S-peptide (residues 1−20) varied between −4 and −6 kcal/mol, with an average of −5.6 kcal/mol and a variance of 0.8 kcal mol. The experimental stability of this shortened ribonuclease decreases by

Table 1: Pairs of Protein Structures, DDMs, Their Characteristics, and Types of Movements

pair	PDB pair	residue included	Biomol[d]	unit cell[e]	RMSDD (Å)	Δ (%)	δ (%)	maximum move	rigid motions	RMSDD/Δ final (Å/%)	type
1	1cll1ctr	138[c]	m		12.83	54.92	4.08	46.7	5	0.44/2.70	tweezers
2	4cd21cd2	202	m		1.25		2.62	11.7	11	0.44/4.74	tweezers
3	4ake1ank	214	m	2	6.45	59.62	16.75	24.8	>12		glove/tweezers
4	1akz1ssp	223	m/tr		1.14	21.89	3.19	5.5	11	0.31/0.61	tweezers
4′									7	0.45/4.07	
5	2tbv2tbv[a]	287	v. shell	3	1.37	32.22	2.67	6.5	3	0.27/1.11	hinge
5′									2	0.43/3.35	
6	1l3f3tmn	316	m/tetr		0.46	3.35	0.38	1.8	2	0.36/0.69	pliers
7	4ape5er2	330	m/d		0.45	5.21	0.54	2.7	2	0.40/2.47	pliers
8	6ldh1ldm	329[b]	tetr		1.25	15.37	4.53	12.3	10	0.41/2.86	tweezers
8′									7	0.45/4.14	
8″[g]									8	0.43/3.70	
9	2gd11gd1	334	tetr		0.49	6.14	0.74	3.1	2	0.36/1.89	pliers
10	1erk2erk	353	m		2.24	27.88	3.81	13.8[f]	>12		glove/tweezers
11	8adh6adh	374[b]	d	1/2	1.05	21.90	2.89	5.7	12	0.45/2.61	tweezers
12	9aat1ama	401	d	2/1	1.20	22.96	7.48	6.7	7	0.36/2.52	tweezers
12′									3	0.43/4.12	
13	16pk13pk	415[b]	m/d	1/4	3.07	45.61	4.39	13.3	>12		glove/tweezers
14	1v4t1v4s	424	m		6.90	49.24	5.66	38.9	>12		glove/tweezers
15	4cts1cts	437	d	2	1.58	26.30	12.20	9.6	>12		glove/tweezers
16	1hrd1bgv	449	h	4	1.89	32.85	1.15	8.5	5	0.22/0.01	pliers
16′									2	0.35/1.84	
17	1lfh1lfg	691[b]	m		4.77	41.43	1.51	24.9	11	0.37/2.00	tweezers
17′									7	0.45/3.19	

[a]There are three molecules in the 2tbv unit cell; two of them (A and B) do not have 102 residues or C^{α} atoms on the N-terminus, and the third (C) molecule is missing only 66 residues. We used A and C (shortened) molecules with the same number of resolved residues in the calculation. [b]For entry 13, three N-terminal and two C-terminal residues, which dramatically change their conformation, were not included in the fitting and RMSDD evaluation; such unincluded residues can be denoted by −3N, −2C for entry 13 and similarly indicate unincluded residues in other structures: −4N, −1C for entry 17; −4N, −7C for entry 17′; −5C for entry 8; −4N for entry 11. [c]To make molecules comparable, residues 1−3 and 148, missing in 1cll, and residues 76−80, missing in 1ctr, are not included in the DDM comparison. [d]Likely but not fully reliable oligomerization state of the biologically active molecule, estimates from PQS. Notations in the table: m, monomeric; d, dimeric; tr, trimeric; tetr, tetrameric; h, hexameric. [e]Number of molecules in the crystallographic unit cell. An empty cell means 1; 1/2 means that the first PDB entry (e.g., 8adh) has one molecule in the unit cell and the second (e.g., 6adh) has two molecules in the unit cell. [f]Nine N-terminal and three C-terminal residues move very far in a nonrigid manner; the value is given for the maximum DD closer to the middle of the chain. [g]Characteristics of the transformation identical to 8′ except dividing fragment 96−104 into rigid fragments 96−100 and 101−104.

8.1 kcal/mol, relative to the entire intact protein (*43*). This gives a measure of reliability of the absolute values of the calculated results and of stability differences within the same molecule. Additional similar statistics will be provided elsewhere.

Functional "Motions", Their Evaluation, and Classification. Functional (or protein association-induced) "motions" might or might not involve an actual hinge. A significant motion might arise either from a continuous deformation of a flexible fragment of a chain (like a bent spring) or from a series of separate hinges (e.g., rotations around non-neighboring single bonds). The result of either of these can be an accumulated large motion at a distance, and in this work, we are focusing on such significant accumulated changes large distances from the local deformations, rather than on the local backbone changes. In particular, these large distant motions can result from a series of smaller motions or main chain deformations that yield a large cumulative remote motion [e.g., as occurs in citrate synthase (*3*)].

We retain the name hinge motion for motions enabled by hinges in the main chain but not involving the grasping of a target, e.g., as in molecules of the viral capsid of 2tbv (*6*, *44*). If a motion grabs a target (e.g., substrate) and brings remote protein parts into contact, we shall liken it to the closing of the tips of tweezers and call this a tweezers motion. A motion in which remote protein parts lock onto a target but where their tips do not form a close contact we shall instead call a pliers motion or, more generally, a chopsticks motion (implying more than one fulcrum position). If one functional conformation cannot be transformed into another (within the coordinate uncertainty) by a large (a dozen) set of rigid-body motions of its fragments, we shall call the entire motion a glove tweezers/chopsticks motion. These are the main motion types observed in our Results.

One can easily imagine a rather large motion of contacting surfaces relative to one another that would not involve shear. An example of such a motion might be the motion of a rocking chair (on a flat floor or on a curved surface), and we call this a rocking motion. If the knobs-into-holes arrangement changes, we can liken it instead to a gear motion. Elbow motion had already been introduced in the literature (*45*).

Here we classify "motions" using only the characteristics of DDMs and CDDMs (in one case below we also use a SM). More detailed (or alternative) motion types can be suggested and used as detailed analysis of the conformational changes (motions) in proteins progresses.

Coarse-Grained (elastic) Modeling of Protein Dynamics. Existing approaches often elucidate intermediate conformations in protein motions by using an interpolation procedure (*8*). We found (*25*, *26*) that such a procedure often leads to severe atomic overlaps and thus yields partially erroneous descriptions for pathways of motion that are unlikely to provide a clear picture of protein repacking during a motion. The interpolation method for coarse-grained models relying on the elastic network representation of dynamics (*25*, *26*) appears to avoid most of the severe atomic overlaps exhibited by the coordinate interpolation methods (*8*), yielding a better understanding of protein motion pathways. Here we represent some from the series of 100 stepwise snapshots of the coarse-grained ENM

Table 2: Rigid Fragment Motions in the Transformation of Holo- to Apoprotein Structures

	PDB pair	no. of residues[a]	fragment for initial[b] molecular superposition	fragments moved as individual rigid bodies after initial molecular superposition[c]
1	1cll1ctr	138 m	24−70	74−138, 101−109, 128−138, 3−11
2	4cd21cd2	202 m	190−202	17−21, 39−41, 42−44, 45−47, 48−81, 82−85, 86−115, 116−146, 147−151, 152−162
3	4ake1ank	214 m		more than a dozen, large nonrigid deformation
4	1akz1ssp	223 m	174−223	3−7, 8−35, 36−48, 49−63, 64−76, 77−121, 122−132, 133−148, 149−170, 171−173
4′			174−223	3−7, 8−48, 49−63, 64−76, 122−148, 171−173
5	2tbvA2tbvC	286 v	174−286	2−164, 167−170
5′			174−286	2−164
6	1l3f3tmn	316 m	127−316	1−126
7	4ape5er2	330 m	1−194	195−330
8	6ldh1ldm	329	3−95	96−104, 105−108, 109−121, 122−215, 216−218, 219−223, 224−305, 306−324, 233−236
8′			3−95	96−104, 105−108, 109−121, 216−218, 219−223, 306−324
8″[d]				96−100, 101−104, 105−108, 109−121, 216−218, 219−223, 306−324
9	2gd11gd1	334	160−334	1−159
10	1erk2erk	353 m		more than a dozen, large nonrigid deformation
11	8adh6adh	374	5−54	55−65, 66−96, 97−101, 102−182, 183−292, 293−298, 299−322, 323−354, 355−363, 364−368, 369−374
12	9aat1ama	401	228−319	382−401, 350−381, 320−349, 2−13, 14−33, 34−36 (fragment 37−227 does not move)
12′			47−319	320−401, 13−46
13	16pk13pk	415 m		more than a dozen, large nonrigid deformation
14	1v4t1v4s	424 m		more than a dozen, large nonrigid deformation
15	4cts1cts	437		more than a dozen, large nonrigid deformation
16	1hrd1bgv	449	1−203	204−372, 373−393, 394−431, 432−449
16′			1−203	204−431
17	1lfh1lfg	691 m	5−86	87−92, 93−138, 139−142, 143−250, 251−329, 330−337, 338−417, 418−420, 421−423, 424−690
17′			5−86	87−250, 330−337, 338−417, 418−420, 421−423, 424−684

[a]Monomers are marked with an m after their length. [b]The transformation fitting this fragment is applied to the entire molecule. [c]Each fragment is fit by an individual transformation, not applied to the rest of the molecule. In some lines, the listed fitted fragments do not cover the entire sequence; the gaps indicate fragments that were accurately placed in the initial fitting, and no subsequent individual fitting led to a significant improvement. [d]Transformation identical to that of entry 8′ except dividing fragment 96−104 into rigid fragments 96−100 and 101−104.

simulations of protein movements, superimposed on the reference wire structure of the pair, along with the corresponding DDMs. Comparisons of these two representations allow us to better identify conformational intermediates for protein functional movements.

RESULTS

Rigid and Nonrigid Movements, Their Classification, and Fragment Stabilities. Results of our calculations for 17 function-related protein movements are summarized in Tables 1−3. Four proteins (thermolysin, lactate dehydrogenase, citrate synthase, and adenylate kinase) were chosen because there were previous studies of their structures or stabilities. Others were randomly taken from the PDB by searching for apo−holo, apo−tertiary, or open−closed pairs of structures with a functional relevance having been suggested in the corresponding publications.

Below we represent and comment on the DDMs of some of the "motions" studied, to demonstrate their (not readily obvious from the tables) variety and the DDMs' interpretive power. Additional DDMs are included in the Supporting Information. We describe the pairs of end structures for each "motion" and the results of the rigid-body motions, characterizing the function-related conformational changes, in Table 1. The sequences of these rigid-body motions are listed in Table 2. We also present examples of simplified stability maps (SM), which provide a simple picture of any recurring stability patterns, where details are given in Table 3. Similarly classified conformational changes are grouped together below. In each group, focus is concentrated on the most interesting cases, with the rest mentioned briefly or relegated to the Supporting Information.

(*i*) *Hinge "Movement" with Small Shear* (*viral shell protein, 2tbv, molecules A and C*). The DDM of 2tbvA2tbvC (Figure 1a) from Tomato Bushy Stunt Virus shows two mostly black triangles (2−164 and 174−286), which are likely to be rigid fragments (see Methods). In the first step, we move the entire structure 2tbvC by fitting its rigid fragment 174−286 to the same fragment in 2tbvA as described in Methods. This first move does not change the DDM or RMSDD. After fragment 2−164 had been moved, the RMSDD of structure 2tbvA2tbvC becomes 0.43 Å, and Δ = 3.35%, which are within the coordinate uncertainty threshold. Thus, it is a true simple hinge with an ~3.5 Å translation and 22° rotation, consistent with earlier results (*6*). An additional rigid-body fitting of fragment 167−170 reduces the RMSDD to 0.27 Å and Δ to 1.11% (Table 1). However, this additional rigid-body movement improves the fit already below the threshold of the coordinate uncertainty and, thus, might be unimportant. The total number of rigid-body motions for each protein pair is shown in the second to last column of Table 1. (Entries in Tables 1−3 are arranged in order of increasing protein chain length. Thus, 2tbv is entry 5.) An example of the output of sequential rigid-body fittings is shown in section 2 of the Supporting Information.

