PROTEINS: Structure, Function, and Genetics 29:1–14 (1997)

RESEARCH ARTICLES

Protein Domain Movements: Detection of Rigid Domains and Visualization of Hinges in Comparisons of Atomic Coordinates

Willy Wriggers and Klaus Schulten*
Department of Physics and Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois

ABSTRACT **The activity of many proteins induces conformational transitions by hinge-bending, which involves the movement of relatively rigid parts of a protein about flexible joints. We present an algorithm to identify and visualize the movements of rigid domains about common hinges in proteins. In comparing two structures, the method partitions a protein into domains of preserved geometry. The domains are extracted by an adaptive selection procedure using least-squares fitting. The user can maintain the spatial connectivity of the domains and filter significant structural differences (domain movements) from noise in the compared sets of atomic coordinates. The algorithm subsequently characterizes the relative movements of the found domains by effective rotation axes (hinges). The method is applied to several known instances of domain movements in protein structures, namely, in lactoferrin, hexokinase, actin, the extracellular domains of human tissue factor, and the receptor of human growth factor. The results are visualized with the molecular graphics package VMD (Humphrey et al., J. Mol. Graphics 14(1):33–38, 1996). Applications of the algorithm to the analysis of conformational changes in proteins and to biomolecular docking are discussed. Proteins 29:1–14, 1997.** © 1997 Wiley-Liss, Inc.

Key words: protein architecture; hinge-bending; lactoferrin; hexokinase; actin; human tissue factor; human growth factor

INTRODUCTION

Proteins are empowered by their intrinsic flexibility to a wide spectrum of biochemical function in catalysis, regulation, protein assembly, and cell motility. Advancing from the premise that the function of proteins, which is controlled by physiologic agents, is associated with the spatial arrangement of structural elements, researchers have addressed aspects of protein architecture and domain movements for many years.[1] A number of recent studies demonstrated that many conformational changes in proteins can be characterized schematically as rigid-body movements of segments of preserved structure (for an exhaustive review, see ref. 2). Of special interest, in this respect, are hinge-bending movements, in which rigid domains are connected by flexible joints that tether the domains and constrain their movement. Hinge-bending is believed to allow an induced fit of molecular surfaces in protein assembly and ligand docking.[3,4] In a conformational analysis of protein structures, the identification of hinge axes and their corresponding rotation angles permits a useful representation of protein domain movements.

We have developed the algorithm *Hingefind*** to detect effective rotation axes when structures of identical or related proteins are compared. For a given pair of protein conformations two tasks arise in the comparison: (1) extracting the rigid (i.e., geometrically similar) domains from the sets of three-dimensional coordinates; (2) visualizing the relative movements of the domains identified as rotations about hinges.

Methods to identify similar substructures in proteins have been applied before in studies of protein architecture[1] and protein evolution.[6] Early algorithms were based on a computationally demanding search in rotational space[6,7] or on a comparison of structural fragments between the structures.[8] The

**A documented version of *Hingefind* can be obtained by anonymous ftp to ftp.ks.uiuc.edu in the directory pub/hingefind or on the World Wide Web at URLftp://ftp.ks.uiuc.edu/pub/hingefind/hingefind.html. The program is currently implemented in X-PLOR[5] and VMD script languages. VMD is a free molecular graphics program (Figs. 10–13), available by anonymous ftp to ftp.ks.uiuc.edu in the directory pub/mdscope/vmd.

*Correspondence to: Klaus Schulten, Department of Physics and Beckman Institute, University of Illinois at Urbana-Champaign, 405 North Mathews Avenue, Urbana, IL 61801.

Received 30 September 1996; Accepted 11 March 1997

performance of rotational search techniques has since been improved by means of a prescreening of the structures using the sequence of dihedral angles.[9] Another class of comparison methods evolved from sequence alignment (dynamical programming) methods,[10–13] which proved to be more tolerant of sequence insertions and deletions. Yet, other methods use interatomic distances instead of Cartesian coordinates[10,12,14,15] or compare secondary structure elements.[13,16,17] More recently, pattern detection algorithms[14,18] have been optimized to search for geometric fragments in a database.[14] Present-day advancements in structure comparison allow one to search for similar substructures of a protein in the entire database of known protein structures.[15,19] Given these already existing techniques that permit a homology-based alignment of protein sequences, we limit the present study to cases in which for the substructures sequence alignments in terms of pairs of corresponding residues are available.

Knowledge of the corresponding residues simplifies the problem of optimal superposition in three dimensions. Procedures for optimizing the least-squares superposition of protein domains by excluding poor fitting residues have been used for a long time.[20] Lesk[1] has formulated a sieving routine that minimizes the rms deviation of a domain by subsequent elimination of atoms that lie far apart in the superposition. A computationally more demanding method, which uses difference-distance matrices to identify rigid domains, has recently been introduced by Nichols et al.[22] We will show that these techniques are unsuitable for the intended investigation of multiple rigid domains and their movements. We have developed, therefore, a new routine, termed *adaptive selection*, which fully partitions a protein into geometrically preserved domains.

The second task at hand is the identification of hinge axes. Rigid-body movements of the domains relative to each other can be determined either by least-squares superposition[23,24] or using the domain's principal axes of inertia.[25] Our approach is to reduce rigid-body transformations to rotations about effective rotation axes. Such a reduction of the movement is not possible in general. Proceeding from the premise that the movement of a domain relative to a reference domain is constrained by a tethering joint, we will demonstrate how a hinge axis can be nevertheless identified.

We begin our study by introducing the implementation of the algorithm. We then discuss the accuracy of the hinge-bending movements identified by *Hingefind*. Finally, we determine and visualize domain movements in structures of lactoferrin, hexokinase, and actin, as well as the extracellular domains of human tissue factor and of the receptor of human growth factor.

METHODOLOGY

In this section, we address the partitioning method that extracts the rigid domains by iterative superposition of the protein structures. We then discuss the dependence of the domain size on the tolerance ϵ for coordinate sets perturbed by Gaussian white noise of standard deviation σ. We formulate an equation that will allow us to estimate the local noise level σ in comparisons of authentic protein structures. Finally, we introduce a geometric technique to determine the effective rotation axes of the motions exhibited by the rigid domains.

Extracting Rigid Domains

We seek to identify rigid domains of preserved geometry in a comparison of two protein structures. A measure of geometric similarity should permit a certain imprecision of the atomic coordinates. For this purpose, we superimpose a given subset of atoms using the Kabsch least-squares method,[23,24] which is, for a sufficiently large subset, insensitive to local fluctuations in the position of atoms.

