PROTEINS: Structure, Function, and Genetics 30:144–154 (1998)

Systematic Analysis of Domain Motions in Proteins From Conformational Change: New Results on Citrate Synthase and T4 Lysozyme

Steven Hayward and Herman J.C. Berendsen*
BIOSON Research Institute, Laboratory of Biophysical Chemistry, University of Groningen, Groningen, The Netherlands

ABSTRACT **Methods developed originally to analyze domain motions from simulation [Proteins 27:425–437, 1997] are adapted and extended for the analysis of X-ray conformers and for proteins with more than two domains. The method can be applied as an automatic procedure to any case where more than one conformation is available. The basis of the methodology is that domains can be recognized from the difference in the parameters governing their quasi-rigid body motion, and in particular their rotation vectors. A clustering algorithm is used to determine clusters of rotation vectors corresponding to main-chain segments that form possible dynamic domains. Domains are accepted for further analysis on the basis of a ratio of interdomain to intradomain fluctuation, and Chasles' theorem is used to determine interdomain screw axes. Finally residues involved in the interdomain motion are identified. The methodology is tested on citrate synthase and the M6I mutant of T4 lysozyme. In both cases new aspects to their conformational change are revealed, as are individual residues intimately involved in their dynamics. For citrate synthase the beta sheet is identified to be part of the hinging mechanism. In the case of T4 lysozyme, one of the four transitions in the pathway from the closed to the open conformation, furnished four dynamic domains rather than the expected two. This result indicates that the number of dynamic domains a protein possesses may not be a constant of the motion. Proteins 30:144–154, 1998.** © **1998 Wiley-Liss, Inc.**

Key words: hinge bending; X-ray conformers; Chasles' theorem

INTRODUCTION

Large proteins are often built of domains. As techniques in X-ray crystallography improve and larger proteins are tackled, proteins without domain structures are likely to become a minority in the crystallographic data bank. Although relatively few in number, proteins in more than one conformation are also being submitted to this data bank.[1,2] Some proteins such as a mutant of T4 lysozyme crystallize in more than one conformation revealing a hinge-bending motion,[3–5] and it is thought that these different conformations are actually snapshots of states accessible to the protein under normal conditions.[5–7] If this is the case, such snapshots provide valuable experimental information on functionally important motions. Others proteins, such as citrate synthase[8,9] are found in an open conformation in the uncomplexed state, but in a closed state when complexed with a ligand. Again as techniques improve, the proportion of structures in the data bank representing more than one conformational state is expected to increase.

In a review of the structural mechanisms involved in domain motions Gerstein and colleagues[2] have classified the types of structural changes into hinge and shear motions, and shown how these "primitive motions" combine to produce different kinds of domain motions. Such analyses are of great value as an understanding of the relationship between interdomain interactions and structures, and the types of domain motions that they produce will be necessary for future protein engineering applications. In addition, the accurate recognition of sites involved in interdomain motions may identify possible target sites for drugs, especially if these sites have greater polymorphism than active sites.

The different definitions of the word "domain" in the literature, arises from the differing concepts among the various branches of protein study. If one has two conformations of a protein, or can furnish more than one conformation by doing simulation, then one has the possibility to define a domain on the basis of groups of residues moving in a concerted fashion. Such a dynamic definition is useful because the function of many proteins can be related to the concerted motion of its parts.

However, the identification of domains from conformers is not a trivial problem as one has to be able

Contract grant sponsor: BIOTECH, European Union.
*Correspondence to: Professor H.J.C. Berendsen, Biophysical Chemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands. E-mail: berends@chem.rug.nl
Received 1 December 1996; Accepted 1 August 1997

to make a distinction between conformational changes arising from interdomain motions, and those arising from intradomain motions. A number of methods such as the difference–distance method,[10] the static-core method[11] and the deformation-plot analysis[12] among others, are already in use to determine domains from conformational change. The difference–distance methods are based on the idea that intradomain atomic distances should be largely unchanged in comparison to interdomain atomic distances. However, depending on the location of the hinge axis even interdomain atomic distances may also remain unchanged, making difference–distance plots difficult to interpret.

Here we introduce an alternative approach based on the theory of rigid body kinematics that is an extension of a method we have applied recently to determine the dynamic domains of lysozyme from simulation data.[13] In the next section, we introduce the method, and then demonstrate it on citrate synthase and the M6I mutant of T4 lysozyme.

METHODS

The underlying idea behind the method is that domains can be identified by their differing rotational properties. As in the previous paper, a rotation–orientation matrix can be defined as the matrix of inner products, for all residue pairs, of the unit rotation vectors that describe the transformation from one structure to another. For example, in the previous paper,[13] the two domains of human lysozyme were identified from a normal mode analysis by plotting the rotation–orientation matrix which revealed two domains rotating in approximately opposite directions. Once domains are identified, interdomain screw axes and other properties pertaining to the domain motion can be determined. However, as long as human judgment is required to recognize the domains from a plot of the rotation–orientation matrix, the methodology remains limited, especially in the case of large multidomain proteins. Here we present a method that no longer requires any human intervention and is applicable to multidomain proteins. This is achieved by use of a clustering algorithm to identify dynamic domains from clusters of rotation vectors in the rotation space.

Identification of Dynamic Domains

Rotation vectors

The only ingredients necessary for application of the method are a structure plus a set of displacement vectors showing how each atom or subset of atoms moves in the conformational change. Such a set of displacement vectors is given directly from a normal mode analysis,[14–18] and can be determined from a molecular dynamics simulation by performing a multivariate analysis of the trajectory.[18–23] A set of displacement vectors can also be determined from a pair of X-ray conformers after performing a least-squares best fit. Although we are, in principle, free to choose any subset of atoms from the two X-ray conformers on which to perform the best fit, it is clear that the only unbiased choice is to perform a least-squares best-fit *over the whole of the protein* (strictly speaking, this should be a mass-weighted best fit, but actually here only the atoms: N, CA, CB, and C, in the protein data bank (PDB) nomenclature, are used, each taken to have equal mass). The resulting set of displacement vectors is one consistent with the internal coordinate system,[24,25] just as in the case of normal mode analysis, although the internal coordinate system only strictly exists for infinitesimal deformations.