Stabilities of continuous fragments of 2tbvA (Table 3) are depicted in Figure 3c. It shows that 2tbv has two stable continuous domains, encompassing most of the chain (the exact locations of

Table 3: Stable Fragments and CATH/SCOP Domains in Pairs of Apo- and Holoprotein Structures

	protein	no. of residues[a]	state	PDB	fragments with stabilities (kcal/mol) and CATH domains
1	calmodulin	138 m	apo	1cll	1−138(>10), 12−70(−2), 101−138(−1), 120−138(−1)
			holo	1ctr	9−70(−1), 75−138(−1), 119−138(−1)
			CATH		1−69, 68−138
2	dehydrofolate reductase	202 m	apo	4cd2	1−202(>10), 54−81(−3), 48−120(0)
			holo	1cd2	1−202(>5), 54−81(0)
			CATH		1−202
3	adenylate kinase	214 m	apo	4ake	1−214(>10), 25−102(−3), 1−107(−1), 111−169(−3), 194−214(−1)
			holo	1ank	1−214(>5), 26−103(−2), 1−109(0), 122−159(0), 196−214(1)
			CATH		1−214
4	DNA-uracil glycosylase	223 m	apo	1akz	1−223(−1), 12−33(0), 85−105(0), 185−200(0)
			holo	1ssp	3−223(−9), 23−210(−5), 145−185(0), 88−109(0), 185−200(0)
			CATH		1−223
5	Tobaco Bushi Stant virus three molecules in unit cell	287 v	molecule A	2tbv	2−162(−5), 171−277(−13)
			molecule C	2tbv	3−161(−6), 172−278(−15)
			CATH		not available
6	thermolysin	316 m	apo	1l3f	1−316(−18), 3−148(−12), 76−316(−16), 136−316(−17), 233−316(−16), 255−316(−6)
			holo	3tmn	1−316(−3), 1−148(−3), 76−316(−6), 136−316(−11), 234−316(−12), 255−316(−2)
			CATH		6−154, 155−315
7	endothiopepsin	330 m	apo	4ape	1−330(>10), 1−173(−2), 18−125(−12), 18−148(−10), 183−330(−2), 197−313(−11)
			holo	5er2	1−330(>10), 1−173(−5), 18−129(−14), 18−149(−13), 183−330(−2), 197−314(−11)
			CATH		2−170, 171−326
8	lactate dehydrogenase	329	apo	6ldh	1−329(>10), 21−96(−10), 109−162(−5), 21−162(0), 163−236(−3), 255−303(−1)
			holo	1ldm	1−329(>5), 21−96(−9), 109−162(−5), 165−235(−6), 252−301(−1)
			CATH		1−162, 163−329
9	glyceraldehyde dehydrogenase	334	apo	2gd1	1−334(>5), 1−120(−7), 1−96(−10), 93−150(−2), 236−317(−1)
			holo	1gd1	1−334(>5), 1−120(−9), 1−96(−9), 95−145(−3), 236−317(1)
			CATH		1−147, 314−332
10	signal-regulated kinase	353 m	apo	1erk	1−353(>10), 1−50(−2), 101−318(−2), 101−233(−3), 172−294(−5)
			holo	2erk	1−353(>5), 3−50(−4), 101−318(2), 122−214(−2), 177−295(−4)
			CATH		1−100, 330−353
11	alcohol dehydrogenase	374	apo	8adh	6−364(−3), 178−294(−15), 192−293(−15), 172−322(−4)
			holo	6adh	6−364(−3), 178−293(−13), 193−293(−13), 172−322(−4), 261−317(−1)
			CATH		1−178, 318−374
12	Asp aminotransferase	401	apo	9aat	1−401(>5), 68−304(−5), 80−102(−2), 321−401(−14)
			holo	1ama	1−401(>5), 68−304(−9), 80−102(−2), 321−401(−13)
			CATH		13−46, 47−319, 320−401
13	phosphoglycerate kinase	415 m	apo	16pk	1−415(3), 3−192(−8), 194−404(−18)
			holo	13pk	2−413(−26), 3−193(−23), 195−414(−21)
			CATH		5−192, 199−406
14	glukokinase	424 m	apo	1v4t	1−424(>10), 40−58(−1), 110−130(0), 270−379(−6), 255−341(−10), 271−391(0)
			holo	1v4s	1−424(>10), 270−379(−7), 261−338(−13), 338−404(−1)
			SCOP[b]		1−181, 182−426
15	citrate synthase (dimer)	437	apo	4cts	51−417(−5), 275−381(−4)
			holo	1cts	51−417(−2), 277−381(−6)
			CATH		1−276+392−423, 277−391
16	Glu dehydrogenase	449	apo	1hrd	1−449(>5), 1−53(0), 54−194(−9), 206−376(−17), 227−321(−21), 410−434(−5)
			holo	1bgv	1−449(>5), 1−53(−2), 54−195(−10), 206−376(−13), 226−321(−18), 410−433(−4)
			CATH		52−187, 207−373, 1−51+425−449
17	lactoferrin	691 m	apo	1lfh	37−74(−2), 91−256(−14), 109−210(−20), 376−408(−1), 435−594(−9), 453−546(−10)
			holo	1lfg	37−74(0), 91−256(−13), 109−210(−16), 376−408(0), 435−594(−10), 453−546(−11) +S−S, −9.2, −5.9, −3.4, −9.2, −6.2
			CATH		1−91+251−339, 92−250, 340−434+595−691, 435−594

[a]Monomers are marked with an m after their length. [b]CATH unavailable. For the rest, see Table 1 and the figure legends.

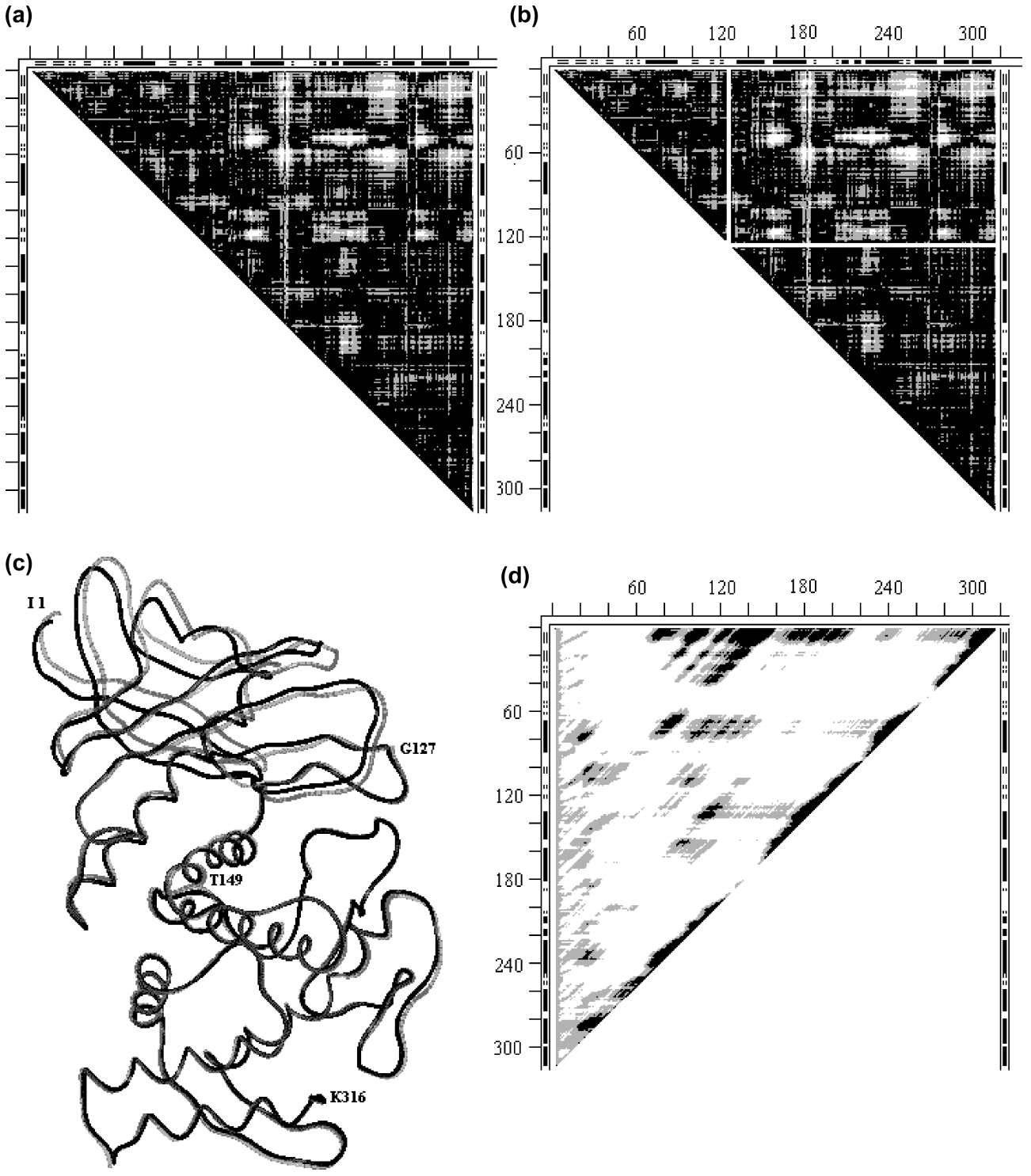

FIGURE 4: DDM, wire superposition, and SM for thermolysin. (a) DDM 1l3f3tmn. For notations, see Figure 1. Helices and β-strands are marked in three DDM's borders; ticks on the left and top are at intervals of 20 residues, and some residue numbers are provided alongside the ticks when helpful. (b) A white horizontal line is drawn in the same DDM right under all white spots; intersecting this line at the diagonal a white vertical line is drawn up. These two white lines form borders of two mostly black (no white spots) triangles which might change their position as rigid bodies in function-related conformational transitions. (c) Wire superposition of 1l3f (black) and 3tmn (gray) by fragment 127−316. (d) SM for 1l3f. For notations, see Figure 3c.

the most stable fragments are listed in Table 3). However, these two domains do not join in a single stable structure (that would be indicated by a black patch in the top right corner of the SM), suggesting that in a monomer they move independently, as might be required in a virus shell structural protein.

(*ii*) *Pliers/Chopsticks "Movements"* (*entries 6, 7, 9, and 16 in Tables 1−3*). DDM 1l3f3tmn for thermolysin apo to holo "motion" is shown in Figure 4a. The movement is a simple pliers motion because the substrate is grabbed without the "jaws" of the active site closing upon each other, forming a new contact, and because the δ of the CDDM does not indicate any shear motion.

The location of nearly black triangles in the DDM is shown in Figure 4b. We draw a horizontal white line just below the cluster of white spots in the DDM and then a vertical white line from the

intersection of the horizontal line with the DDM's diagonal. This yields two mostly black triangles behaving as rigid fragments, 1−126 and 127−316 (in agreement with ref *46*), in the function-related conformational transition ("movement"). We superimpose fragment 127−316 and apply the transformation to the entire holo structure; then we fit fragment 1−126. After these two transformations, the RMSDD decreases to 0.36 Å and Δ to 0.73%. The translation of the center of mass of fragment 1−126 is 1.3 Å, and rotation around the axis passing through this center is 6° (almost the same as the value of 5° reported in ref *46*). It is clear from Figure 4c that the white spots in the DDM represent not local distortions but a systematic, albeit not large, difference between the apo and holo structures. We could instead draw a vertical white line to the left of the cluster of white spots in the DDM of Figure 4a and then draw the horizontal line. This would lead us to the mostly black triangles for domains comprised of residues 1−148 and 149−316 (not shown). The first fragment also corresponds to the stable domain (Table 3). Using this fragment for the initial fitting followed by the fitting of the second fragment leads to very similar results, possibly indicating a nonuniqueness of the pathway for this functional conformational change.

In contrast to 2tbv, 1l3f has not only stable domains (agreeing reasonably well with the CATH assignment) but an entire highly stable structure, including all residues 1−316 (see the large black spot in the top right corner of Figure 4d and Table 3). Interestingly, a comparison of the stabilities of the fragments of apo- and holothermolysin (entry 6 in Table 3) suggests that the stabilization from the protein chain alone significantly decreases upon substrate binding, this loss possibly being traded for the stabilization provided by the protein−substrate interactions and perhaps by the accompanying oligomerization suggested by PQS (*41*).