Let $\vec{x}_n'$ and $\vec{y}_n (n = 1, 2, \ldots, N)$ be two given vector sets (representing atomic coordinates of a protein in two conformations) and s_n the weight corresponding to each n. We use a reduced representation of the protein geometry in terms of its C_α atoms that we consider sufficient for protein structure comparisons. For this special case, mass-weighing is not necessary. The weights $s_n \in \{0, 1\}$ are membership parameters that describe the selection of a subset of atoms for superposition. The Kabsch least-squares method[23,24] then fits $\vec{x}_n$ to $\vec{y}_n$, yielding a vector set $\vec{x}_n'$, which minimizes the function $\Sigma_n\, s_n(\vec{x}_n' - \vec{y}_n)^2$. The fitted $\vec{x}_n'$ is related to $\vec{x}_n$ by the rigid-body transformation

$$\vec{x}_n' = \mathbf{U}\vec{x}_n + \vec{v} \tag{1}$$

where $\mathbf{U} = (\mathbf{u_{ij}})$ is an orthogonal matrix and $\vec{v} = \Sigma_n\, s_n(\vec{x}_n' - \vec{x}_n) = \Sigma_n\, s_n(\vec{y}_n - \vec{x}_n)$ connects the centroids of the selected subsets $\vec{x}_n$ and $\vec{x}_n'$ (or $\vec{y}_n$). The geometric conformance of individual atoms with the subset can be measured by the distance δ_n between corresponding atoms in the two structures after superposition

$$\delta_n = \|\vec{x}_n' - \vec{y}_n\|. \tag{2}$$

The idea to rank atoms by their geometric conformance with a large part of the protein is the basis of Lesk's sieving routine, which eliminates poor matching residues to minimize the rms deviation between compared protein structures.[1,26,27] The Lesk method allows one to extract a well-fitting substructure in situations in which only a single preponderant domain exists (the rigid core of the protein). However, the method is unsuitable in situations with multiple rigid domains: The Lesk algorithm assumes that the

initial superposition will provide a good measure of conformance, but there is no a priori knowledge about the number and orientation of rigid domains. The measure of conformance depends on the atom subset chosen for superposition; the Lesk method may, thus, eliminate atoms that are part of a well-preserved domain if they happen to match poorly in the initial superposition.

To solve the multiple domain problem, the atoms selected for superposition and the measure of conformance need to be optimized simultaneously. One possible way to optimize the selection of atoms was described in an early version of the sieving method,[21] in which superpositions of domains were started at known well-fitting subsets and, subsequently, extended to include additional unknown well-fitting atoms. In the following, we present a new algorithm that combines and extends features of the existing sieving methods[1,21,26,27] to automatically find rigid domains without a priori knowledge of suitable subsets for superposition.

Well-fitting subsets are extracted iteratively in steps $t = 0, 1, 2, \ldots$. A search for a new rigid domain is started by choosing an initial set of membership parameters, $s_n(0)$, representing a localized seed subset of C_α atoms

$$s_n(0) = \begin{cases} 1 & \text{if } \|\vec{x}_n - \vec{x}_\lambda\| < \rho \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

where $\lambda \in \{1, 2, \ldots, N\}$ is an atom index and ρ is the radius of the initially selected seed subset. We show in the Performance section that for practical applications, λ can be randomly chosen and a radius $\rho = 15$ Å is suitable for comparisons involving C_α atoms.

The atom composition of this subset, described by the membership parameters $s_n(t)$ $(t = 0, 1, 2, \ldots)$, will then adapt to superpositions of the structures in an iterative two-step process:

1. Superimpose the structures using $s_n(t-1)$ and compute $\delta_n(t-1)$ from Equations 1 and 2.
2. Add new atoms, whose conformance meets a tolerance ϵ, to the subset; exclude atoms from the subset that do not meet the tolerance due to changes in the superposition, i.e., choose

$$s_n(t) = \begin{cases} 1 & \text{if } \delta_n(t-1) < \epsilon \\ 0 & \text{otherwise} \end{cases}. \quad (4)$$

Figures 1 and 2 demonstrate how this *adaptive selection* routine has been implemented. The two steps, Equations 1 and 2, are repeated until the subset converges (convergence loop, Fig. 1). Once a rigid domain is found, it is subtracted from the set of atoms used in the comparison. Searches for rigid domains are repeated for the remaining atoms until the protein is completely partitioned into well-fitting substructures.

Two options are available for the mode of partitioning (Fig. 2): "slow" mode partitioning, which maintains the spatial connectivity of the changing segment, and "fast" mode partitioning, which is significantly faster while sacrificing the connectivity of the segments.

Effect of Noise and Changing Tolerance

In Figure 3, we illustrate the ϵ-dependence of the size of the largest found domain ("fast" partitioning) in comparisons of the structure of a protein (lactoferrin, 691 amino acid residues) to structures that have been obtained by adding Gaussian noise of standard deviation σ to each of the atomic coordinates of the reference structure. Let Δx_i^j be the i-th component $(i = 1, 2, 3)$ of the noise-related displacement of the j-th atom from its original position $(j = 1, \ldots, N)$. Assuming a large number of atoms N, the following formulae for the atom displacements hold by virtue of the law of large numbers:

$$\frac{1}{N}\sum_{j=1}^{N} (\Delta x_i^j) \approx 0 \; (i = 1, 2, 3),$$

$$\frac{1}{N}\sum_{j=1}^{N} (\Delta x_i^j)^2 \approx \sigma^2 \; (i = 1, 2, 3). \quad (5)$$

The rms difference between the original and the noisy structure is accordingly (for large N)

$$\text{rms} = \sqrt{\frac{1}{N}\sum_{j=1}^{N} [(\Delta x_1^j)^2 + (\Delta x_2^j)^2 + (\Delta x_3^j)^2]} \approx \sqrt{3}\sigma. \quad (6)$$

For $\epsilon \gg \sigma$, the domain orientation will be insensitive to fluctuations in the atom positions. Given the Gaussian distribution of atomic displacements, we can then express the domain size D by the probability to find an atom within a sphere of radius ϵ about its position in the reference structure:

$$D(\epsilon, \sigma) = \frac{1}{(2\pi\sigma^2)^{3/2}} \int_0^\epsilon \exp\left(-\frac{r^2}{2\sigma^2}\right) 4\pi r^2 \, dr. \quad (7)$$

Figure 3 shows that the data agree well with Equation 7 at tolerances above the noise level. For $\epsilon \ll \sigma$, however, we observe deviations from Equation 7 that are more pronounced for larger σ. This behavior is caused by local inhomogeneities in the atom positions. At tolerances below the noise level, the domain will only consist of few atoms and the least-squares fit is no longer insensitive to fluctuations in the positions. Due to the tendency of the domains to