Given a structure and a set of displacements vectors it is possible to determine the rotation vectors (see below) of individual residues by calculating the curl of the displacement vector field at each residue as was described in the previous paper. Here, since localized rotations often obscure the effects of whole domain rotations, we have found it necessary to calculate the rotation vectors of short main-chain segments. Main-chain segments generated using a sliding window of a chosen length were displaced from their initial conformations to their final conformations according to their atomic displacement vectors. By superimposing each main-chain segment in its initial conformation onto its final conformation by least-squares best fitting, we can determine the displacement vectors representing the *rigid body* displacement of the segment. Rotation vectors are then calculated for each segment.

Each rotation vector can be represented by a "rotation point" in the three-dimensional rotation space, whose coordinates are the components of the rotation vector. If we consider an ideal domain moving as a perfect rigid body, all the rotation vectors corresponding to residues or main-chain segments within the domain will be identical, and all the rotation points will coincide at a single point representing the rotation of the whole domain. If a protein comprises n ideal domains each with a different rotation vector, we will see n rotation points. In the real world the rotation points of the main-chain segments of a domain protein will not coincide according to domain, due to intradomain variation, but should at least form clusters in the rotation space (for an example, see Fig. 5). If we have a method to identify these clusters we can identify dynamic domains.

Clustering algorithm and definition of dynamic domains

We use the K-means clustering algorithm[26] to determine the clusters of rotation points. The K-means method divides the points in K clusters, where K is a specified number. The obvious problem with this method is that we get as many clusters as

we specify. In other words, we require the process to stop automatically when some physically meaningful criterion is not satisfied. One criterion we use is minimum domain size in number of residues. This, however, is not sufficient.

We are interested in the determination of physically meaningful quantities concerning interdomain motions. We therefore only analyze the relative motions of pairs of domains connected through the main chain. We should also not analyze the relative motion of a connected pair of domains if the intradomain fluctuation is greater than the interdomain fluctuation. If we can quantify the intradomain and interdomain fluctuations, then our other criterion should be based on the ratio of these two quantities.

For simplicity consider a two-domain protein in two conformations. If we do a whole-protein best fit to generate a set of displacement vectors, $\Delta\mathbf{d}_i$, then, as described in the previous paper, the relative contributions to the total mean-square fluctuation (MSF) from the interdomain and interdomain motions can be calculated, for

$$\sum_i |\Delta\mathbf{d}_i|^2 \approx \sum_i |\Delta\mathbf{d}_i^{\text{ext}}|^2 + \sum_i |\Delta\mathbf{d}_i^{\text{int}}|^2 \qquad (1)$$

and the ratio

$$\frac{\sum_i |\Delta\mathbf{d}_i^{\text{ext}}|^2}{\sum_i |\Delta\mathbf{d}_i^{\text{int}}|^2} \qquad (2)$$

could be used as our criterion. Here the superscript "ext" (external to the domains) refers to the interdomain contribution, and "int" (internal to the domains) refers to the intradomain contribution. However, this quantity depends on the relative domain sizes. For example, if one domain is much larger than the other, the intradomain motion within the large domain will far outweigh the amount of external motion in the small domain. Ideally, we require a quantity that does not depend on the relative sizes of the domains. This is achieved by first assigning masses of $1/N_A$ to atoms in domain A, where N_A is the number of residues in domain A, and masses of $1/N_B$ to atoms in domain B, where N_B is the number of residues in domain B. If we then perform a whole-protein "mass"-weighted best fit, the quantity,

$$\chi = \sqrt{\frac{\left(\sum_i^{N_A} \left|\Delta\mathbf{d}_i^{\text{ext}}\right|^2 / N_A\right) + \left(\sum_i^{N_B} \left|\Delta\mathbf{d}_i^{\text{ext}}\right|^2 / N_B\right)}{\left(\sum_i^{N_A} \left|\Delta\mathbf{d}_i^{\text{int}}\right|^2 / N_A\right) + \left(\sum_i^{N_B} \left|\Delta\mathbf{d}_i^{\text{int}}\right|^2 / N_B\right)}} \qquad (3)$$

is a measure of the ratio of interdomain motion to intradomain motion independent of the relative sizes of the domains. In this work we accept a domain pair for the analysis of their interdomain motion only if χ exceeds 1.0. Note that χ is not a measure of the rigidity of the domains. It sometimes makes sense to analyze the external motion of an object even if it has considerable internal motion.

From the clustering algorithm we take the dynamic domains to be those corresponding to the clusters at the level of the division where all domains are larger than the minimum domain size, and where the number of domains for which *all* connected domain pairs have χ values greater than 1.0, is a maximum. In order to find this maximum the value of K in the clustering algorithm is set to a large number (e.g., 100). Subsequently, all connected dynamic domain pairs are analyzed for their interdomain motions as described below.

One remaining problem is that if two domains happen to have the same rotation vectors they are not distinguishable. At present we check that all residues within a domain, as determined above, are contacting each other, i.e. that they all form one compact domain. Another solution to this problem is described in the Conclusions section.

Determination of the Interdomain Screw Axis

Consider a protein with two dynamic domains α and β. In order to find a possible hinge axis we should determine the motion of one domain relative to the other. If we fit the two conformations of domain α together using a least-squares best-fit routine, reorienting the whole structures accordingly, then we have the initial and final positions of domain β in its movement relative to domain α (we will always call the domain we fix in space α). If we superimpose the initial conformation of domain β onto the final conformation of domain β, and consider the displacements vectors between atoms in the initial and superimposed conformation of domain β, we then have a set of $\Delta\mathbf{d}_i^{\text{ext}}$'s which give the atomic displacement vectors corresponding to the general displacement of domain β as a rigid body relative to domain α.