DDM 4ape5er2 (endothiopepsin) is shown in Figure 1a; DDM 1hrd1bgv (glutamate dehydrogenase) and rigid-body movements of its fragments were discussed in detail in ref *10*. The DDM of 1hrd1bgv (Figure 12a′) is similar to that for 2tbv (Figure 1a), with three vertical white strips added (for details, see ref *10* and Tables 1−3). The DDM for 2gd11gd1 (glyceraldehyde dehydrogenase) can be found in Figure S1 of the Supporting Information.

These three proteins have stable domains but no stability for the entire molecule. Details of calculated fragment stabilities can be found in Table 3. Locations of predicted stable domains are in reasonable agreement with CATH. Sequences of rigid fragment movements transforming each holo structure into the corresponding apo structure within coordinate uncertainty limits are listed in Table 2.

DDMs of 4ape5er2 (Figure 1b) and 2gd11gd1 (Figure S1) are quite similar in their general appearance to DDM 1l3f3tmn (Figure 3a) and require as simple rigid-body transformation. They all have low RMSDD and Δ values (0.46 and 3.35, 0.49 and 6.16, and 0.45 and 5.41, respectively) just above the coordinate uncertainty thresholds (< 0.46 and < 5, respectively). PQS suggests that 5er2 undergoes dimerization, 2gd1 and 2gd2 are tetramers, while 1hrd and 1bgv are suggested to be hexamers. There are some differences (*10*) for 1hrd1bgv between the rigidly moving fragments (Table 2) and domains (Table 3).

(*iii*) *Tweezers Movements* (*entries 1, 2, 4, 8, 11, 12, and 17 in Tables 1−3*). The DDM for calmodulin (1cll1ctr) (Figure S2a of the Supporting Information) resembles the DDM of 2tbv (Figure 1a). However, its CDDM (Figure S2b of the Supporting Information) shows formation of new long-range contacts

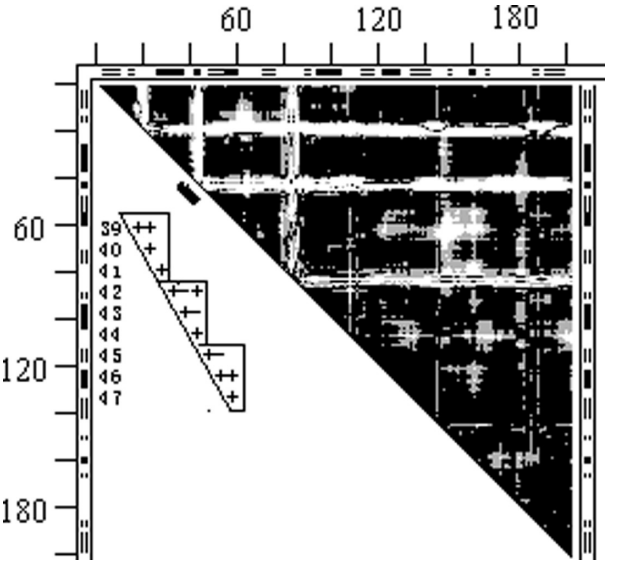

FIGURE 5: DDM 4cd21cd2 (dehydrofolate reductase). For notations, see the legend of Figure 1. The inset shows fragment 39−47 of the same but larger symbol coded DDM with plus for black, minus for gray, and space for white. This fragment is marked with a short thick black line along the part of the diagonal of the bitmap DDM occupied by the corner of the second L-shaped band.

(enclosed in small rectangles), which leads us to classify this movement as tweezers, more shear than 2tbv, and the largest RMSDD despite being the smallest structure (Table 1). Stabilities of calmodulin fragments are marginal (as well as of the entire molecule), and their localization is in a reasonable agreement with CATH (see Table 3). After the first two rigid-body fittings (Table 2), the RMSDD decreases from its initial value of 12.83 to 1.17 Å, and after further fittings, it decreases to 0.85 Å. However, residues 1 and 2 at the N-terminus, residue 138 at the C-terminus, and residues 71−73, around the fragment without a reported structure, move incoherently (they are not parts of any black triangles indicating a rigid block). After their deletion from the structures that were compared, we obtained an RMSDD of 0.44 Å and a Δ of 2.7%. Note that N- and C-terminal residues often stick out and away from the rest of the monomeric structure and do not contribute to its stability, as well as the visible ends of some loops that are not visible in the crystal. We usually delete these in our comparisons.

DDM of the tweezers 4cd21cd2 [dehydrofolate reductase (Figure 5)] presents a black-and-white pattern different from those of all previous DDMs. Instead of large white spots on a black background, it shows a series of rather wide (up to 10 residues) L-shaped bands with their right angle bends touching the diagonal. These diagonal-touching areas have twisted borders often requiring more than one small rigid body to be fitted as is shown by three triangles in the insert from a larger DDM (see Figure 5). Note that there are a few white spaces within these triangles. However, the rigid-body fitting of each fragment corresponding to these triangles yields an rmsd of < 0.4 Å, suggesting that one should check how bad the effects of a few white DDs in a triangle are for the fitting. Except for these L-shaped bands and a few white spots or bulges on the bands' edges, the DDM is rather black. Results of the series of fittings are listed in Table 1 and their sequences in Table 2. This tweezers motion also has some shear. Its entire structure is not stable, but it has a marginally stable fragment. There is no agreement with the CATH domain assignments.

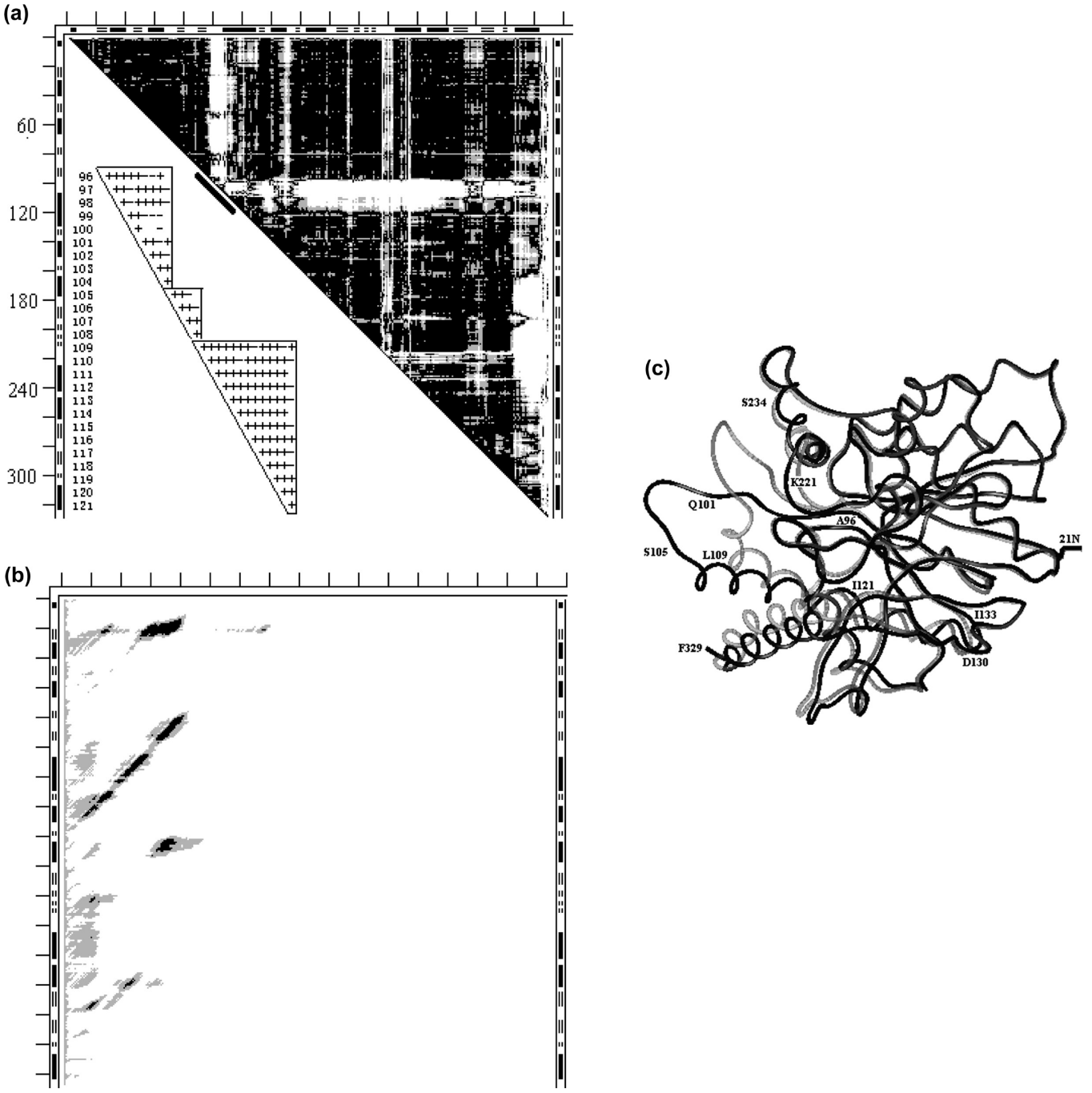

FIGURE 6: DDM, SM, and wire superposition for lactate dehydrogenase. (a) DDM 6ldh1ldm (for notations, see the legend of Figure 1). The inset shows fragment 96−121 of the same but larger symbol coded DDM with plus for black, minus for gray, and space for white. This fragment is marked with a short thick black line along the part of the diagonal of the bitmap DDM occupied by the corner of the largest L-shaped band. (b) SM 6ldh (for notations, see the legend of Figure 3c). (c) Wire superposition of 6ldh and 1ldm by fragment 3−95. Residue numbers directly refer to the black wire (6ldh) and only approximately to the nearby gray wire (1ldm).

The DDM of tweezers 1akz1ssp [DNA-uracil glycosylase (Figure S3 of the Supporting Information)] appears like an enhanced version of the pliers DDM 2gd11gd1 in Figure S1. The number of white spots is greater, and they are larger in size. There is also one relatively narrow L-shaped band. The pair allows a rigid-body fitting within coordinate uncertainty (Tables 1 and 2). The molecule has a marginally stable whole apo structure, with the stability significantly increasing with substrate binding. The single stable domain agrees with CATH (Table 3). Binding, according to PQS, is accompanied by trimerization.

The next tweezers DDM represents a milestone in the studies of protein functional motions (*4*) where it was suggested that a hinge at residues 96−100, another hinge at residues 105−110, and a kink around residue 119 produce one large motion in lactate dehydrogenase.

The DDM 6ldh1ldm [lactate dehydrogenase (Figure 6a)] transparently translates these suggestions into the terms of our methodology, allowing some improvements. This DDM resembles the DDM of 4cd21cd2 (Figure 5) and is characterized by similar large numbers (7−10) of rigid-body transformations and the maximum movement magnitudes (Table 1).