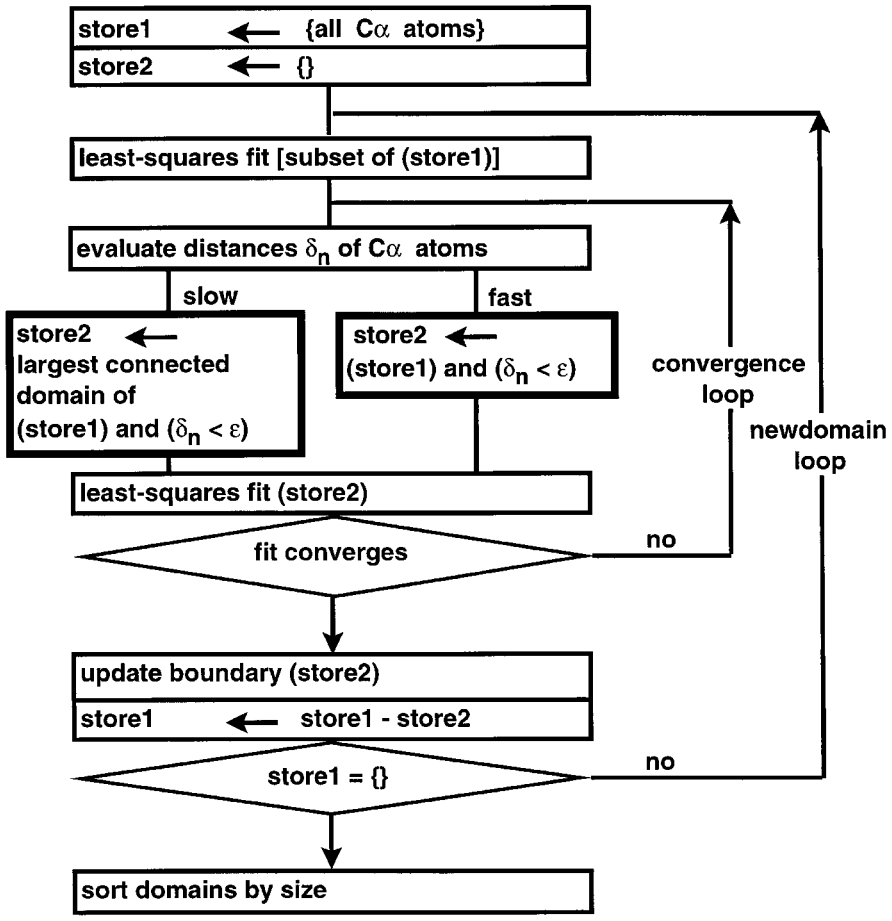


Fig. 1. Flowchart of the adaptive selection procedure. The outer newdomain loop searches for new domains until the protein is fully partitioned. The inner convergence loop determines a new domain by adaptive selection superposition. The set store1 contains all α-carbons in the region complementary to the domains already found and is initially full. The set store2 contains selected α-carbons of the new domain ($s_n = 1$, Equation 4) and is initially empty. A least-squares fit of a small localized seed subset of store1 initializes the convergence loop. In the next step, the distances between corresponding C_α atoms in the two superimposed structures are computed (geometric conformance). In "fast" mode partitioning, all α-carbons that are within ϵ distance will be assigned to store2. In case of "slow" mode partitioning, the spatial connectivity of store2 is also maintained (see Fig. 2). The structures are superimposed by store2. After convergence of the procedure, the boundary of store2 is updated to include "borderline" atoms from earlier found domains which conform better with store2.

converge to better-fitting subsets, the size of subsets (domain size) exhibiting geometric conformance increases above the value given by Equation 7.

We want to formulate an expression that for $\epsilon \ll \sigma$ quantitatively describes how the domain size depends on the tolerance. To find such an empirical formula, we formally replace the upper integration limit "ϵ" in Equation 7 by "$\epsilon + \Delta\epsilon(\epsilon, \sigma)$." $\Delta\epsilon\ (\epsilon, \sigma)$ describes the ϵ-deviation of the data in Fig. 3 from Equation 7. The following simplifying assumptions yield an empirical function for $\Delta\epsilon$:

1. $\Delta\epsilon$ depends solely on ϵ and σ, not on size and geometry of the protein.
2. $\Delta\epsilon$ is proportional to ϵ for a given domain size (linear response of the deviation); because the domain size is determined by the fraction ϵ/σ (Fig. 3, Equation 7), we expect $\Delta\epsilon/\epsilon$ to be constant for constant ϵ/σ.
3. $\Delta\epsilon/\epsilon$ is (on average) a decreasing function of ϵ/σ (Fig. 3, Equation 7), and $\Delta\epsilon/\epsilon$ vanishes (on average) for $\epsilon/\sigma \gg 1$ (deviation vanishes for high tolerance/noise ratio).

In Figure 4, we have plotted the relative deviation $\Delta\epsilon/\epsilon$ against ϵ/σ. We have used the lactoferrin data (691 residues) from Figure 3 and have included similarly obtained data of actin (375 residues). The data can be fitted to an empirical function $\Delta\epsilon/\epsilon = \exp(-3.9\epsilon/\sigma)$. No assertions are made regarding any analytic origin of this function. However, Figure 4 demonstrates that assumptions 1–3 hold very well for the two investigated proteins. We thus arrive at the desired empirical formulation for the domain

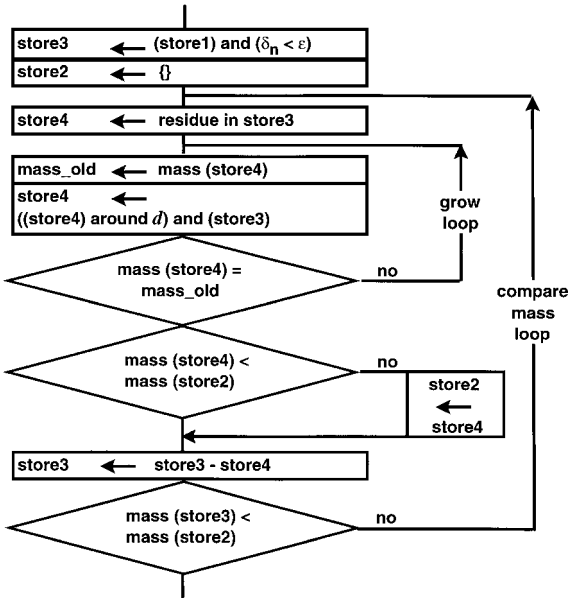


Fig. 2. Flowchart of spatial connectivity maintenance in "slow" mode partitioning. The compare mass loop cycles through the α-carbons that meet the tolerance criterion (store3). The grow loop starts to grow a connected subset starting from a seed atom (store4). The loop exits when no more neighboring residues in store3 within a distance d of store4 are found. Then, store2 is updated to identify the subset with the largest mass after the compare mass loop exits. A maximum connectivity distance $d = 6$ Å is practical for C_α comparisons.