The basis of the method we use to determine a possible hinge axis is Chasles' theorem, which states that the most general displacement of a rigid body is a screw motion about a unique screw axis (see footnote on page 163 of Goldstein[27]). This can be expressed mathematically as:

$$\Delta\mathbf{d}_i^{\text{ext}} = (1 - \cos\Delta\theta)(\mathbf{n}(\mathbf{n}\cdot\mathbf{r}_i^s) - \mathbf{r}_i^s) + \sin\Delta\theta(\mathbf{n}\times\mathbf{r}_i^s) + \Delta\tau_\parallel\mathbf{n} \qquad (4)$$

where $\mathbf{r}_i^s$ is the atomic position vector from an arbitrary origin on the interdomain screw axis, $\Delta\theta$ is the angle through which the body is rotated, $\mathbf{n}$ is the unit vector along the axis, and $\Delta\tau_\parallel$ is the translation

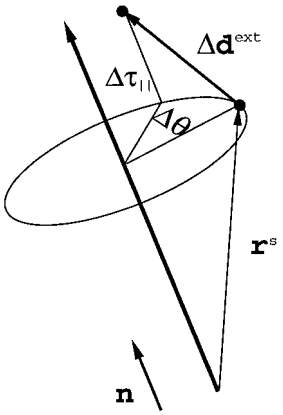

Fig. 1. Illustration of the meaning of $\Delta\theta$, $\Delta\tau_{\|}$, **n**, and $\mathbf{r}^s$ and their relation to $\Delta\mathbf{d}^{ext}$ in a general rigid body displacement.

along the axis (see Fig. 1). Note that there are six parameters specifying the rigid body displacement: two for the orientation of the axis, two to specify its location, one to specify the magnitude of rotation, $\Delta\theta$, and finally one to specify the magnitude of the translation along the axis, $\Delta\tau_{\|}$. This theorem has been used before, without citation, in X-ray crystallographic refinement,[28] in an analysis of the domain motion in liver alcohol dehydrogenase[29] and lactoferrin,[11] and possibly in a number of other studies of this nature.

The path domain β takes in going from its initial position to its final position may be totally different from that implied by Equation (4), but if domain β undergoes a rigid body rotation about a physical hinge axis without any translation, $\Delta\tau_{\|}$ will be zero and the axis of Equation (4) will coincide with the hinge axis. This is the crucial point behind the usage of Equation (4). If we can determine all six parameters, and they reveal an axis inconsistent with a hinge axis, then we will know that the domain motion is not hinge bending. Consider, for example, the case where domain β performs a shear motion. In this case we will find that $\Delta\theta = 0$ and we will know that domain β is translating with respect to domain α. If the interdomain screw axis is located somewhere inside one of the domains or outside the body of the protein, again we will know that this is not consistent with a hinge bending motion. Such an interdomain screw axis will reveal that domain β is not only rotating with respect to the interdomain region but also translating in a direction perpendicular to the rotation axis. Only when the screw axis is located between the two domains, and near to connecting regions, can we assume that the parameters we have determined are those describing a physical hinge axis. We will return to this point later where we give a more precise definition of the various types of axes.

Method to Determine Rotation Vectors and Other Parameters of Screw Motion

Both for the determination of the dynamic domains and for the determination of the interdomain screw axis we need to have a method to determine rotation vectors from two conformations. One common method is to diagonalize the rotation matrix from the best fit of these two conformations. In our previous analysis we developed the method of determining the rotation vector of a residue by calculating the curl of the displacement vector field at that residue. Here again we use this method to calculate the rotation vectors of residues, main-chain segments, and of whole domains. In the case of the motion of domain β relative to domain α, this is simply done by evaluating the curl of displacement vector field of the $\Delta\mathbf{d}_i^{ext}$.

The curl is a differential operator that is applied to a vector field. The method we use to evaluate this quantity is described in the previous paper.[13] Although our method of calculating the curl is based on a linear approximation, because we have removed all nonlinearity by performing best fits, our method gives exactly the same results as the usual method of diagonalizing the rotation matrix. Note that the curl needs only to be calculated at one residue of domain β as it has the same value over the whole domain now that internal variations have been removed. Also, the curl can be evaluated directly at any residue without having to perform best fits as was done in the previous paper to calculate the rotation–orientation matrix from the normal mode analysis results. Applying the curl to Equation (4) and using elementary vector analysis one can show that:

$$\nabla \times \Delta\mathbf{d}^{ext} = (2 \sin \Delta\theta)\,\mathbf{n} \tag{5}$$

Once **n** has been determined, $\Delta\tau_{\|}$ can be found using:

$$\Delta\tau_{\|} = \Delta\mathbf{d}^{ext} \cdot \mathbf{n} \tag{6}$$

and finally the location of the interdomain screw axis can be determined. Now we have determined all six parameters for the interdomain screw axis for the motion of domain β relative to domain α.

The length of the rotation vector, $\Delta\theta$, is given by the inverse sine of half of the amplitude of the vector determined by calculating the curl. As we cannot directly distinguish between an angle of $\Delta\theta$ or $180-\Delta\theta$, we initially take $\Delta\theta$ to be the lower value and check that it can reproduce the displacement vector field. If this is not the case we take the $180-\Delta\theta$ value.

Method to Determine Residues Involved in Interdomain Motion

The method to determine dynamic domains assigns residues to one domain or another. However, it