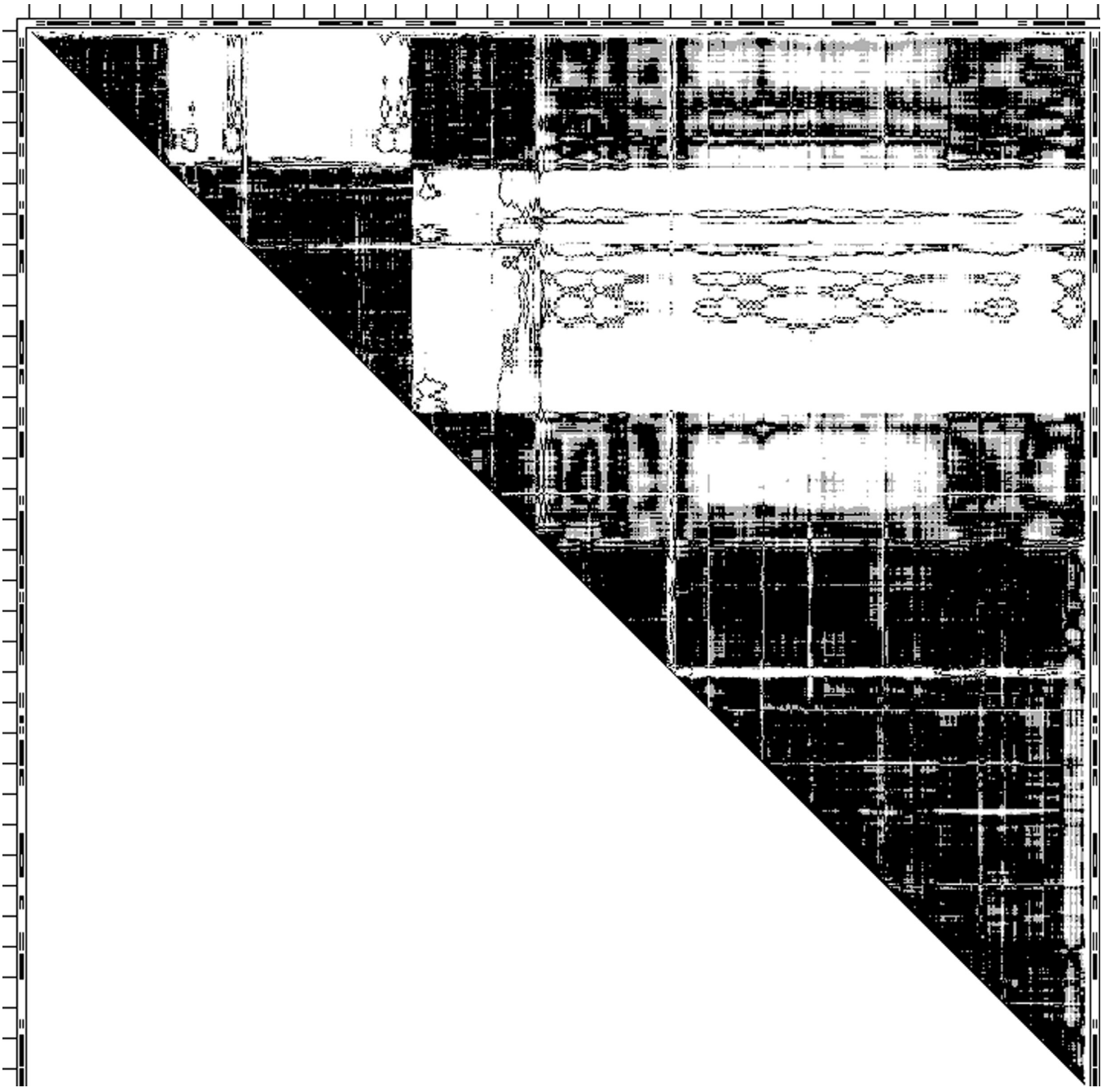

FIGURE 7: DDM 1lfh1lfg (lactoferrin). For notations, see the legend of Figure 1.

The five C-terminal residues do not form a rigid fragment (no black triangle at the bottom of the diagonal). If we exclude them from the DDM, then its RMSDD decreases from 1.35 Å (Table 1) to 1.25 Å.

Three major white areas in Figure 6a are two L-shaped bands, the top one of them being rather thick, and a ragged vertical band on the right edge of the DDM. The hinges and the kink described by Gerstein and Chothia (*4*) can be seen as two small overlapping black triangles at the corner of the L-shaped band and one more right under it (the top small triangle has a short intrusive white line, visually interrupting it). The first four rigid motions (including the initial one) reduce the RMSDD to 0.64 Å. This constitutes a major part of the decrease in the RMSDD toward the threshold of coordinate uncertainty. The following rigid transformations (Table 2, entries 8 and 8′) yield the RMSDD and Δ for the entire molecule (Table 1, entries 8 and 8′) within the coordinate uncertainty threshold. However, the fit of fragment 96−104 (an important part of the active site "lid") is rather poor with an rmsd of 0.99 Å. The center of mass of this fragment translates 5.7 Å, and then the fragment rotates by 36°. The inset in Figure 6a shows details of the DDM for fragment 96−121. It shows that the line for residue 100 contains three white spaces out of five DDs in the triangle 96−104. This suggests dividing it into two triangles, 96−100 and 101−104, which have no white DDs at all. The rmsds for these two fragments decrease to 0.79 and 0.36 Å, with center of mass translations of 2.5 and 9.8 Å and rotations of 21° and 60°, respectively. Analysis of Figure 6c shows that while black loop 101−104 is rather flat, its gray counterpart shows a significant upward curvature. Transforming fragment 96−104 as one (the loop assignment in Table 1 of ref *4*) or two (96−100 and 101−104) fragments, reflecting this curvature, leads to a 1.2 Å difference (of the total of 11.9 Å) in the change of

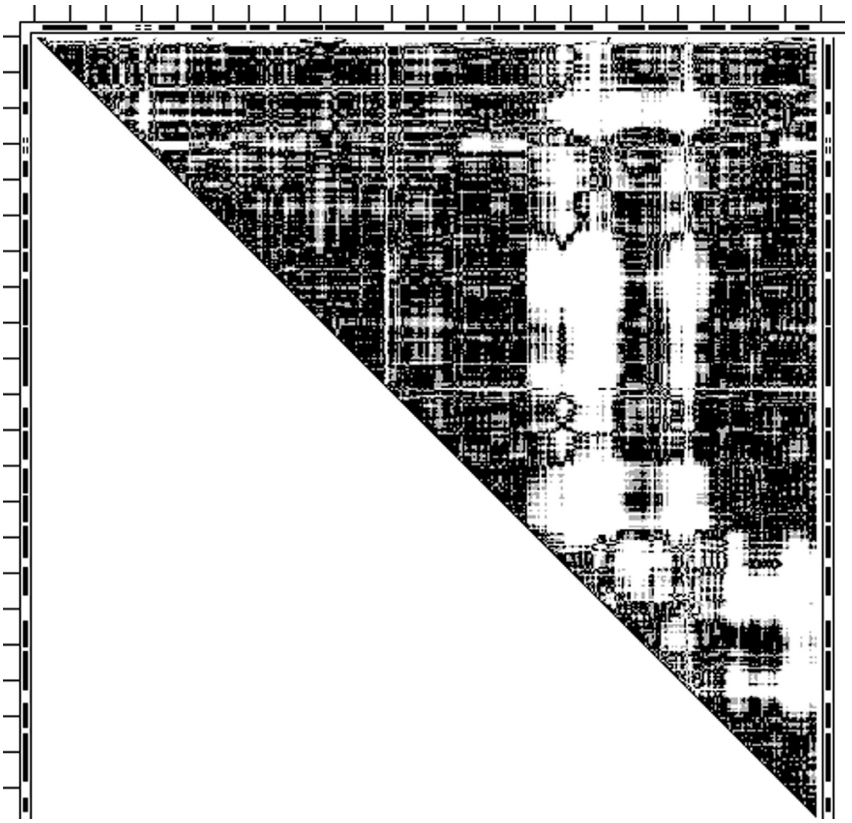

FIGURE 8: DDM 4cts1cts (citrate synthase). For notations, see the legend of Figure 1.

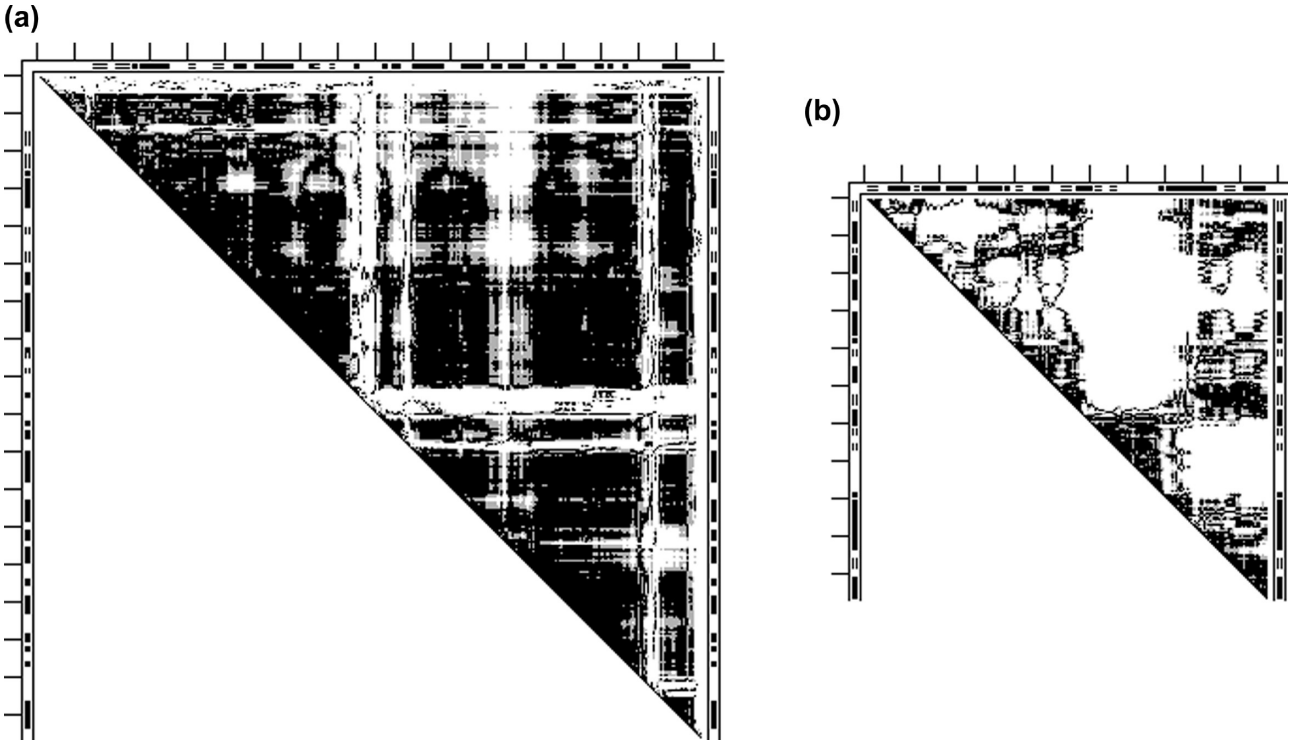

FIGURE 9: DDMs for two kinases: (a) 1erk2erk (signal-regulated kinase) and (b) 4ake1ank (adenylate kinase). For notations, see the legend of Figure 1.

the position of atom C^{α}_{103} in the 6ldh to 1ldm conformational transition.

A comparison of the DDM in Figure 6a and the residue numbering along the wire structure in Figure 6c reveals that all movements of the fragments indicated in the DDM are coordinated with the "movement" of loop 96−121 through direct interactions of these fragments with this "moving" loop.

6ldh and 1ldm do not possess the stability of the whole molecule (Figure 6b), which might be unimportant if the biological molecule is a tetramer, as suggested by PQS (see also