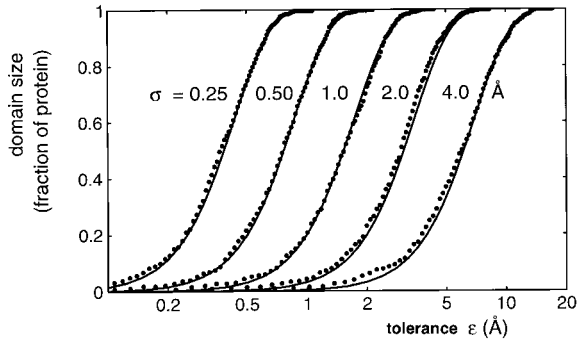


Fig. 3. Effect of tolerance ϵ on the size of the first found domain. The data points have been obtained by means of "fast" mode partitioning of lactoferrin (691 residues) compared with atomic coordinates blurred by Gaussian noise of standard deviation σ. The solid curves correspond to Equation 7.

size $D(\epsilon, \sigma)$ for small tolerances

$$D(\epsilon, \sigma) = \frac{1}{(2\pi\sigma^2)^{3/2}} \int_0^{f(\epsilon,\sigma)} \exp\left(-\frac{r^2}{2\sigma^2}\right) 4\pi r^2\, dr,$$

$$f(\epsilon, \sigma) = \epsilon(1 + \exp(-3.9\epsilon/\sigma)). \quad (8)$$

We will use Equation 8 in the Results section to estimate the noise level in comparisons of structures of lactoferrin, hexokinase, actin, and human tissue factor/human growth hormone receptor.

Locating Effective Rotation Axes

To locate the hinges, we strive to represent the relative movement of rigid domains as effective rotations. We assume that rigid domains have been identified using the adaptive selection procedure above and that the vector sets $\vec{x}_n$ and $\vec{y}_n$ have been superimposed by a chosen reference domain using the Kabsch method. The relative movements of any domain relative to the reference domain can then be characterized as a rigid body transformation with rotation matrix $\mathbf{U}$ and transformation vector $\vec{v}$ (Equation 1). To express this rigid-body transformation as a rotation with matrix $\mathbf{U}'$ about a pivot point $\vec{t}$, we write formally

$$\vec{x}'_n = \mathbf{U}'\vec{x}_n - \mathbf{U}'\vec{t} + \vec{t}. \quad (9)$$

Comparing Equation 1 and 9, we identify $\mathbf{U}' = \mathbf{U}$, which yields a formal solution for the pivot point

$$\vec{t} = (\mathbf{1} - \mathbf{U})^{-1}\vec{v}. \quad (10)$$

Unfortunately, the equation cannot be solved in three dimensions: The orthogonal matrix $\mathbf{U}$ has eigenvalue 1 by virtue of Euler's theorem[28] and $det(\mathbf{1} - \mathbf{U}) = 0$. Thus, in three dimensions, a rotation about a hinge, which would retain all the degrees of freedom of a rigid-body movement, cannot be constructed.

In Figure 5, we introduce an approximation of the effective rotation in three dimensions that works well in most practical cases. From the rotation matrix $\mathbf{U}$ (Equation 1), we can extract by standard linear algebra methods an axis vector $\vec{r}$, which is the eigenvector corresponding to eigenvalue 1,[28] and a rotation angle α about this axis (Fig. 5). The underlying assumption of the approximation is that the relative movement of two rigid domains is constrained by a flexible joint connecting the domains. The effective rotation axis, therefore, should be almost parallel to the plane bisecting the centroid-connecting vector $\vec{v}$ (Fig. 5) (i.e., the helical component of the rotation about $\vec{v}$ is small). We can project in good approximation the rotation onto the bisecting plane that yields a new rotation axis $\vec{r}'$ and a projection angle β between $\vec{r}$ and $\vec{r}'$:

$$\vec{r}' = \vec{r} - \left(\frac{\vec{r}\cdot\vec{v}}{v^2}\right)\vec{v}, \quad (11)$$

$$\beta = \cos^{-1}\left(\frac{\vec{r}\cdot\vec{r}'}{rr'}\right). \quad (12)$$

We decompose the rotation about $\vec{r}$ into two consecutive rotations about $\vec{r}'$ and $\vec{v}$. The new rotation angle α′ about $\vec{r}'$ is[29,30]

$$\alpha' = 2\tan^{-1}\left(\cos(\beta)\cdot\tan\left(\frac{\alpha}{2}\right)\right). \quad (13)$$

Using this component of the rotation, one can now construct the effective rotation axis shifted by a

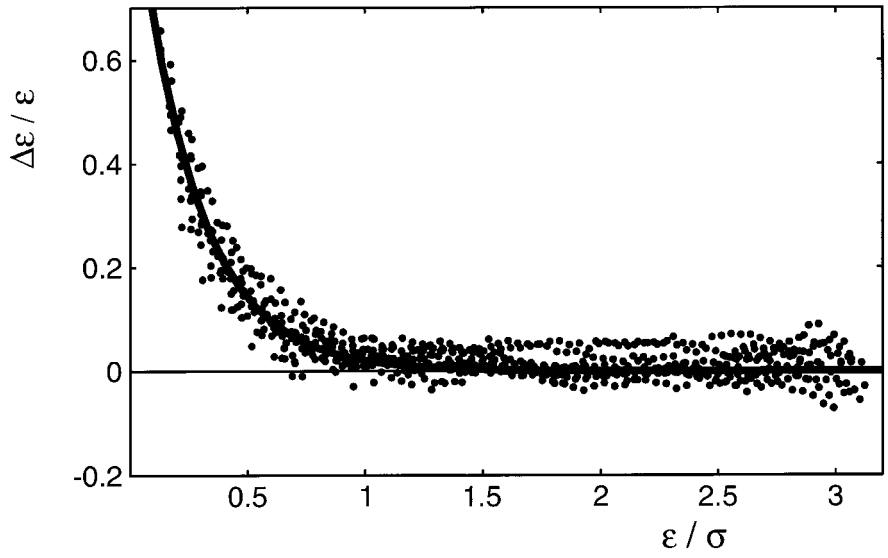


Fig. 4. Relative deviation of actin and lactoferrin domain size data from the large tolerance limit of D (ϵ, σ) (Equation 7). The relative deviation in ϵ-direction of Fig. 3 has been plotted against the relative tolerance ϵ/σ. The solid curve has been fitted to the data (Equation 8).