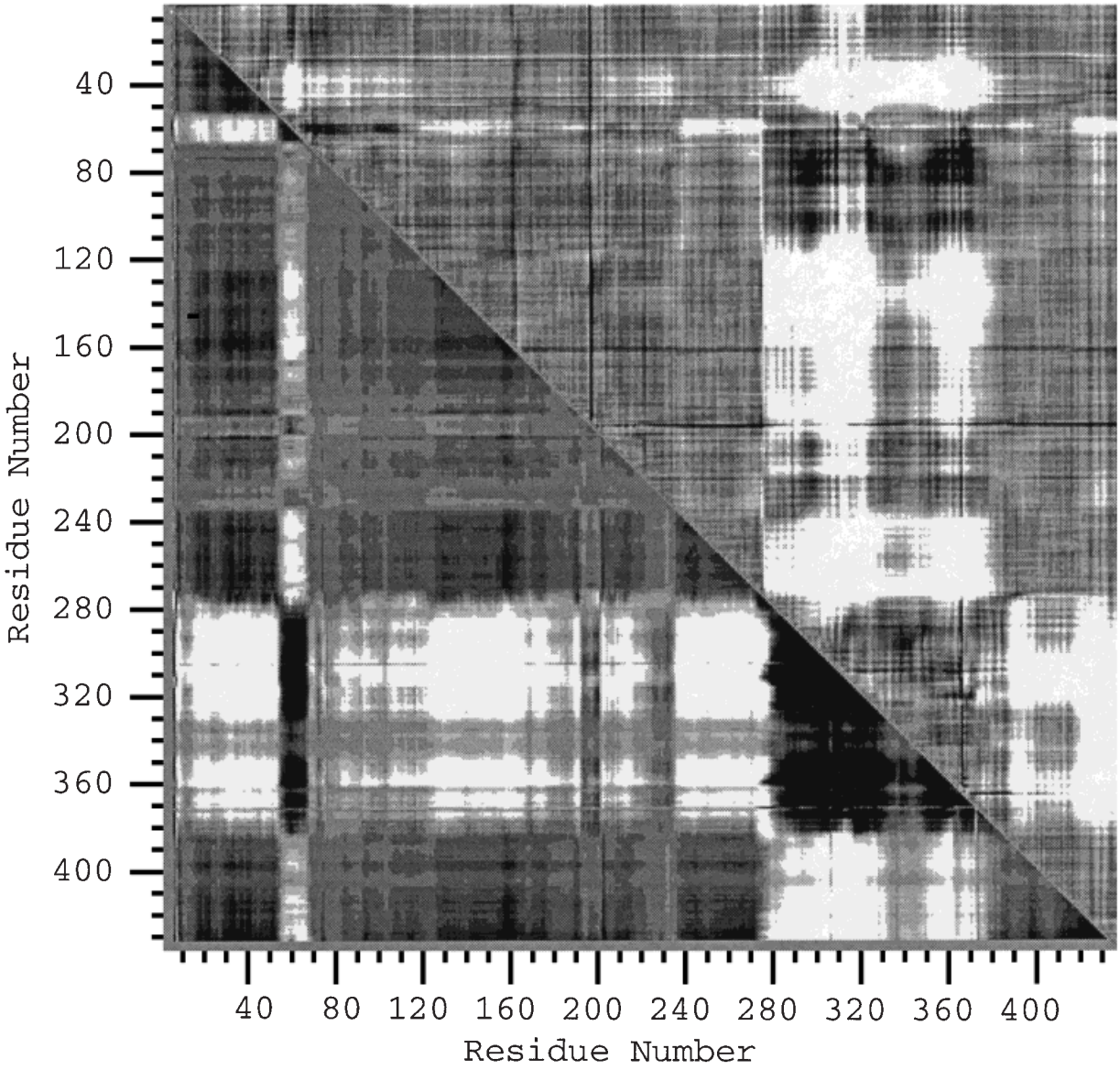

Fig. 2. **Lower half:** Gray-scale plot of rotation–discrimination matrix for citrate synthase. A whole protein best fit of the open and closed conformations of citrate synthase (PDB codes: 1CTS and 2CTS) is made to determine the atomic displacement vectors. Corresponding main-chain segments from a sliding window of 11 residues are then superimposed in order to determine the displacement vectors representing the rigid body displacement of each segment. Rotation vectors are calculated for each segment and assigned to the residue at the center of the window. Inner products are calculated for all pairs and normalized by the standard deviation of the distribution. The normalized inner product of the *i*th residue with the *j*th residue is plotted as a gray scaled dot at coordinate *i,j*. The black dots indicate normalized inner products $>$ 1.0, white dots normalized inner products <-1.0. **Upper half:** Difference–distance matrix plot for citrate synthase. Difference–distance distribution has been normalized by the standard deviation of the distribution. Black dots indicate that the normalized difference in distance between the two CA atoms in the closed conformation and the open conformation is $>$1.0, a white dot indicates it is <-1.0, gray dots indicate in-between values.

is clear that some of the residues in the connecting regions between two domains do not really belong to either domain but are involved in the interdomain motion. Here we present a method to determine these interdomain residues. The method is similar to that described in the previous paper, but now we make it more objective.

We slide a window through the whole of domain α, and then through the whole of domain β. At each window position we determine the rotation vector of the corresponding main-chain segment and calculate the component of the rotation vector along the direction of the interdomain screw axis for this domain pair. The residue at the center of the window is assigned the rotation vector component of the main-chain segment corresponding to the window. In domain α the average value of the component should be about zero as domain α is the one fixed in space. In

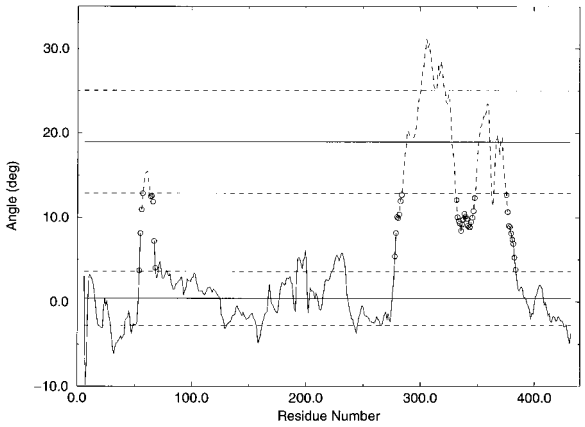

Fig. 3. Determination of residues involved in the interdomain motion for citrate synthase. A sliding window 11 residues in length generates main-chain segments whose rotation vectors are determined. The component of these rotation vectors along the interdomain screw axis is calculated and plotted against the residue at the center of the window. Those from domain α are indicated by the continuous line, those from domain β by the broken line. The horizontal continuous lines indicate the averages of the rotation vector components of the two domains, the broken horizontal lines indicate the averages plus and minus their corresponding standard deviations. The residues involved in the interdomain motion are indicated by circles. They are those at the interdomain boundaries plus those neighboring residues, with rotation vector components above one standard deviation from the average in the case of the domain α residues, and below one standard deviation from the average in the case of domain β residues.

domain β the average value for the component should be about equal to the rotation angle of the whole of domain β. This means that in the interdomain region we should see a transition. We first calculate the averages and standard deviations of the rotation vector components in domains α and β. The residues involved in the interdomain motion are taken to be those at the interdomain boundaries, plus those neighboring residues that deviate at least one standard deviation from the average of the domain to which they belong.

After determining the interdomain residues we could go back and eliminate these from the analysis to determine the interdomain screw axes. So far our analysis indicates that this has little effect on the location of the interdomain screw axis.