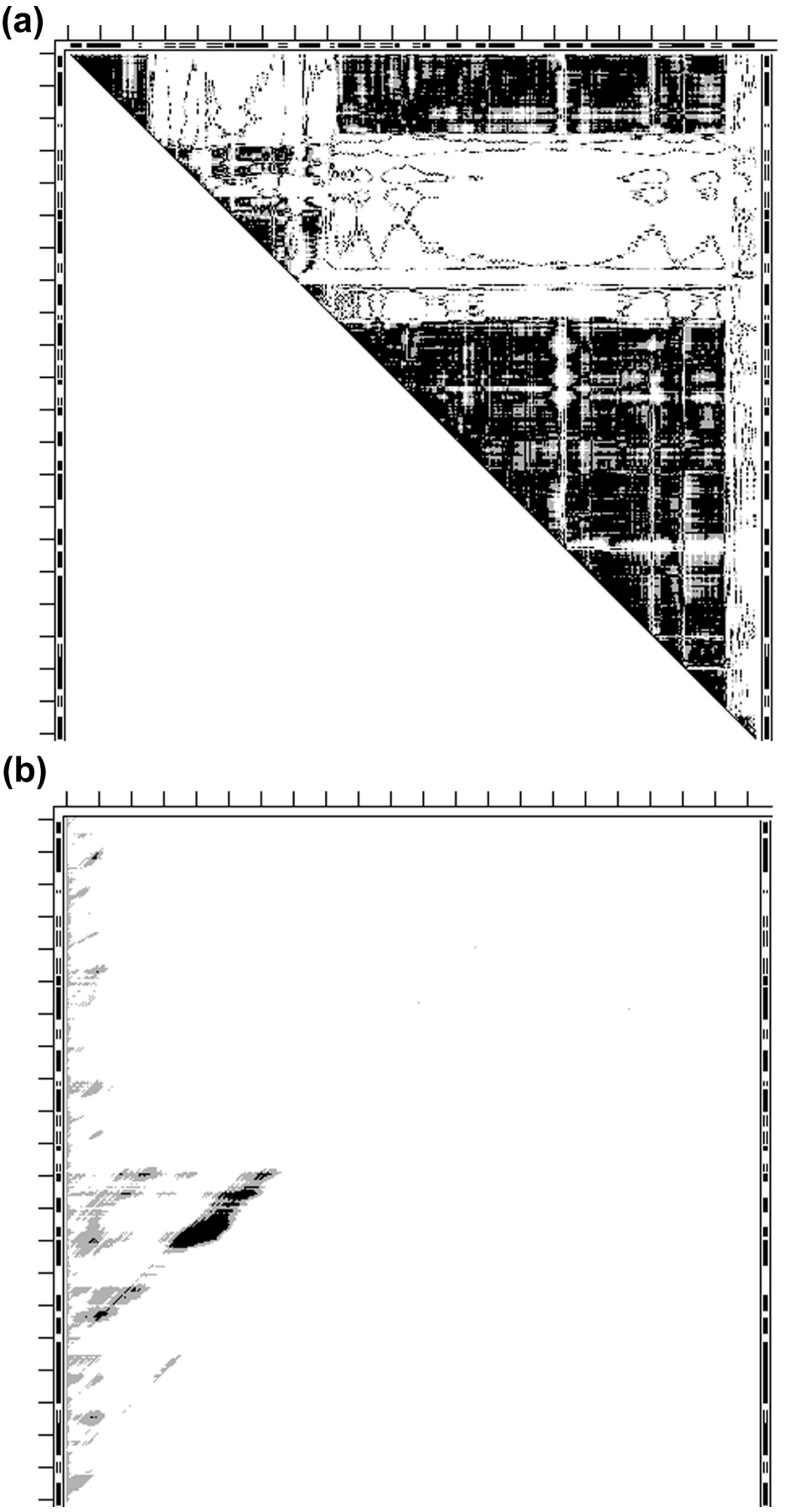

FIGURE 10: Glucokinase. (a) DDM 1v4t1v4s (for notations, see the legend of Figure 1). (b) SM 1v4t (for notations, see the legend of Figure 3c).

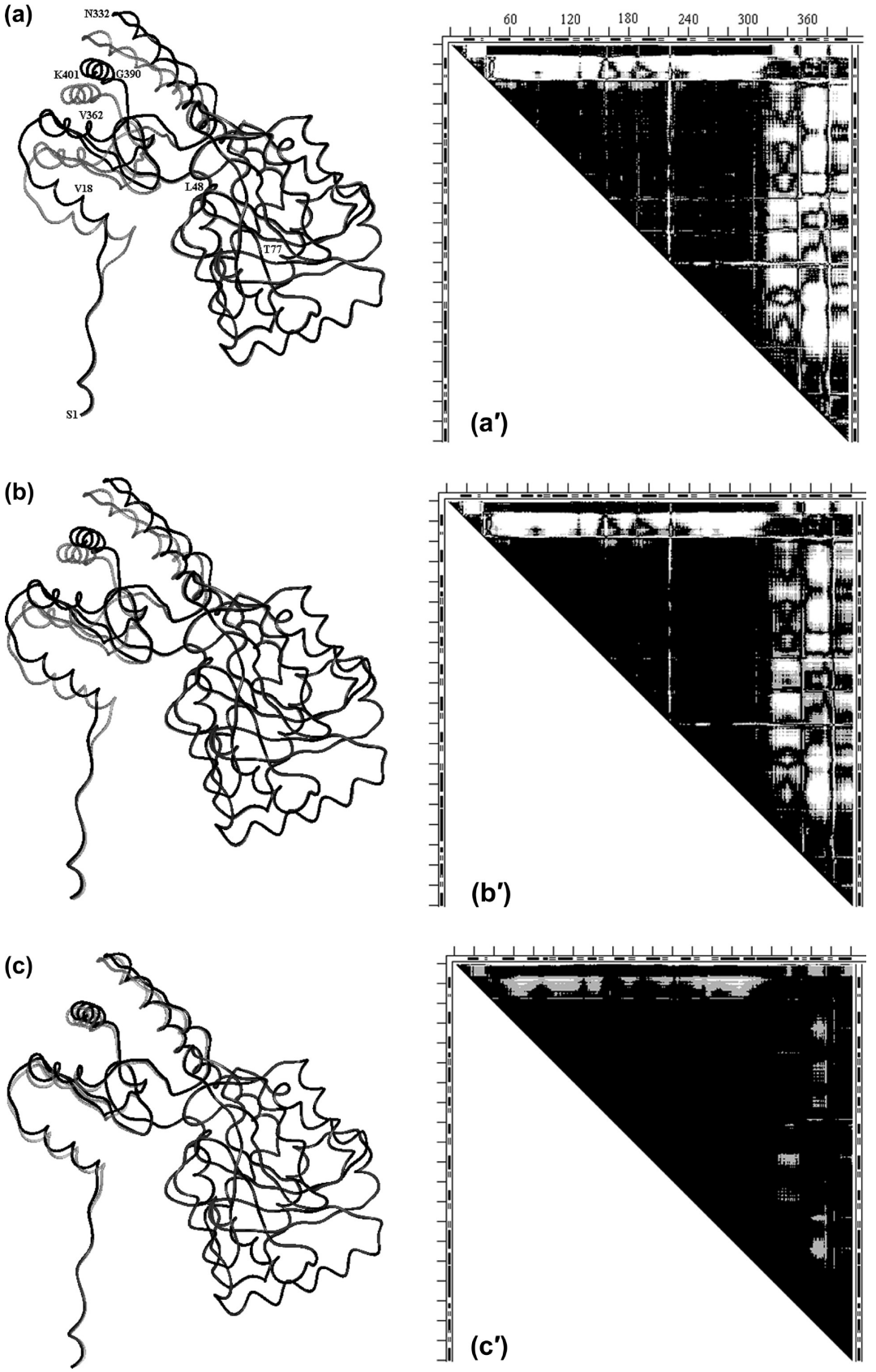

FIGURE 11: (a−c) Three snapshots (numbers 1, 40, and 80 of 100 total) from the elastic coarse-grained dynamics movie of the motion pathway of Asp aminotransferase for each of conformations of 1ama (gray) changing into 9aat (black), which is completed in snapshot 100. In all snapshots, rigid fragments 228−305 are superimposed. Wire rendering is done with *MOLE* (G. P. Privalov). Residue numbers (panel a only) directly refer to the black wire (9aat) and only approximately to the nearby gray wire (1ama). (a′−c′) Three DDMs 9aat1ama where the conformation of 1ama is taken to correspond to the three snapshots (1, 40, and 80) from panels a−c. DDM for snapshot 100 would be completely black because 1aam would completely change its conformation to that of 9aat.

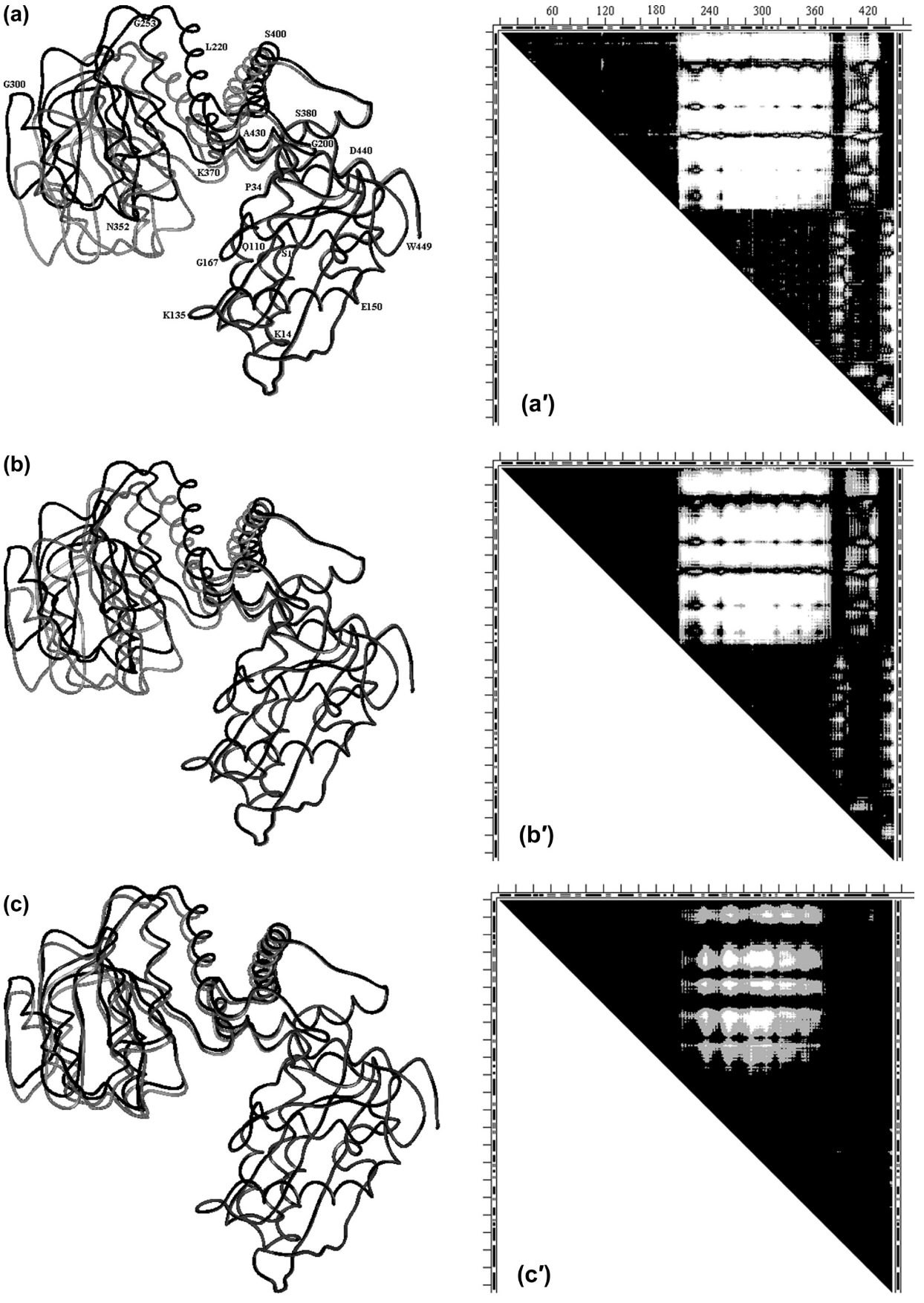

FIGURE 12: (a−c) Three snapshots (numbers 1, 40, and 80 of 100 total) from the elastic coarse-grained dynamics movie of the motion pathway of Glu dehydrogenase for each conformation of 1bgv (gray) changing into 1hrd (black) which is completed in snapshot 100. In all snapshots, rigid fragments 1−204 are superimposed. Residue numbers (panel a only) directly refer to the black wire (1hrd) and only approximately to the nearby gray wire (1bgv). (a′−c′) Three DDMs 1hrd1bgv where the conformation of 1bgv is taken to correspond to the three snapshots (1, 40, and 80) from panels a−c. DDM for snapshot 100 would be completely black because 1aam would completely change its conformation to that of 1hrd.

Discussion). However, both structures exhibit four stable fragments (black areas in Figure 6b). Table 3 lists the locations of only the most stable fragments. Rigidly moving fragments (Table 2) are not stable. This occurs in a majority of proteins. Note that two CATH domains lump together two stable fragments each (Table 3).

The DDMs of 8adh6adh (Figure S4 of the Supporting Information) and 9aat1ama (Figure 11a′) have somewhat similar vertical stripes on their right sides, and Figure S4 also resembles Figure 12a′, which has an upper group of white spots coalesced into a solid rectangle. 6adh can be transformed by a series of rigid-body movements into 8adh (Tables 1 and 2). PQS suggests that 8adh and 6adh are dimeric biomolecules (Table 1). 8adh and 6adh are both moderately stable as a whole and also have stable domains. There is no agreement between the locations of the stable fragments and the CATH domain assignments (Table 3).

The conformational transition 9aat1ama was described in detail in ref *10* (see also Tables 1 and 2). PQS suggests that both 9aat and 1ama are biological dimers (Table 1). Neither 9aat nor 1ama has an overall stable fold, but each has two stable domains that agree reasonably well with two of the three CATH domain assignments (Table 3).