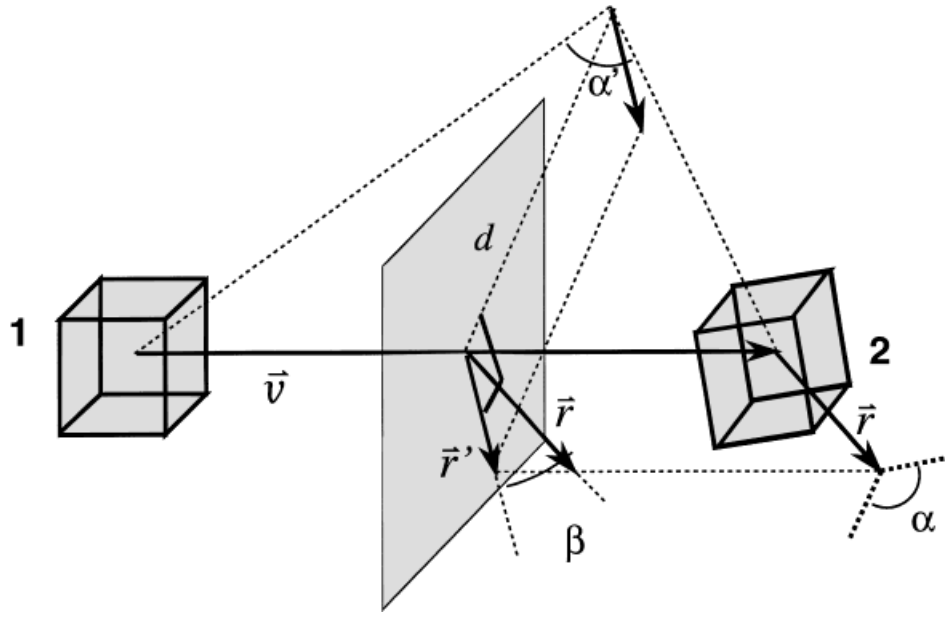


Fig. 5. Construction of the effective rotation axis. A cube schematically represents the moving protein domain in positions 1 and 2. The vector $\vec{v}$ connects the center of mass of the domain in the two positions. The rigid body rotation axis $\vec{r}$ is projected onto the bisecting plane of $\vec{v}$. The projected axis $\vec{r}'$ is then shifted a distance d on the bisecting plane. A rotation about the constructed effective rotation axis superimposes the center of mass of the domain with the center of mass in position 2.

distance d on the bisecting plane (Fig. 5), in which

$$d = \frac{v/2}{\tan(\alpha'/2)}. \tag{14}$$

A rotation with angle α' about the axis then transforms the center of mass (COM) of the reference domain in position 1 onto the COM of the domain in position 2. The projection, thus, maintains, relative to the least-squares fit, the removal of the COM difference between the sets, but approximates the rotation.

The quality of the approximation can be assessed by the magnitude of the projection angle β. The angle

Table I. Uniqueness of the Largest Found Domain in a Comparison of Two Crystal Structures of the Protein Lactoferrin[33,34] (691 Residues) at 1.2 Å Tolerance, Using Fast Mode and Slow Mode Partitioning at Various Seed Set Radii ρ (Equation 3)

Seed set radius ρ	Mode	Test cases	Domain size		Membership of residues		
			Mean	SD	100%	50–99%	1–49%
10Å	Fast	688	325	3	318	9	4
10Å	Slow	687	325	5	298	29	4
15Å	Fast	691	325	3	318	9	4
15Å	Slow	691	325	5	315	13	3
20Å	Fast	691	323	4	311	14	6
20Å	Slow	691	325	4	315	13	3
25Å	Fast	691	328	4	310	21	2
25Å	Slow	691	325	4	315	13	3

*The domain size (number of residues) is averaged over the test cases of different initial seed indices λ (Equation 3). The mean and the standard deviation (SD) of the domain size are given. Only test cases in which the domain converged to a subset of >200 atoms have been considered, resulting in <691 cases for ρ = 10 Å. The uniqueness of the found domain can be assessed by the membership distribution of all residues.

β is also the angle between our effective rotation axis and the classic screw axis of rigid-body movement.[28,31,32] The screw axis is a shifted rotation axis that does not eliminate the COM translation completely to retain all the degrees of freedom of a rigid-body movement. A theorem attributed to Chasles[31] (which directly follows from Euler's theorem[28]) states that a unique position for the rotation axis can be found for which the residual translation vector is parallel to the rotation axis. Any rigid-body movement can be so described as a helical twist about the axis, accompanied by a helical rise along the axis.[32] Ideally, β vanishes for true rotations about a hinge, and our effective rotation axis becomes identical to the screw axis.

Another useful quantity to assess the approximation of the movement by the effective rotation axis is the error Δ in terms of the rms differences of the domain in the least-squares fit, rms_{ls}, and in the effective rotation, rms_{al}, relative to the COM displacement v

$$\Delta = \frac{\mathrm{rms}_{al} - \mathrm{rms}_{ls}}{v}. \qquad (15)$$

PERFORMANCE

In this section, we demonstrate the general performance properties of *Hingefind*. First, we discuss the uniqueness of the subsets found by adaptive selection, using crystal structures of the protein lactoferrin (691 residues). Subsequently, we discuss the accuracy of found effective rotations for domain movements of the protein actin (375 residues).

Convergence of Found Domains

To investigate how the found rigid domains depend on the seed set radius ρ and the seed atom index λ (Equation 3), we have compared two crystal structures of the protein lactoferrin[33,34] at 1.2 Å tolerance (cf. Results section), using fast and slow (spatial connectivity) mode partitioning. On a Hewlett-Packard 735/125 workstation, a single comparison of the two lactoferrin structures required about 20 seconds in "fast" mode and about 60 seconds in "slow" mode.

We have determined the largest rigid domain for seed subsets of radii ρ = 10, 15, 20, and 25 Å, each starting from one of the 691 C_α atoms of the protein (λ = 1, . . . , 691). Table I shows the mean and standard deviation of the size of the found domain averaged over λ. Seed index–dependent fluctuations are relatively small (three to five residues) compared with the total domain size (about 325 residues). The table also shows the number of residues that we found in 100% (common residues), 50–100% (likely members), and 1–49% of the test cases with different λ. Of special interest is the number of residues common to all λ-dependent instances of the domain (100% column). Table I demonstrates that this number is generally very close to the domain size, with the exception of the small seed set radius (ρ = 10 Å), when there is an uncertainty of membership of about 10% of the residues in slow mode partitioning. Also, for this small seed set radius the domain did not converge for every λ to a well-defined subset of >200 atoms.

The uniqueness of the found domain is most pronounced for ρ = 15 Å: For this radius, 98% and 97% of the domain residues have been found to be independent of λ, for fast and slow mode partitioning, respectively. In practical applications of the method, the index λ can, therefore, be randomly chosen.