Effective Hinge Axis, Closure Axis, and Twist Axis

As mentioned above it is possible for an interdomain screw axis to be located anywhere in space. If there is translation of domain β with respect to the interdomain region, and this translation is perpendicular to the rotation axis, then the interdomain screw axis will not be located at the interdomain region. If, however, domain β only rotates with respect to the interdomain region due to real physical constraints imposed by the interdomain residues, then the interdomain screw axis will be located at these interdomain residues, and the term hinge axis could be reasonably applied. However, the term "hinge axis" is not precisely defined in the literature, and has been often used with regards to specific dihedral angle rotations of interdomain residues irrespective of whether they are located near the interdomain screw axis or not. In order to avoid confusion we define an "effective hinge axis" to be an interdomain screw axis that passes near *any* of the interdomain residues. In fact, we make this more precise here, and call an axis an effective hinge axis only if it passes within 3 Å of the CA atom of any of the residues shown to be involved in the interdomain motion by the method described above.

As before we also characterize effective hinge axes into two extreme types: those parallel to the line joining the centers of mass of the two domains, and those perpendicular to this line. The former are called twist axes, the latter, closure axes. Any axis can be decomposed into components parallel or perpendicular to this line and a percentage measure of the degree of its closure or twist motion can be defined from the square of the projection on either axis.[13]

RESULTS AND DISCUSSION

Citrate Synthase

The structural data bank contains both an open structure (PDB code: 1CTS) and a closed structure (2CTS) of citrate synthase.[8] The enzyme functions as a homodimer. Although the monomer has a fairly obvious two-domain structure, the exact demarcation between the domains is unclear and therefore provides a good test for the method. The conformational change has previously been considered to be the rigid body rotation of a small domain about a large domain by 18 degrees with hinge points at His 274 and Gly 275.[8,9]

In order to demonstrate the soundness of the basic idea of using rotation vectors to identify domains, the method is first compared to a commonly used method to determine domains: the difference–distance matrix method.[10] The difference–distance matrix method is based on the idea that intradomain atomic distances should be largely unchanged in comparison to interdomain atomic distances. However, as was mentioned in the Introduction, depending on the location of the hinge axis, even interdomain atomic distances may also remain unchanged, and the plot may be unclear as regards the demarcation of the domains. Figure 2 shows the difference–distance matrix plot in comparison with a plot of the "rotation–discrimination matrix." The rotation–discrimination matrix differs from the rotation–orientation matrix introduced in the previous paper in that we take the inner products between the full rotation vectors instead of their unit vectors. The difference in clarity between the plots of the rotation–discrimination matrix and the difference–distance

TABLE I. Domain Motion Data for Citrate Synthase and Mutant T4-Lysozyme

Protein transition	Citrate synthase:	T4-lysozyme				
	Unliganded—liganded	W–A	A–B	A–B	A–B	C–D
Domain α (residues)	1–55, 67–279, 335–337, 342–345, 378–437	12–72	14–71	14–71	87–162	12–70
Domain β (residues)	55–66, 280–334, 338–341, 346–377	1–11, 73–162	1–13	72–86	72–86	1–11, 69–162
Angle of rotation (deg)	19.2	9.1	18.3	17.3	13.2	5.3
Translation along axis (Å)	−0.01	0.17	−0.7	−0.14	−0.5	−0.1
Residues involved in motion	54–57, 64–68, 278–284, 332–348, 376–384	11–12, 70–73	12–15	65–75	83–92	11–12, 65–71
Residues within 3 Å of axis*	56, 65, 234, 381, 382, 384	11, 70	13, 14, 29, 67, 68	44–46, 70, 71, 92, 93	82, 84, 85, 88, 98, 99, 160, 161	67, 70–73
% Closure motion*	85	98	94	39	51	96

*Based on first structure.

matrix is obvious. The rotation–discrimination matrix plot has a clear blocking structure barely distinguishable in the difference–distance plot. Although we have used a sliding window 11 residues in length here, the variation in window length has little effect on the location of the boundaries visible in this plot, but mainly influences its overall clarity.

Before application of the full process to determine the dynamic domains in citrate synthase, we must not only choose a window length, but also decide on a minimum domain size. Obviously the minimum domain size should be adjusted according to the size of the protein. For the citrate synthase monomer, which has 437 residues, we choose a minimum domain size of 40 residues. Table I summarizes the results of the full application of the methodology to citrate synthase, and Figure 3 shows how the interdomain residues are determined from the components of the rotation vectors of the main-chain segments along the direction of the interdomain screw axis.

The interdomain screw axis of citrate synthase is indeed an effective hinge axis as defined above. Figure 4 shows the open structure, colored to show the dynamic domains, the residues involved in the interdomain motion, and the effective hinge axis itself. A previously unnoticed feature of the motion is that the two-stranded antiparallel beta sheet between residues 56 and 64, belongs dynamically to the small domain, although it is connected to the large domain. The effective hinge axis passes between the two stands of the beta sheet, near residues 54–57 and 64–68, and near residues 380–384, all shown to be involved in the interdomain motion. It is tempting to think that the binding of the sheet to the small domain helps determine the direction of the effective hinge axis and so control the interdomain motion. In both the open and closed structure, the carbonyl oxygen of Pro60 makes a hydrogen bond with the main-chain amide hydrogen of Arg324 located in the small domain. In addition, in the open conformation salt bridges are formed between Arg324 and Asp61, and Lys325 and Glu62. In the closed structure,

Fig. 4. Citrate synthase in open conformation. The two dynamic domains as determined by the procedure are colored cyan (α) and red (β), and the residues found to be involved in the domain motion are colored green. The effective hinge axis is colored cyan to indicate that both it and the cyan colored α domain are fixed in space, with the arrow direction indicating the direction of rotation of the small domain by the right-hand rule. Note that the red beta sheet although connected to the α domain is dynamically part of the β domain. The axis passes between the two strands of this sheet. For details see Table I. Figure created using ''Molscript''[33] and ''Raster3D''.[34]