The largest structural pair we have studied is 1lfh1lfg, and it also undergoes tweezers motion (Figure 7). The sequences of moving rigid fragments, fitting structures 1lfh and 1lfg of lactoferrin, are listed in Table 2. Removing the four N-terminal residues and one C-terminal residue from the final DDM leads to a DDM having characteristics within the coordinate uncertainty threshold (Table 1). A less accurate fitting and the removal of the seven C-terminal residues lead to fitting within the threshold but close to its limit.

This large protein does not exhibit stability for the entire folded chain even with the large number of stable fragments additionally stabilized by its 16 S−S bonds (Table 3). Two continuous CATH domains agree well with two of the stable fragments, while the two discontinuous CATH domains do not (Table 3). PQS identifies lactoferrin as monomeric.

An important lesson from DDM 1lfh1lfg fragment 90−250 is that there is no principal difference between L-shaped bands produced by the movements of loops and large fragments in the middle of the chain: the difference is just in the thickness of the band (in 1lfh1lfg, the triangular corner of the L-band happens to be rigid, but this is not proven to be a general rule).

(*iv*) *Glove Movement* (*entry 15 in Tables 1−3*). The first glove DDM 4cts1cts for citrate synthase, analyzed below, also represents a milestone in the studies of protein functional motions (*3*) that has been repeatedly described (*3*, *6*, *47*, *48*). It was suggested that the functional movement in citrate synthase involves small shifts and rotations of five packed helices (shear) in the small domain (residues 272−387) accumulating into a 10 Å shift and 28° rotation relative to the rest of the protein. It was also suggested that the large interface in the biological dimer precludes having simple hinge motions in this protein.

White areas in the DDM of Figure 8 indicate significant conformational changes within the small domain as well as its "movement" relative to the rest of the structure. Twelve rigid-body movements in the small domain alone do not reduce its RMSDD below 0.54 Å or Δ below 6.33% because deformations of helices and loops result in relatively short fragments with poorly retained geometry. The RMSDD of 0.61 Å was obtained for the untransformed large domain (with the small domain and the five N-terminal residues excised from the structure). Thus, even the "large domain" undergoes significant conformational changes beyond the coordinate uncertainty. All attempts to fit the entire holo form to the apo form by rigid-body movements fail to lead to an RMSDD below 0.67 Å, meaning that we have to classify the citrate synthase conformational change as a glove motion with large shear and tweezers closure.

The stabilities of fragments of unliganded (4cts) and liganded (1cts) citrate synthase are listed in Table 3. Small domains formed by five helices (residues 274−386) are stable in both 4cts (−4 kcal/mol) and 1cts (−6 kcal/mol) as well as almost the entire structure, including residues 51−415, of residues 1−437. Calculated stabilities are −5 kcal/mol for 4cts (apo) and −2 kcal/mol for 1cts (holo). Thus, substrate binding is accompanied by the destabilization of the entire (residues 51−405) protein and some stabilization of the small domain. This suggests that the destabilization of the entire protein chain upon this binding can probably be compensated by the binding free energy which has not been included in our calculations. Furthermore, citrate synthase is a biological dimer, and dimerization may provide additional stabilization. Thus, the stability redistribution is somewhat different in citrate synthase compared to thermolysin. However, in both cases, ligand binding seems to destabilize the largest most stable part of the monomer, possibly storing energy for throwing out products.

However, in contrast to thermolysin, we cannot say that stable domains move as rigid bodies in the functional conformational change of citrate synthase. A stable small domain of citrate synthase is not rigid but changes significantly through shear, and the other part, usually termed a large domain, is unstable by itself and also undergoes significant internal deformations (also see Discussion).

(*v*) *Glove Movements in Kinases* (*entries 3, 10, 13, and 14 in Tables 1−3*). In humans, ∼1.7% of all genes encode protein kinases (*49*) and approximately half of them map to disease loci (*49*). Thus, it is of special interest to look at some typical functional motions in a few kinases (including an example of less numerous nucleotide kinases).

The DDM 1erk2erk (Figure 9a) of signal-regulated kinase resembles the DDM 6ldh1ldm (Figure 6a) of lactate dehydrogenase and the DDM of 4cd21cd2 of dehydrofolate reductase (Figure 5), all showing a number of L-shaped bands. The major difference is that the number of such bands is larger in Figure 9a and the corner of the largest of them at the DDM's diagonal does not have small overlapping black triangles responsible for the largest group of rigid-body moves seen in Figure 6a for lactate dehydrogenase. In kinase (Figure 9a), this corner of the white L-band represents a large nonrigid conformational deformation. A similar lack of small black triangles, representing short rigid segments, in other white corners of this DDM adds to the number of parts undergoing nonrigid deformation during the functional movement. This resembles a schematic case in Figure 2a. As a result, we could not find a series of 12 or fewer rigid-body transformations that would convert one functional conformation of this kinase, 2erk, to the other, 1erk (Table 1). Neither 2erk nor 1erk possesses stability for the entire structure, but both have three moderately stable fragments. Their positions in the chain do not agree with the CATH domain assignments (Table 3).

The DDM 4ake1ank (Figure 9b) of adenylate kinase exhibits a complex irregular shape of white patches, with many small and only a few larger black triangles along the diagonal. These predominant small black triangles (short rigid segments) resemble the idealized Figure 2b. In some places, there are white areas

along the diagonal, which correspond to nonrigidly deforming chain segments. Both factors make it impossible to transform 1ank into 4ake within the coordinate uncertainty threshold with 12 or fewer rigid-body motions.

Apo structure 4ake shows three marginally stable folding domains (Table 3) that do not include a significant part of the protein chain, and only one of them is present in holo form 1ank. The entire folded protein chain is unstable. The stable fragments in 4ake or 1ank do not agree with CATH domains (Table 3).

The DDM 16pk13pk of phosphoglycerate kinase (Figure S5a of the Supporting Information) shows two rather solid large black triangles suggesting the presence of large rigid blocks. However, the predominantly white area beyond these black triangles has a rather complex structure, with white areas touching the DDM's diagonal somewhat similar to the DDM in Figure 9b. Attempts to transform 13pk into 16pk within coordinate uncertainty by more than a dozen rigid-body movements failed.

The SM of 13pk (Figure S5b of the Supporting Information) shows a very stable entire holo structure, together with very stable N-terminal and C-terminal folding domains formed by approximately half of the protein chain each. However, the stability of the entire bisubstrate 16pk is only marginal (Table 3). PQS suggests that the bisubstrate form is monomeric and the holo form is tetrameric. No PGK from a particular organism has been crystallized in both the apo and holo forms. The location of stable fragments is in good agreement with the CATH assignments.

The DDM 1v4t1v4s of glucokinase (Figure 10a) indicates a highly flexible glove movement as in adenylate kinase. Attempts to transform 1v4s into 1v4t within coordinate uncertainty by more than a dozen rigid-body movements failed. The SM of 1v4t (Figure 10b) shows only a relatively short stably folded fragment (Table 3) in a PQS-predicted monomer, with the remainder of the structure being unstable and thus mobile. For this protein, simple mechanical analogues for a description of the functional "movement" seem to be lacking. A possible fitting analogy might be a "boa movement" (boa is a genus of snakes folding around a target to constrict it). The location of stable fragments does not agree with the SCOP domain assignments.

While the four kinases studied above indicate nonrigid functional movements, this may not be the case for all kinases. Our preliminary results for the small protein guanylate kinase (186 residues) indicate that apo form 1ex6 can be transformed to holo form 1ex7 with 10 rigid-body movements.

Coarse-Grained Modeling of Protein Dynamics and DDMs. It is often quite difficult to clearly identify in a static picture the local differences between two conformations of the same protein even if they are superimposed on the basis of the rmsd fit. Overcoming this difficulty was one of the motives for developing "morph" movies of protein motions: human eyes see movements more readily than static differences. However, static pictures are preferable in print. Therefore, we compare three snapshots from a coarse-grained movie of the motion pathway (*25*) for each of two protein pairs, 9aat1ama and 1hrd1bgv, in two forms: (a) superpositions of the reference wire apo structure and the wire structures in the snapshots and (b) the DDMs corresponding to these reference−snapshot structural pairs. The generation of the DDMs and the required (initial) superposition of two conformations of the same protein are performed as described in Methods.

Conformations from snapshots 1, 40, and 80 were chosen for each protein for comparison with the reference conformation of the corresponding protein. Because we have fit large rigid fragments of the two structures and have carefully chosen the viewing orientations of the pairs, it is possible to see some differences on the left side of the structure in Figure 11a−c. They, however, would be difficult to detect in many other projections. Because of the intricacies of the chain fold, it may also be difficult to assign these changes to positions along the chain. At the same time, the DDMs show easily identifiable changes in the intensity and shape of the nonblack areas of Figure 11a′−c′. These changes can also be easily projected on the sequence and secondary structure marked on the DDMs, and we used this to choose the residue number labels for the wire structures. This is also the case for changes that can be noticed on the left side of the snapshots in Figure 12a−c and the corresponding DDMs in Figure 12a′−c′. In 9aat1ama, N-terminal fragment 14−36 folds onto the C-terminal domain, and in 1hrd1bgv, C-terminal fragment 432−449 folds on the N-terminal rigid fragment.

It should be noted here that white in the DDMs indicates only that DDs in two structures differ by more than 1 Å. Thus, many changes within these spots are not visible. A development of the movement can be shown in a single cumulative MDDM (Movie DDM) by painting in different colors those areas that become white at different stages of the simulation. In a movie, it might be useful to show the wire structural superpositions and DDMs (like Figure 11a′−c′ or 12a′−c′) or MDDMs side by side for each snapshot, with a possibility of pausing both pictures at any frame.

DISCUSSION

Utility of Objective Thresholds and DDMs. Results presented in this work depend on the values of the coordinate uncertainty thresholds within which we consider two structures to be identical. We derived these values in our previous work (*10*). Currently, this provides us with useful objective thresholds. It is quite remarkable that the range of coordinate uncertainties, derived by using this set, does not overlap but just borders the range of similarly measured coordinate differences for function-induced protein coordinate changes. While these thresholds might change somewhat upon further use, we have shown here that currently they can serve as reasonable guidelines.

One of these two thresholds, RMSDD, is slightly below 0.5 Å, which is taken by us as the threshold value for DDs with smaller values to be colored black in the DDMs. The second threshold, Δ, reflects the number of DDs larger than 1 Å. We color DDs between 0.5 and 1 Å gray and above 1 Å white in the DDMs. Such DDMs, as shown above, facilitate the easy identification of protein fragments that undergo conformational transformations as rigid bodies. A slightly more careful analysis of the DDM allowed us to easily improve the accuracy of the original (*4*) description of the conformational change in functional loop 96−121 of lactate dehydrogenase, resolving four instead of two rigid fragments. This has led to the non-negligible 1.2 Å difference in the position of C^{α}_{103} near the active site. The more accurate positioning is due to the use of new techniques (*10*) allowing much easier and transparent identification of rigid fragment "movements" compared to previous techniques (e.g., compare to Figure 3 of ref *4*). Finding that functionally important loop 96−121 consists of four rigid fragments also challenges a popular perception of a loop usually acting as a solid lid (*1*). In fact, very often loops do not have any rigid substructures but "move" in a flexible fashion, rather like wet rags thrown on to

cover some part(s) of a protein (e.g., as we have seen in the kinases studied here).

What Should Be Considered a "Very Flexible Structure" or a "Major Movement"? The article on movements in adenylate kinase (*50*) calls it "a highly flexible protein", based on the large amplitude of the domain motion of ~30 Å. Maximum DDs agree rather well with higher RMSDDs but might lead to some confusion when related to the number of rigid-body motions required for transforming one of the structures into another (Table 1). The latter, however, directly shows how many chain fragments move while retaining their internal structure. It is obvious that a cabin with a door swinging on a hinge, even with large amplitude, is less flexible than caterpillar tracks, which in turn are less flexible than a bicycle chain, which is less flexible than a rope. This implies that flexibility increases with the number of rigid fragments and decreases with their size and might require some scaling of the cutoff number of rigid-body movements based on protein size. We have used a uniform cutoff of 12 movements, but we tried up to 20 movements for a few large proteins, finding that the transformations were still unable to superimpose two structures within the coordinate uncertainty threshold. More experience and an improved understanding may eventually answer this question and facilitate an automation of the fitting process.