Accuracy of the Effective Rotation

To investigate the accuracy of the found effective rotation, we have compared structures of actin from

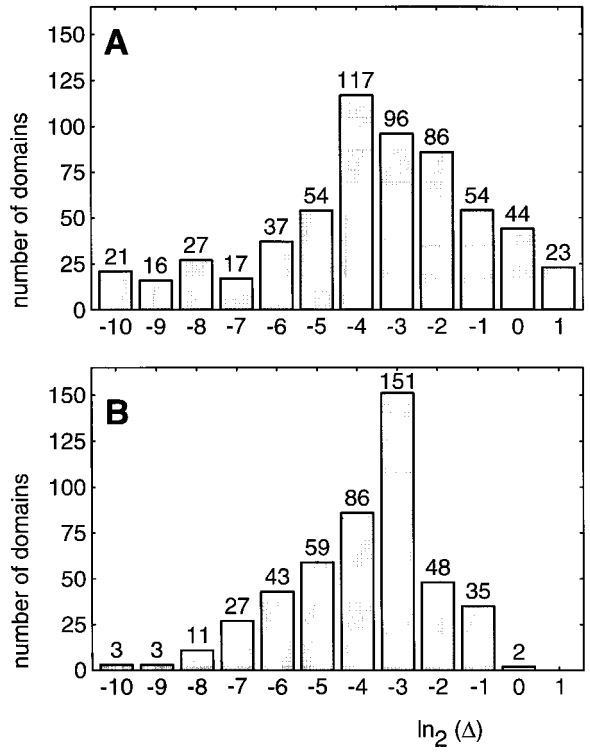


Fig. 6. Logarithmic distribution of relative errors Δ (Equation 15). The bar corresponding to $ln_2(\Delta) = -n$ represents the number of cases in the interval $[2^{-n-1}, 2^{-n}]$. **A:** "Fast" mode partitioning (605 test cases). **B:** "Slow" mode partitioning (spatial connectivity maintained, 476 test cases).

MD simulation[35] and from the PDB entry 1ATN[36] and incremented the tolerance ϵ from 20% to 150% of the initial rms difference (2.8 Å). Only domains comprising >15 residues have been considered for an evaluation of the effective rotation axis. The relative rigid-body movements have then been determined by superimposing the largest domain of the protein structures, yielding 605 relative movements sampled in "fast" partitioning and 476 movements sampled in "slow" partitioning (Fig. 6A,B): Spatial connectivity reduces the number of found domains. About 80% of the "slow" mode cases and 65% of the "fast" mode cases exhibit relative errors <12.5%, indicating a good representation of the corresponding movements as an effective rotation. Most movements with errors >50% have been eliminated when we used the "slow" mode algorithm (Fig. 6A,B).

Of interest is the dependence of the relative error (for the "slow" mode cases) on the tolerance ϵ. Figure 7 shows a window of low error Δ at tolerances between 60% and 80% of the initial rms difference of the two structures. It is obvious that the tolerance should be smaller than the initial rms difference if locally preserved domains are to be detected. If the tolerance chosen is too high, the largest found domain will make up most of the protein, leaving only small structural fragments for the other domains that are possibly disoriented. However, if the tolerance is below the local noise level in the structure comparison, any found domain will be small (Fig. 3) and disoriented due to local structural inhomogeneities. In both cases, the movements are then more likely to be poorly described as effective rotations. Thus, outside the window (Fig. 7), errors can become large and the algorithm yields poor approximations of the Kabsch least-squares fit.

Figure 8 demonstrates the correlation of the error Δ with the projection angle β. More than half of the test cases in the survey show a projection angle <20°. For these cases of good hinge approximation, the relative error remains confined to values <17%.

RESULTS

In this section, we visualize domain movements for four sample proteins that exhibit significant conformational differences: The proteins lactoferrin and hexokinase allow us to demonstrate the performance of *Hingefind* in comparisons of classic structures from X-ray crystallography. The versatility of the program is further exemplified by an investigation of domain movements arising in MD simulations of the protein actin and by a comparison of two homologous cytokine receptors, the extracellular domains of human tissue factor and of the receptor of human growth factor.

We apply the "fast" mode partitioning method and Equation 8 to estimate the noise level in the comparisons of structures. The tolerance of the partitioning is optimally chosen to filter significant conformational differences (the domain movements) from the noise. The structures are then compared by "slow" mode partitioning at optimum tolerance. The results of the comparisons of protein structures with *Hingefind* are visualized by the molecular graphics program *VMD*.[47]

Lactoferrin

Crystal structures have been determined for an iron-bound and an iron-free form of the iron-transporting protein (691 residues). Each N- and C-terminal lobe contains two domains, denoted N1, N2 and C1, C2. In the iron-bound form[33] (PDB entry 1LFG), N1 and N2 are moved together with an iron bound between them. The same is true for C1 and C2. In the iron-free form[34] (PDB entry 1LFH), N1 and N2 are separated, exposing a deep cleft between them. The similarity of the domain structure in the two conformations renders lactoferrin an ideal candidate for testing *Hingefind*.

Hexokinase

The monomer of yeast hexokinase comprises 458 residues. The structure contains two lobes: the large

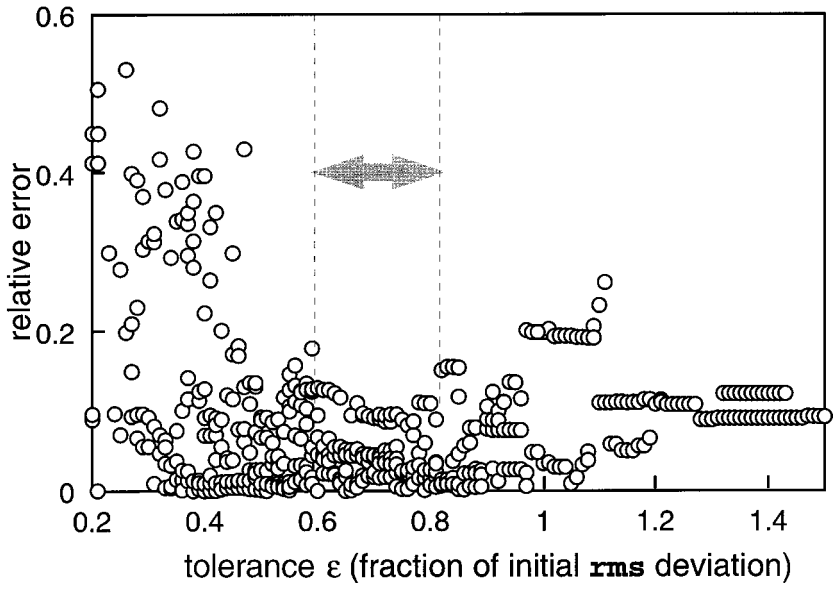


Fig. 7. Dependence of the relative error Δ on the tolerance ε (476 actin "slow"-mode test cases). The window of low Δ values (gray double arrow) corresponds to the optimum range of tolerance for the investigated system.