Fig. 5. **Top:** The A conformation of the M6I mutant of T4 lysozyme from PDB file 150L, showing four dynamic domains according to the A–B transition. The N-terminal domain is colored cyan, the C-terminal domain, yellow. The red domain is located at the N-terminal helix and the orange domain at the C terminus of the long interdomain helix. The green regions indicate residues involved in the interdomain motions. The horizontal cyan arrow indicates how the orange domain moves relative to the blue domain again by the right-hand rule, and the other cyan arrow indicates how the red domain moves relative to the blue. The yellow arrow indicates the motion of the orange domain relative to the yellow domain. All axes are effective hinge axes. For details see Table I. **Bottom:** Rotation vectors in the three-dimensional rotation space, colored as in the upper figure. The clustering algorithm determines the dynamic domains from this distribution of points. Figure created using ''Molscript''[33] and ''Raster3D''.[34]

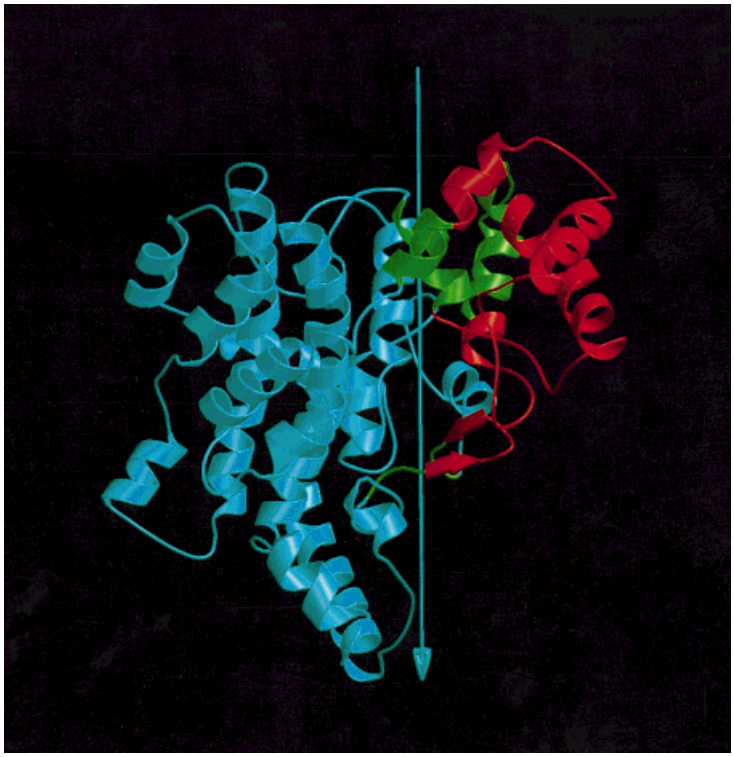

Fig. 4.

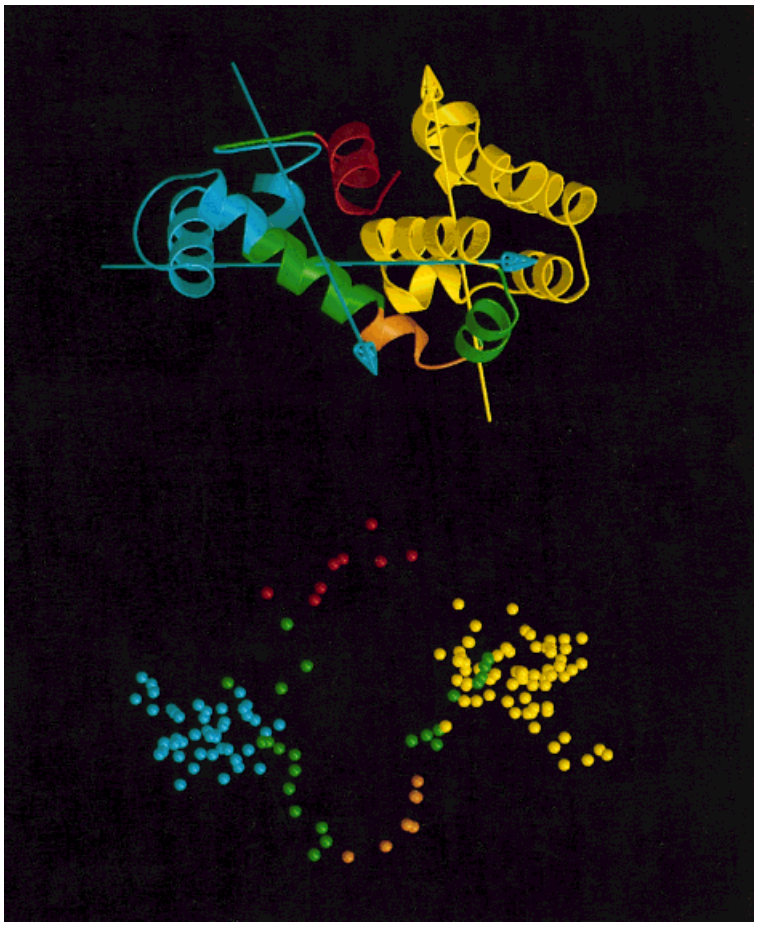

Fig. 5.

however, the salt bridge is between Asp61 and Lys325. Any mutation that frees this beta sheet from its noncovalent binding to the small domain would be expected to have a profound effect on the dynamics, and consequently on the function of citrate synthase. Unfortunately, we have no experimental information on the effects of mutations at these sites as mutation studies made on citrate synthase have been mainly targeted to active site residues.[30]

As mentioned at the beginning of this section it has previously been thought that the hinge point is located at residues 274 and 275.[8,9] Although these residues may be involved in the interdomain rotation they are in fact roughly 15 Å from the effective hinge axis and the term "hinge point" is surely inappropriate. Gerstein and colleagues[2] in their review of the structural mechanisms behind domain motions classify citrate synthase as an example of a protein undergoing domain closure through shear motions. They are referring to the detailed structural rearrangements occurring locally in the interdomain region. These shear motions obviously produce an effective hinge axis.

Mutant T4 Lysozyme

For the M6I mutant of T4 lysozyme five different structures are known, one identical to the wild-type structure and termed here "W", and four structures from a single crystal, termed "A," "B," "C," and "D" as they occur in the PDB file, 150L.[3] The W conformation is the most closed conformation, with A, B, C, and D being progressively more open.