Functional "movements" in proteins are often described in the literature as "mainly hinge" or "mainly shear" (*5*, *6*). Classifications into "major" and "minor" movers were also used (*4*). One might be legitimately interested not in all details but only in a major component of a functional movement. However, it is often difficult to unambiguously define what is major. Is it a movement of the rigid fragment with the largest number of residues or perhaps a movement of a shorter fragment over a longer distance? There could be also various "weighted" combinations of these and other criteria. One question that can be easily answered with our methodology would concern the movements that are required to make all DDs smaller than, e.g., 2 Å. Changing this threshold is easy and might provide useful information. From what we have seen though, this would obscure and exclude from a determination the functional "motion" found in thermolysin (*46*), with the maximum DD being 1.8 Å, as well as introduce chain breaks of up to 2 Å.

Protein Stability and Compactness of X-ray Structures. It is clear that for a protein to be compact it must be stable; otherwise, it would have many alternative conformations which would require a larger volume to accommodate all of them. Because we ignored interactions with substrates or cofactors and between subunits in oligomers, which can provide important stabilization, we considered compactness only for monomeric apoproteins. We should also recall the significant simplifications in these calculations and therefore realize that our estimates of stability may require more validation against available experimental data.

In our set of 17 proteins (Tables 1−3), 10 proteins (1cll, 4cd2, 4ank, 1akz, 1l3f, 4ape, 1erk, 16pk, 1v4t, and 1lfh) do not form oligomers in their apo form crystals according to PQS.

1l3f (thermolysin) is the most interesting among the other monomeric apoproteins in our list (Tables 1−3). It has a high calculated stability for the entire protein (−18 kcal/mol) which is very likely to be beyond the inaccuracy of the method. Therefore, thermolysin is likely to be compact in solution. However, it is also noted that apothermolysin crystallization required Zn ions (*46*), and it is not clear what their role is in stabilizing the structure. The high *B* factors in the apothermolysin 1l3f do suggest some residual movement in the crystal (*46*).

Another interesting feature of the calculations for thermolysin is a 15 kcal/mol decrease in the stability of the entire protein chain upon substrate binding (1l3f → 3tmn). This is likely to be beyond any possible errors. Calculations for another holo structure of thermolysin (8tln) yield results within 2 kcal/mol. This is suggestive of a mechanism in which substrate binding energy is stored in a strain (destabilization) of the protein that might subsequently be used for the ejection of the reaction product. Certain confidence in our method comes from its use for predictions of stable domains and subdomains in thermolysin (*18*), which were then successfully isolated and characterized confirming the predictions (*35*, *51*).

Of the first four apo monomers, only one, 1akz, has a marginal stability. The increase in stability by 8 kcal/mol in the holo form may still lie within the largest possible error bounds of the method; however, the sign of the change suggests that the apo monomer is stable and, thus, compact. The three other monomeric proteins have marginally stable domains, within the method's error bounds, but instabilities of the entire structures for these are worse than 10 kcal/mol. We can safely infer that it is rather unlikely that these monomers are stable and compact. They might be stabilized by the crystal structures. After the substrate binds, they can be stabilized by the binding energy. This suggests that these proteins either are incompletely folded or have domains moving independently in solution before they recognize and then bind substrate. Note that adenylate kinase exhibits three marginally stable substructures in the monomeric apo form 4ake and only one in the holo form 1ank. This may suggest that intrachain contributions to protein stability might be traded for stabilization by substrate binding. This "stability trade-off" resembles the hypothesis of "energetic counterweight balancing" for storage of the substrate binding energy in adenylate kinase by shifting the region of higher mobility in the catalytic cycle (*52*). Mobile sites were located (*52*) according to the *B* factors and cannot be straightforwardly related to the change in domain stability. However, it is encouraging that we find a similar effect despite its small magnitude (Table 3). A similar but larger stabilization trade-off was seen above in thermolysin.

Three apo monomers (4ape, 1erk, and 1v4t) do show the instability of the entire protein above 10 kcal/mol, with two (4ape and 1v4t) showing rather stable folded fragments and one (1erk) not so stable (−2 to −5 kcal/mol). Thus, they are unlikely to be compactly folded in solution.

Monomeric 16pk has a hardly stable (3 kcal/mol) structure. However, it is not a fully open apo form but some intermediate conformation. A fully open apo form from another organism (1vjd) also shows an only marginal stability of −2 kcal/mol. Attempts to crystallize both fully open and closed structures from the same organism were not successful. Reported experimental domain stability studies indicated that the C-terminal domain is more stable than the N-terminal one (*20*). While our calculation may overestimate the stabilities of these domains, they show that the C-terminal domain is significantly more stable than the N-terminal one for both 16pk and 1vjd, in close agreement with the experiment.

The largest monomeric protein in our set, 1lfh, is strongly stabilized by 16 disulfide cross-links. However, even accounting for that does not reveal any stability for the entire molecule. As with many other proteins not exhibiting stability of the whole molecule, it seems likely that its domains move freely relative to

one another until they catch a substrate (and/or cofactor) that stabilizes the entire protein. For oligomeric proteins, such stabilization might also be provided by intersubunit interactions. Thus, nine of ten monomeric proteins studied here either seem to be only marginally stable or have folded stable substructures that are mobile in the apo form. This would enhance the access of the substrate to the active site.

For nonmonomeric apoproteins (2tbv, 6ldh, 2gd1, 8adh, 9aat, 4cts, and 1hrd), it is impossible to infer compactness on the basis of stabilities calculated for monomers because of the lack of clarity in treating the thermodynamics of protein binding and association. For small molecules, binding and association depend on the concentration and thus on the loss of rotational and translational entropy. It was estimated (*53*) that dimerization of a protein should cost ~15 kcal/mol in lost entropy, according to classical theory for small molecules. However, it was noted (*19*) that proteins do refold after being cut into two or three parts. This would supposedly cost 15−30 kcal/mol in lost translational and rotational entropy of independent fragments and thus prevent folding into a stable structure, which without cuts usually has a stability no better than −15 kcal/mol. This paradox was experimentally supported by calorimetric measurements of the unfolding and dissociation entropy of a homodimeric protein and of its mutant with covalently linked monomers (*54*). The entropies were practically identical in both cases, demonstrating that classical entropic terms do not play any role in protein association.

However, experimental studies of subunits and fragments of lactate dehydrogenase (*20*) allow some comparisons with our stability calculations for subunits of this tetrameric protein. The dimer-of-dimers structure of this protein is stabilized by the N-terminal decapeptide of each of the subunit polypeptide chains. After this decapeptide is being cleaved off, a metastable "proteolytic dimer" is obtained, which shows anomalously high flexibility and gains enzymatic activity only in the presence of structure-making salts. The monomer shows a nativelike structure, accessible only as a short-lived, proteolytically sensitive kinetic intermediate on the pathway of reconstitution. The separate NAD- and substrate-binding domains are unstable but still sufficiently structured to recognize one another and to pair correctly. Upon joint reconstitution, they form active, nicked dimers. However, renaturation of the separate domains and subsequent mixing are unsuccessful. Thus, domains may assist each other in advancing to their native structure by mutual stabilization through specific interactions.

Our calculations for 6ldh show four stable fragments which do not include the N-terminal decapeptide. Each of two CATH domains includes two of these four stable fragments. However, the first pair of stable fragments is calculated to form a larger folded fragment with a doubtful stability of 0 kcal/mol, corresponding to a usually assigned lactate dehydrogenase N-terminal domain. The second pair shows no calculated stability better than 10 kcal/mol, very strongly suggesting the lack of any stability for the entire C-terminal domain. Thus, results of our stability calculations agree with the experimental instability of individual domains having stable subdomains, possibly sufficient for domain−domain recognition, as well as with the instability of the monomers.

All of these correlations and agreements with experimental data strongly indicate that our method for calculating the stability of protein fragments can be rather reliably used for developing plausible suggestions deserving of further verification.

Stable Fragments, Rigidly Moving Fragments, and Continuous, Discontinuous, and CATH Domains. Some recent papers identified domains as structural modules that move rigidly relative to one another during protein function (*21*). The number of rigid-body motions required to transform the conformation of one functional state of a protein into that of another state is given in Table 1. Numbers of calculated stable fragments for the same protein can be found in Table 3. For all four kinases in our set, the number of required rigid-body motions exceeds 12 while the calculated number of stable fragments is significantly smaller. The situation is similar for 4cd21cd2, 1akz1ssp, 6ldh1ldm, 8adh6adh, and 1lfh1lfg, which require from five to 11 rigid-body movements and have significantly fewer calculated stable fragments. 9aat1ama (entries 12 and 12′) has only two nonoverlapping stable fragments but requires at least three rigid fragments for fitting. For 1cll1ctr, there is a partial agreement between the marginally stable fragments (Table 3) and localization of rigidly moved fragments. However, some stable fragments are broken into several rigid fragments, and an unstable N-terminal fragment moves as a rigid body. For 2tbv, 1l3f3tmn, 4ape5er2, and 2gd11gd1, the agreement is reasonable; they have two or three nonoverlapping stable fragments and a similar number of rigid fragments. For the pair 1hrd1bgv, the numbers of rigid and stable fragments are either similar (entry 16) or, in a less accurate fit, there are more stable than rigid fragments (entry 16′).

Thus, for only five of 17 proteins do we find reasonable agreement between the rigid independently moving fragments and the nonoverlapping stable fragments. This may mean that, just as in one stable domain of citrate synthase, parts of a stable domain can shift significantly without the loss of stability of the entire domain in contrast to the view of full rigidity for folded domains (*1*). Generally, our results strongly suggest that rigid and stable fragments do not coincide in a majority of cases.

All earlier publications referred to small and large domains in citrate synthase (*3*, *6*, *47*, *48*). Interestingly, we found no fragments of the large domain having a stability better than 0 or 1 kcal/mol, and these are usually helical hairpins. We repeated stability calculations, excising the sequence (residues 272−389) of the small domain with two or three extra loop residues on each side from its structure [unfolded fragments with fixed ends are entropically very destabilizing (see eq S3 of the Supporting Information)]. This led to the appearance of an instability of 2 kcal/mol for the large domain. Thus, the large discontinuous domain seems to be unstable without the small domain and apparently may assemble on its preformed surface. Of course, stabilization by quaternary structure is also possible.

CATH does not agree with continuous stable domain localization for 8adh (see the previous section of Discussion), poorly agrees for 2gd1, lumps together into one domain two stable fragments separated by an unstable loop in 6ldh, two unstable fragments in citrate synthase, 4cts, and in 1lfh, and does not agree for the three kinases (4ake, 1erk, and 1v4t).

CATH identifies a discontinuous domain corresponding to two marginally stable fragments in 1hrd1bgv, which might represent a variation of the discontinuous domain of 4cts already discussed above. In 1lfh, CATH assigns two discontinuous domains, one of which does not include any stable fragments and another that overlaps with only one such stable fragment. These discontinuous domains might be a kind of "afterthought" in the folding process, when other stable domains can stabilize them, as in citrate synthase. Thus, there are significant disagreements between the

CATH domain locations and our computed stable domains for eight of 17 proteins.

Do Large Interfaces in Dimers Preclude Hinge Motions? Our results allow to justifiably challenge the earlier idea (*3*) that the large interface in a dimer precludes simpler hinge or pliers motions in citrate synthase. According to PQS, dimerization buries 4170 Å^2 of ASA per 1cts monomer and 4600 Å^2 of ASA per 4cts monomer. 1hrd buries 3924 Å^2 of ASA per monomer in the oligomerization, and 1bgv buries 3946 Å^2 according to PQS. The RMSDDs for 1hrd1bgv and 4cts1cts are also comparable [1.89 and 1.58 Å, respectively (Table 1)]. Nevertheless, 1hrd1bgv (Figure 12) is not constrained to undergo its functional movement through a simple pliers mechanism. Judging from changes in the intersubunit buried area between 4cts and 1cts, the relative monomer−monomer movement is larger in the 4cts to 1cts conformational change than in the 1hrd to 1bgv conformational transition, where the intersubunit buried area practically does not change. In the dimeric 9aat1ama, the intersubunit buried areas per monomer are slightly lower (3270 and 3500 Å^2, respectively). However, it has the third highest δ (estimating shear) as shown in Table 1. We concluded that 9aat1ama undergoes a functional conformational transition through a rather simple tweezers motion.

Why Do Proteins with Closely Related Functions Exhibit Very Different Functional Motions? There were some previous efforts to understand this (*46*). However, most efforts were restricted to protein classification based on their not-so-well-defined domain structures (*36*, *37*) or to determining whether a function-driven conformational change involves large hinge or shear motions (*6*).