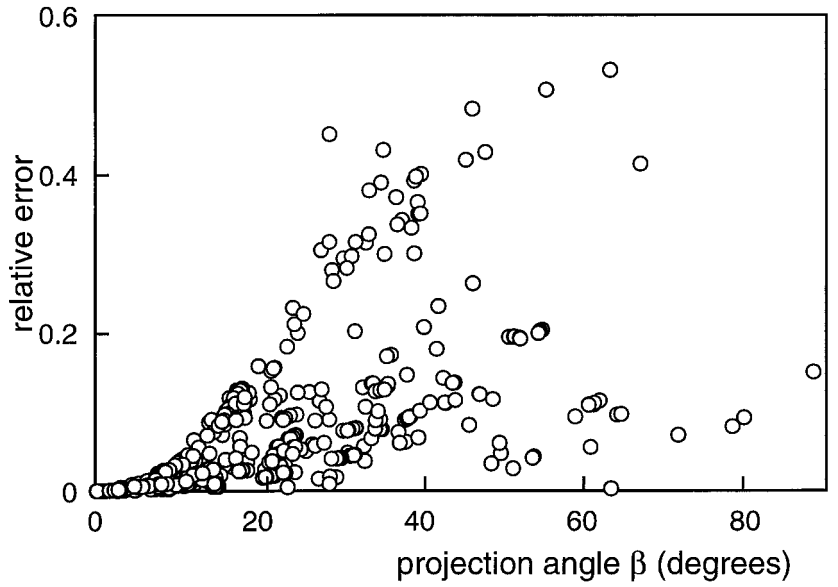


Fig. 8. Dependence of the relative error Δ on the projection angle β (476 actin "slow"-mode test cases); in more than half of the test cases, the projection angle is <20°.

lobe (residues 1–58, 187–458) and the small lobe (residues 59–186). The active site is in a cleft between them. The liganded[37,38] (PDB entry 1HKG) and unliganded[39] (PDB entry 2YHX) forms of the enzyme differ in the relative spatial disposition of the domains. On binding glucose, they move toward each other, burying the glucose almost completely. We will show that the cleft closure in hexokinase is accompanied by tertiary structure rearrangements similar to the movements in computer simulations of actin.

Actin

The G-actin monomer consists of 375 amino acid residues, forming four subdomains that are oriented around the active site and give the protein a cloverleaf appearance. An ATP or ADP nucleotide is bound in the center of the protein, and subdomains 2 and 4

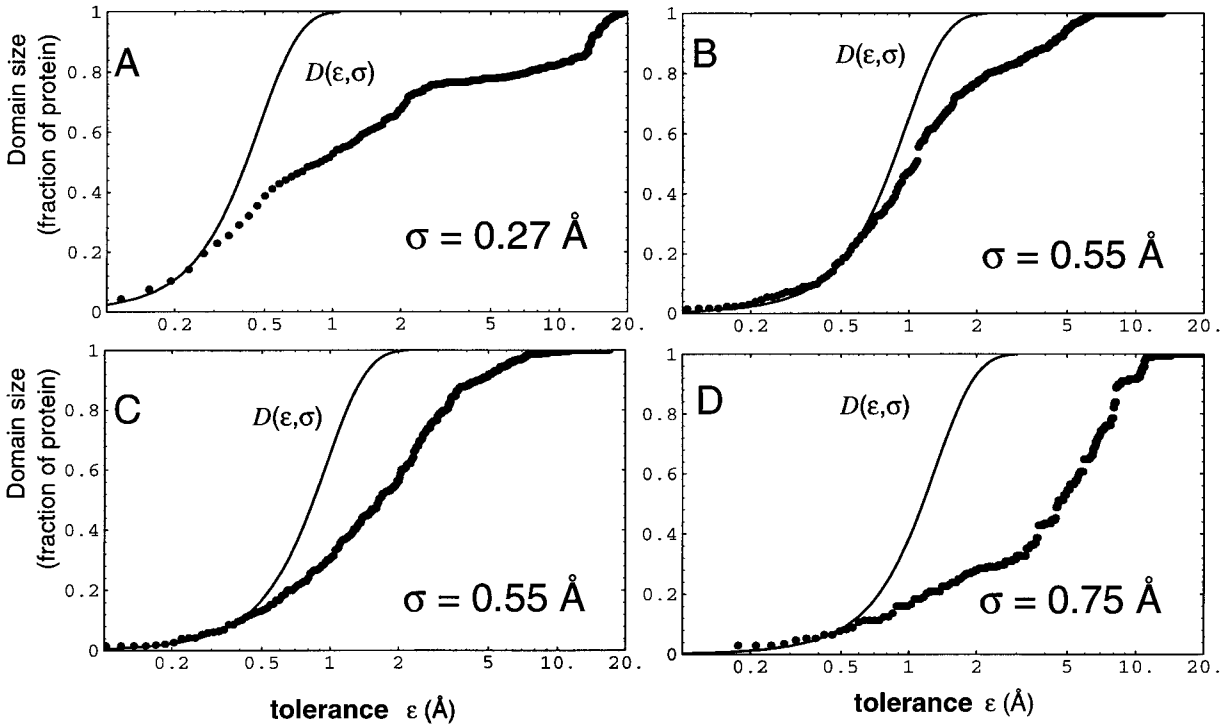


Fig. 9. Effect of changing tolerance ϵ on the domain size. **A:** Lactoferrin. **B:** Hexokinase. **C:** Actin. **D:** Tissue factor/growth hormone receptor. The function $D(\epsilon, \sigma)$ given by Equation 8 has been fitted to the data at small tolerance. The resulting values of σ are given.

are separated by the nucleotide binding cleft. Crystal structures of actin[36,40,41] are very similar,† which motivated us to simulate actin.[35] We compare the structure from MD simulation to the crystal structure of the complex with DNase 1[36] (PDB entry 1ATN).

Human Tissue Factor/Human Growth Hormone Receptor

The extracellular domains of human tissue factor[43,44] (PDB entry 1HFT) and of the receptor of human growth hormone[45] (PDB entry 3HHR) belong to a superfamily of cytokine receptors.[46] The structural geometry of the biopolymers consists of two immunoglobulin-like domains connected by a hinge region. Each domain is formed by two antiparallel β-sheets. For the structural comparison, we have identified 168 pairs of corresponding residues (out of approximately 230 residues each) based on sequence alignment.[46] This test case provides an example for the importance of sequence alignment for visualization of structural differences in a comparison of homologous proteins with *Hingefind*.

Domain Movements vs. Intrinsic Noise

Equation 8 provides an expression for the tolerance dependence of the domain size $D(\epsilon, \sigma)$ for atomic coordinates perturbed by Gaussian noise. In a comparison of two authentic protein structures, the asymptotic scaling of the domain size at small ϵ can be used to fit $D(\epsilon, \sigma)$ by adjusting the parameter σ (Fig. 9).