Although we could choose any of a possible 10 pairs, we chose to analyze only those nearest neighbor pairs, that is, the W–A, A–B, B–C, and C–D transitions. In this way we will see a picture of the motion from the most closed conformation to the most open.

T4 lysozyme is a 164-residue protein, and we have chosen a window of length 7 residues and a minimum domain size of 10 residues for the analysis. Table I gives all the details concerning the results, apart from the B–C transition for which no domain pairs had χ values greater than 1.0. The most striking result is that, although for the W–A and C–D transitions the protein has the expected two domains, for the A–B transition it has four dynamic domains. One of the extra domains is the N-terminal helix, the other is located at the C terminus of the long interdomain helix. Figure 5 shows rotation points in the rotation space for the A-B transition and the A structure colored to illustrate the domains, the interdomain regions, and the interdomain screw axes.

We can now describe the complicated motion that T4 lysozyme performs in going from the closed to open conformation implied by these five conformational snapshots. It begins with a closure motion in the W–A transition. It then fractures into a four-domain protein for the A–B transition. After making a largely internal rearrangement in the B–C transition, it finally returns to being a two-domain protein in a final closure motion in the C–D transition. In all cases the interdomain screw axes are effective hinge axes. Residues 11, 12, and 13 appear to be intimately involved in most of these transitions. Again we expect a mutation in this region, and in particular at Gly12, to have a profound effect on the dynamics and function of this enzyme.

Previous studies on T4 lysozyme include one on the same five structures studied here,[3] another on a mutant in which the isoleucine at residue 3 is replaced by a proline,[4] and a further study on a number of mutants.[5] Our study is limited in breadth in comparison to these latter two works, but we have indeed found new results with our method despite this, in particular the emergence of new domains in the A–B transition. This difference is of course due to the difference in the methods used. Other methods used to analyze T4 lysozyme include the difference-distance method, which, as we have seen above in the case of citrate synthase, is rather limited. Another, is the displacement plot method[3,4] in which two conformations are best fitted on the basis of a selected set of CA atoms and the displacements of all CA atoms between their positions in the two conformations are plotted against residue number. The results of this analysis are overwhelmingly biased by the choice of atoms to be fitted. The one-dimensional plot is very difficult to interpret, as it is not clear whether displacements arise from true conformational change or are simply related to the distance from some axis of rotation in a rigid body. Another method used specifically in the determination of regions involved in hinge bending requires the fitting of corresponding backbone segments from a sliding window, and calculating the root mean square deviation, rmsd, between the two fitted conformations.[31,32] The idea is that hinge bending regions will be kinked and show larger RMSDs than the intradomain regions. This may work in the ideal case of two rigid bodies connected by a hinge region, but could fail if there are relatively large intradomain deformations of a short-range nature, and where the hinge bending angle is caused by the accumulation of small deformations over a large hinge bending region.

CONCLUSIONS

The methodology introduced in a previous paper to analyze domain motions in proteins has been improved and extended. It has been tested on citrate synthase and a mutant of T4 lyszoyme. The method determines domains, any existing hinge axes, and residues involved in the interdomain motion. It is applicable in any case where one has more than one conformation of a protein, or equivalently, one conformation plus a set of displacement vectors as in normal mode analysis, or after application of an

essential dynamics analysis or quasi-harmonic analysis of a molecular dynamics simulation. The basic idea behind the methodology is that domains can be identified by their differing rotational properties. There is an overall consistency to the methodology, since the methods used to identify the domains are also used to determine possible hinge axes and residues involved in the interdomain motion.

The search in the three-dimensional rotation space for clusters of rotation vectors can be generalized to the search in the six-dimensional space containing all the six parameters governing the rigid body motion of residues or main-chain segments. Such an implementation will overcome the possible problem caused if domains happen to have identical rotation vectors.

Our method to determine dynamic domains does *not* require the domains to be rigid, but is based on the ratio of interdomain motion to intradomain motion. In this way an objective criterion can be applied to distinguish interdomain fluctuation from intradomain fluctuation.

The determination of the interdomain screw axes is based on the theorem of Chasles from rigid body kinematics. The interdomain screw axes can be located anywhere in space depending on the nature of the interdomain motion. This fact allows us to give a more rigorous definition of a hinge axis. We define an interdomain screw axis to be an effective hinge axis only if it passes near residues shown to be involved in the interdomain motion. The effective hinge axes are further classified by decomposing them into closure and twist axes.

The methods have been tested on two proteins that have already been studied in some depth by others, namely, citrate synthase and T4 lysoyme. In both cases we have gained new insights into their motion. For citrate synthase the beta sheet, although connected to the large domain, is dynamically part of the small domain. This fact cannot be inferred by the commonly used difference–distance matrix plot. The effective hinge axis passes between the two strands of the sheet and it appears, therefore, that it is an essential component of the mechanism controlling the interdomain motion. The sheet is bound to the small domain through hydrogen bonds and salt bridges involving residues Pro60, Asp61, Glu62, Arg325, and Lys325. Mutations of these residues that free the sheet from the small domain would be expected to have a profound effect on the dynamics and function of this enzyme.

For T4 lysozyme, two new domains appeared in the A–B transition that were not seen in the other three transitions. If we accept that the X-ray conformers of T4 lysozyme represent possible snapshots of conformers accessible to a single protein in solution,[5–7] then the number of dynamic domains is not a constant of the motion. This should not surprise us because interactions between parts of a protein will change depending on the protein's conformation. For example, hydrogen bonds between the surfaces of two parts of a protein may be sufficient for them to behave as a single dynamic domain, but if the hydrogen bonds are overtaken by water molecules, these two parts may very well behave as two separate dynamic domains.

As in the case of citrate synthase, we can identify a residue that we expect to be important in controlling the motion of T4 lysozyme, namely Gly12.

The results show that this methodology can detect important features that existing methods have failed to find. The method is quite general and objective and can be automated without human intervention. Further analyses of a larger class of proteins will hopefully reveal principles of interdomain motion in proteins and their functional significance.

We intend to make the program "dyndom" that implements the methodology described in this article available in the near future. Please see information at internet address: http://rugmd0.chem.rug.nl/~steve/steve.html

ACKNOWLEDGMENTS

S.H. is a recipient of a BIOTECH fellowship from the European Union. We thank N. Seki and Frans van Hoesel for helpful discussions.