There are two protein groups with similar functions each in our study: dehydrogenases and kinases. In both groups, we find examples of movements of rigid or stable fragments or domains (e.g., 2gd11gd1, 1hrd1bgv, and 8adh6adh for dehydrogenases; 16pk13pk and less clearly for 4ake1ank in kinases) and of mainly loop movements (e.g., 6ldh1ldm for dehydrogenases and 1erk2erk for kinases). The first type of movement is characterized by large white spots in the DDMs, and the second type is characterized by L-shaped white bands in their DDMs. It is possible that nature uses any working design, as we do.

It was noted previously (*46*) that in dehydrogenases NAD-binding domains are similar but might be located in different parts of the chain. Catalytic domains could be quite different because of the sizes and shapes of the substrates. Not much is currently understood about what determines particular designs and modes of "motion" in dehydrogenases and how that occurs. Tetrameric lactate and dimeric malate dehydrogenases have similar domain structures with, however, some possibly important variations in the locations of stable fragments. The N-terminal decapeptide, holding together two dimers in the tetramer of lactate dehydrogenase (ldh) (*20*), folds as a part into the N-terminal stable fragment of dimeric malate dehydrogenase (mdh), a monomer of which may also have some stability. The N-terminal domains are unstable or only marginally stable in both proteins and are comprised of two independently stable fragments. The second domain in ldh is unstable and is formed by two marginally stable fragments. There is only one stable fragment within the second domain of mdh, but the entire domain is reasonably stable. Furthermore, the fold without the N-terminal fragment also seems to be stable. Are these differences required for function? A few mutations in lactate dehydrogenase turn it into a catalytically effective malate dehydrogenase (*55*). A complication is introduced by the fact that crystallized pairs of apo and tertiary forms of malate dehydrogenases (1b8p1b8u and 2cmd1emd) do not show any motions above the coordinate uncertainty. Is this caused by the choice of improper substrate analogues, or does the crystallization itself play a dominant role? A possible answer is provided by the mRNA capping enzyme (*56*), with one molecule in an open and another in a closed conformation in the same unit cell. Running the biochemical reaction in the crystal showed that only the closed form is activated. Thus, closing of the active site might be a necessary step in many reactions, with the binding happening in an open form often preserved by the crystallization.

Can Elastic Model Dynamics Reveal Barriers or Breaking of Rigid Fragments during the Motions? Some useful insights may be provided by combining elastic network modeling of the functional conformational changes and the location of rigid fragments extracted from the DDM for the end points of the conformational change and from the estimates of fragment stability. The advantage of the sequences of conformational changes, provided by the elastic models compared with earlier interpolation morphs, is the avoidance of significant atomic overlaps during the motions. However, currently the question of whether transitions follow unique paths and how much the paths depend on elastic model parameters is not fully resolved. Another important question is whether such smooth nonoverlapping motions require significant distortions at intermediate stages of fragments that seem to be retaining rigidity in the end points DDM.

Concluding Remark. The methods, results, and discussion presented here are likely only a starting point in expanding the studies and interpretations of the vast new information about protein flexibility phenomena. Nevertheless, they can already provide biochemists with a rather comprehensive approach, allowing for a more transparent, multifaceted, detailed, and accurate understanding and interpretation of the function-related conformational changes in proteins than other earlier proposed approaches.

SUPPORTING INFORMATION AVAILABLE

Additional information, supporting and clarifying methods, results, and conclusions. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Petsko, G. A., and Ringe, D. (2004) Protein Structure and Function, Blackwell, London.
2. Chothia, C., Lesk, A. M., Dodson, G. G., and Hodgkin, C. (1983) Transmission of conformational change in insulin. *Nature 302*, 500–505.
3. Lesk, A. M., and Chothia, C. (1984) Mechanisms of domain closure in proteins. *J. Mol. Biol. 174*, 175–191.
4. Gerstein, M., and Chothia, C. (1991) Analysis of protein loop closure. Two types of hinges produce one motion in lactate dehydrogenase. *J. Mol. Biol. 220*, 133–149.
5. Krebs, W. G., and Gerstein, M. (2000) The morph server: A standardized system for analyzing and visualizing macromolecular motions in a database framework. *Nucleic Acids Res. 28*, 1665–1675.
6. Gerstein, M., Lesk, A., and Chothia, C. (1994) Structural mechanisms for domain movements in proteins. *Biochemistry 33*, 6739–6749.
7. Gerstein, M., and Krebs, W. G. (1998) A database of macromolecular motions. *Nucleic Acids Res. 26*, 4280–4290.
8. Krebs, W. G., Tsai, J., Alexandrov, V., Echols, N., Junker, J., Jansen, R., and Gerstein, M. (2003) Studying protein flexibility in a statistical framework: Tools and databases for analyzing structures and approaches for mapping this onto sequences. *Methods Enzymol. 374*, 544–584.

9. Gerstein, M., and Echols, N. (2004) Exploring the range of protein flexibility, from a structural proteomics perspective. *Curr. Opin. Chem. Biol. 8*, 14–19.
10. Rashin, A. A., Rashin, A. H. L., and Jernigan, R. L. (2009) Protein flexibility: Coordinate uncertainty and interpretation of structural differences. *Acta Crystallogr. 65*, 1140–1161.
11. Wernisch, L., and Wodak, S. J. (2003) Identifying structural domains in proteins. In Structural Bioinformatics (Bourne, P. E., and Weissig, H., Eds.) pp 365−385, Wiley, New York.
12. Veretnik, S., Bourne, P. E., Alexandrov, N. N., and Shyndialov, I. N. (2004) Toward consistent assignment of domains in proteins. *J. Mol. Biol. 339*, 647–678.
13. Wetlaufer, D. B. (1973) Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. U.S.A. 70*, 697–701.
14. Wodak, S. J., and Janin, J. (1981) Location of structural domains in proteins. *Biochemistry 20*, 6544–6552.
15. Rashin, A. A. (1979) An approach to calculation of the folding pathway from protein X-ray structure. *Stud. Biophys. 77*, 177–184.
16. Rashin, A. A. (1981) Location of domains in globular proteins. *Nature 291*, 85–87.
17. Rashin, A. A. (1984) Buried area, conformation entropy, and protein stability. *Biopolymers 23*, 1605–1620.
18. Rashin, A. A. (1984) Prediction of stabilities of thermolysin fragments. *Biochemistry 23*, 5518–5519.
19. Rashin, A. A. (1993) Aspects of protein energetics and dynamics. *Prog. Biophys. Mol. Biol. 60*, 73–200.
20. Jaenicke, R. (1999) Stability and folding of domain proteins. *Prog. Biophys. Mol. Biol. 71*, 155–241.
21. Schneider, T. S. (2004) Domain identification by iterative analysis of error-scaled difference distance matrices. *Acta Crystallogr. D60*, 2269–2275.
22. Chacyn, P. (2006) http://sbg.ci.csic.es/index.html.
23. Yang, L., Song, G., Carriquiry, A., and Jernigan, R. L. (2008) Close correspondence between the motions from principal component analysis of multiple HIV-1 protease structures and elastic network modes. *Structure 16*, 321–330.
24. Echols, N., Milburn, D., and Gerstein, M. (2003) MolMovDB: Analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res. 31*, 478–482.
25. Kim, M. K., Chirikjian, G. S., and Jernigan, R. L. (2002) Elastic models of conformational transitions in macromolecules. *J. Mol. Graphics Modell. 21*, 151–160.
26. Kim, M. K., Jernigan, R. L., and Chirikjian, G. S. (2002) Efficient generation of feasible pathways for protein conformational transitions. *Biophys. J. 83*, 1620–1630.
27. Nishikawa, K., Ooi, T., Isogai, Y, and Saito, N. (1972) Tertiary structure of proteins. I. Representation and computation of the conformations. *J. Phys. Soc. Jpn. 32*, 1331–1337.
28. Kuipers, J. B. (1998) Quaternions and Rotation Sequences, Princeton University Press, Princeton, NJ.
29. Chothia, C., and Janin, J. (1978) Geometrical principles that determine the three dimensional structure of proteins. *Proc. FEBS Meet. 52*, 117–126.
30. Chothia, C., Levitt, M., and Richardson, D. (1981) Helix to helix packing in proteins. *J. Mol. Biol. 145*, 215–250.
31. Horn, B. K. P. (1986) Closed form solution of absolute orientation using union quaternions. *J. Opt. Soc. Am. 4*, 629–642.
32. Maiti, R., Van Domselaar, G. H., Zhang, H., and Wishard, D. S. (2004) Superpose: A simple server for sophisticated structural superpositions. *Nucleic Acids Res. 32*, W590–W594 (Web Server Issue).
33. Coutsias, E. A., Seok, C., and Dill, K. A. (2004) Using quaternions to calculate RMSD. *J. Comput. Chem. 25*, 1849–1857.
34. Kavraki, L. E. (2006) Molecular Distance Measures, version 1.16. http://cnx.org/content/m11608.
35. Vita, C., Dalzoppo, D., and Fontana, A. (1984) Independent folding of the carboxyl-terminal fragment 228−316 of thermolysin. *Biochemistry 23*, 5513–5518.
36. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M. (1997) CATH: A hierarchic classification of protein domain structures. *Structure 5*, 1093–1108.
37. Murzin, A. G., Brenner, S. E., Hubbard, T., and Chothia, C. (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol. 247*, 536–540.
38. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res. 28*, 235–242.
39. Privalov, G. P., and Privalov, P. L. (2000) Problems and prospects in the microcalorimetry of biological macromolecules. *Methods Enzymol. 323*, 31–62.
40. Rashin, A. A., and Rashin, A. H. L. (2007) Surface hydrophobic groups, stability and flip-flopping in lattice proteins. *Proteins 66*, 321–341.
41. Henrick, K., and Thornton, J. M. (1998) PQS: A protein quaternary structure file server. *Trends Biol. Sci. 23*, 358–361.
42. Privalov, P. L. (1979) Stability of proteins: Small globular proteins. *Adv. Protein Chem. 33*, 167–241.
43. Tsong, T. Y., Hearn, R. P., Wrathall, D. P., and Sturtevant, J. M. (1970) A calorimetric study of thermally induced conformational transitions of ribonuclease A and certain of its derivatives. *Biochemistry 9*, 2666–2677.
44. Harrison, S. C. (1980) Protein interfaces and intersubunit bonding. The case of tomato bushy stunt virus. *Biophys. J. 32*, 139–151.
45. Lesk, A., and Chothia, C. (1988) Elbow motion in the immunoglobulins involves a molecular ball-and-socket joint. *Nature 335*, 188–190.
46. Hausrath, A. C., and Matthews, B. W. (2002) Thermolysin in the absence of substrate has an open conformation. *Acta Crystallogr. D58*, 1002–1007.
47. Janin, J., and Wodak, S. J. (1983) Structural domains in proteins and their role in the dynamics of protein function. *Prog. Biophys. Mol. Biol. 42*, 21–78.
48. Remington, S., Wiegrand, G., and Huber, R. (1982) Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. *J. Mol. Biol. 158*, 111–152.
49. Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science 298*, 1912–1934.
50. Berry, M. B., Meador, B., Bilderback, T., Liang, P., Glaser, M., and Phillips, G. N., Jr. (1994) The closed conformation of a highly flexible protein: The structure of *E. coli* adenylate kinase with bound AMP and AMPPNP. *Proteins 19*, 193–198.
51. Jaenicke, R. (1991) Protein folding: Local structures, domains, subunits, and assemblies. *Biochemistry 30*, 3147–3161.
52. Muller, C. W., Schlauderer, G. J., Reinstein, J., and Schulz, G. E. (1996) Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding. *Structure 4*, 147–156.
53. Finkelstein, A. V., and Janin, J. (1989) The price of lost freedom: Entropy of bimolecular complex formation. *Protein Eng. 3*, 1–3.
54. Tamura, A., and Privalov, P. (1997) The entropy cost of protein association. *J. Mol. Biol. 273*, 1048–1060.
55. Wilks, H. M., Hart, K. W., Feeney, R., Dunn, C. R., Muirhead, H., Chia, W. N., Barstow, D. A., Atkinson, T., Clarke, A. R., and Holbrook, J. J. (1988) A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science 242*, 1541–1544.
56. Hakansson, K., Doherty, A. J., Shuman, S., and Wigley, D. B. (1997) X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell 89*, 545–553.