REFERENCES

1. Schulz, G.E. Domain motions in proteins. Curr. Opin. Struct. Biol. 1:883–888, 1991.
2. Gerstein, M., Lesk, A.M., Chothia, C. Structural mechanisms for domain movements in proteins. Biochemistry 33:6739–6749, 1994.
3. Faber, H.R., Matthews, B.W. A mutant T4 lysozyme displays five different crystal conformations. Nature 348:263–266, 1990.
4. Dixon, M.M., Nicholson, H., Shewchuk, L., Baase, W.A., Matthews, B.W. Structure of a hinge-bending bacteriophage T4 lysozyme mutant Ile3-Pro. J. Mol. Biol. 227:917–933, 1992.
5. Zhang, X., Wozniak, J.A., Matthews, B.W. Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme. J. Mol. Biol. 250:527–552, 1995.
6. Mchaourab, H.S., Oh, K.J., Fang, C.J., Hubbell, W.J. Conformation of T4 lysozyme in solution. Hinge-bending motion and the substrate-induced conformational transition studied by site-directed spin labeling. Biochemistry. 36:307–316, 1997.
7. de Groot, B.L., Hayward, S., van Aalten, D.M.F., Amadei, A., Berendsen, H.J.C. Domain motions in bacteriophage T4 lysozyme: A comparison between molecular dynamics and crystallographic data. Proteins (in press).
8. Remington, S., Wiegand, G., Huber, R. Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 angstroms resolution. J. Mol. Biol. 158:111–152, 1982.
9. Wiegand, G., Remington, S.J. Citrate synthase, structure, control, and mechanism. Annu. Rev. Biophys. Biophys. Chem. 15:97–117, 1986.
10. Nishikawa, K., Ooi, T., Isogai, Y., Saito, N. Tertiary structure of proteins. I. Representation and computation of the conformations. J. Phys. Soc. (Jpn) 32:1331–1337, 1972.
11. Gerstein, M., Anderson, B.F., Norris, G.E., Baker, E.N., Lesk, A.M., Chothia, C. Domain closure in lactoferrin. J. Mol. Biol. 234:357–372, 1993.

12. Huang, E.S., Rock, E.P., Subbiah, S. Automatic and accurate method for analysis of proteins that undergo hinge-mediated domain and loop movements. Curr. Biol. 3:740–748, 1993.
13. Hayward, S., Kitao, A., Berendsen, H.J.C. Model-free methods to analyze domain motions in proteins from simulation: A comparison of a normal mode analysis and a molecular dynamics simulation of lysozyme. Proteins 27: 425–437, 1997.
14. Go, N., Noguti, T., Nishikawa, T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc. Natl. Acad. Sci. U.S.A. 80:3696–3700, 1983.
15. Brooks, B., Karplus, M. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. Proc. Natl. Acad. Sci. U.S.A. 80:6571–6575, 1983.
16. Levitt, M., Sander, C., Stern, P.S. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. Int. J. Quant. Chem. 10:181–199, 1983.
17. Case, D.A. Normal mode analysis of protein dynamics. Curr. Opin. Struct. Biol. 4:285–290, 1994.
18. Hayward, S., Go, N. Collective variable description of native protein dynamics. Annu. Rev. Phys. Chem. 46:223–250, 1995.
19. Levy, R.M., Srinivasan, A.R., Olson, W.K., McCammon, J.A. Quasi-harmonic method for studying very low frequency modes in proteins. Biopolymers 23:1099–1112, 1984.
20. Kitao, A., Hirata, F., Go, N. The effects of solvent on the conformation and collective motions of protein: Normal mode analysis and molecular dynamics simulations on melittin in water and vacuum. Chem. Phys. 158:447–472, 1991.
21. Hayward, S., Kitao, A., Hirata, F., Go, N. Effect of solvent on collective motions in globular proteins. J. Mol. Biol. 234:1207–1217, 1993.
22. Garcia, A.E. Large-amplitude nonlinear motions in proteins. Phys. Rev. Lett. 68:2696–2699, 1992.
23. Amadei, A., Linssen, A.B.M., Berendsen, H.J.C. Essential dynamics of proteins. Proteins 17:412–425, 1993.
24. Noguti, T., Go, N. Dynamics of native globular proteins in terms of dihedral angles. J. Phys. Soc. (Jpn) 52:3283–3288, 1983.
25. McLachlan, A.D. Gene duplications in the structural evolution of chymotrypsin. J. Mol. Biol. 128:49–79, 1979.
26. Hartigan, J.A. "Clustering Algorithms." New York: Wiley, 1975.
27. Goldstein, H. "Classical Mechanics," 2nd edit. Reading, MA: Addison Wesley World Student Series, 1980.
28. Muirhead, H., Cox, J.M., Mazzarella, C., Perutz, M.F. Structure and function of haemoglobin III: A three-dimensional Fourier synthesis of human deoxyhaemoglobin at 5.5A resolution. J. Mol. Biol. 28:117–156, 1967.
29. Colonna-Cesari, F., Perahia, D., Karplus, M., Eklund, H., Brändén, C., Tapia, O. Interdomain motion in liver alcohol dehydrogenase: structural and energetic analysis of the hinge bending mode. J. Biol. Chem. 261:15273–15280, 1986.
30. Evans, C.T., Kurz, L.C., Remingtion, S.J., Srere, P.A. Active site mutants of pig citrate synthase: Effects of mutations on the enzyme catalytic and structural properties. Biochemistry 35:10661–10672, 1996.
31. Remington, S.J., Matthews, B.W. A systematic approach to the comparison of protein structures. J. Mol. Biol. 140:77–99, 1980.
32. Birktoft, J.J., Rhodes, G., Babszak, L.J. Refined crystal structure of cytoplasmic malate dehydrogenase at 2.5 A resolution. Biochemistry 28:6065–6081, 1989.
33. Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Cryst. 24:946–950, 1991.
34. Merrit, E.A., Murphy, M.E.P. Raster3D Version 2.0. A program for photorealistic molecular graphics. Acta Cryst. D50:869–873, 1994.