From Figure 9A,B, we have determined the rms value of local noise, $\sqrt{3}\sigma$ (Equation 6), in the two cases in which isoforms of crystal structures have been compared. The rms level of local noise is 0.47 Å and 0.95 Å for lactoferrin and hexokinase, respectively. Both values lie within 10% of values reported elsewhere[30,34,38] for rms differences of the well-fitting domains. For the comparison of the two actin structures, we obtain a noise value of 0.95 Å (Fig. 9C). The noise may reflect contributions from thermal disorder in the simulated structure[35] and from static disorder of the actin crystal structure at 2.5 Å resolution.[36] For the two cytokine receptors, we obtain a rms noise value of 1.3 Å (Fig. 9D). This relatively high noise level reflects uncertainties of

†Note added in proof. Shortly after submission of this manuscript a new crystal structure of actin in an "open" form was reported.[42] The movements exhibited by this structure relative to the earlier structures.[36,40,41] involve rotations of subdomain 4 away from the cleft, opposite to the movements shown in this paper and in ref. 35. An only slight difference in energy between the open and closed states reveals an extreme sensitivity of actin's conformation to changes in the environment.[42] Our structure from MD simulation can therefore be interpreted as one instance in an ensemble of thermally accessible conformations.

sequence alignment[46] and structural differences between tissue factor and growth hormone receptor.

Domain Movements in Trial Proteins

The domain sizes of the compared proteins deviate from $D(\epsilon, \sigma)$ for tolerances above the local noise level due to large-scale movements in the protein structures (Fig. 9). For the visualization of the domain movements, we have chosen a tolerance at which the difference between domain size and $D(\epsilon, \sigma)$ in Figure 9 is largest: These tolerances are 1.2 Å, 1.8 Å, 1.8 Å, and 3.5 Å for lactoferrin, hexokinase, actin, and the cytokine receptors, respectively.

The results of *Hingefind* structural comparisons are depicted for the four cases in Figures 10 to 13. Effective rotation axes and perpendicular COM-connecting lines are shown as tubes in the color of the corresponding domain. The arrow indicates a left-handed rotation that shifts the COM of the domain depicted in white onto the COM of the domain in a colored tube representation. In all cases, we chose the largest found rigid domain as reference domain by which the structures were superimposed.

Lactoferrin

The domains in lactoferrin move essentially as rigid bodies. The three rigid domains found show little disorder and are practically identical to the structural domains of lactoferrin (Fig. 10). The lobes exhibit rms differences between 0.42 and 0.65 Å when superimposed.[34] These values are in agreement with the noise level of 0.47 Å. The axis of rotation of lobe N1 passes through the two β-strands linking lobes N1 and N2. The C-lobe is structurally intact even in the iron-free form. The differences in domain movements in the C- and N-lobe have been interpreted as effects of crystal packing forces.[34] The reported value of the rotation angle (53° for the movement of N1 relative to N2[34]) is consistent with our data (8° and 54° of N1 and N2, respectively, relative to the C-lobe).

Hexokinase

The large lobe of hexokinase appears to be structurally well preserved (Fig. 11). A second rigid region comprises part of the small lobe. The movement exhibited by this domain appears to be rather unusual: The second domain is irregularly shaped and forms no clearly defined interface to the large lobe. The underlying conformational change induced by glucose binding has been the subject of several studies,[37,38,48–50] which suggests that the domain closure in hexokinase is a cumulative result of many small relative movements of secondary structure elements, rather than of a single rigid-body motion of the small lobe. The effective rotation axis coarsely represents the complex conformational change in the small lobe. The angle of the effective rotation (13°, Fig. 11) agrees well with a rotation angle of 12° reported elsewhere.[38,49]

Actin

The largest rigid domain comprises most residues of actin's subdomains 1 and 3 (Fig. 12). Subdomain 2 relaxes in the simulated structure by rotating about a hinge at the interface to subdomain 1. The relaxation is due to the missing DNase1 of the complex in the crystal structure.[35] Several rather small domains and the relatively large number of disordered residues on the surface of the protein (120 of 375 residues) indicate that the simulated structure exhibits tertiary structure rearrangements that are the result of thermal fluctuations.[35] The most prominent conformational change is a movement of rigid regions in actin's structural subdomain 4, which closes the nucleotide binding cleft of the protein (Fig. 12) as discussed in ref. 35.

Human Tissue Factor/Human Growth Hormone Receptor

The case of the two different cytokine receptors demonstrates the challenges of comparing two homologous proteins of deviating sequence. Even at a low partitioning tolerance of 3.5 Å, about one-third of the residues appear to be disordered (Fig. 13). This result may be due to a mismatch in residue pairing from sequence alignment,[46] as well as due to geometric differences in the tertiary structure of the domains.[43–45]

The different orientation of the domains of the two receptors has been attributed to the structure of the interface between the domains: The hydrophobic core in tissue factor is continuous throughout the interface, in contrast to the growth hormone receptor. As a consequence, the ability of certain residues near the interface to contribute to ligand binding differ significantly between the two receptors.[43] The hinge-bending angle of 46° agrees well with the reported value of 45°.[43]

CONCLUSIONS

The algorithm *Hingefind* can be used to partition a protein into domains of preserved geometry and to determine effective rotation axes of the movement. *Hingefind* allows one to investigate domain movements in proteins as demonstrated for the structures of lactoferrin, hexokinase, actin, and two cytokine receptors (human tissue factor and human growth hormone receptor). The algorithm should become a useful tool for the characterization of domain movements in other crystallographic structures and for the analysis of structures from molecular dynamics simulations.

Methods that extract rigid domains in a comparison of two protein structures have been developed before. The simple sieving routine by Lesk[1] is limited to situations in which only a single well-preserved domain exists. Nichols et al.[22] have introduced a distance matrix-based algorithm that would also

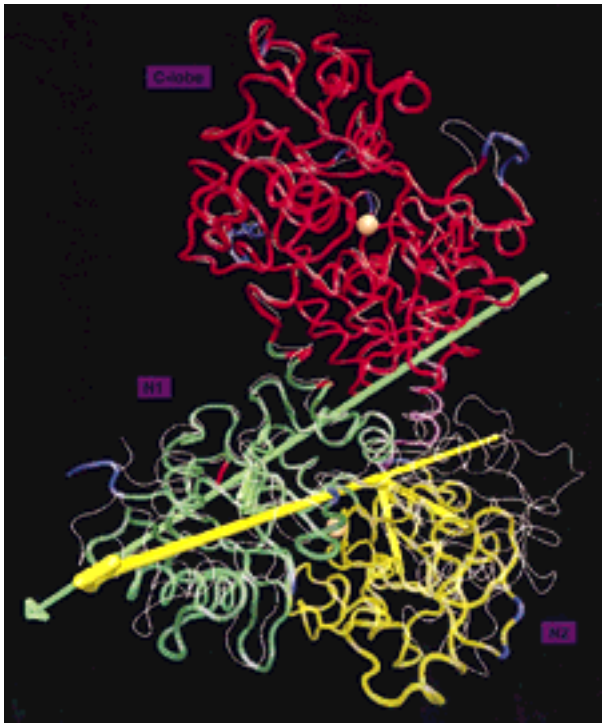


Fig. 10.

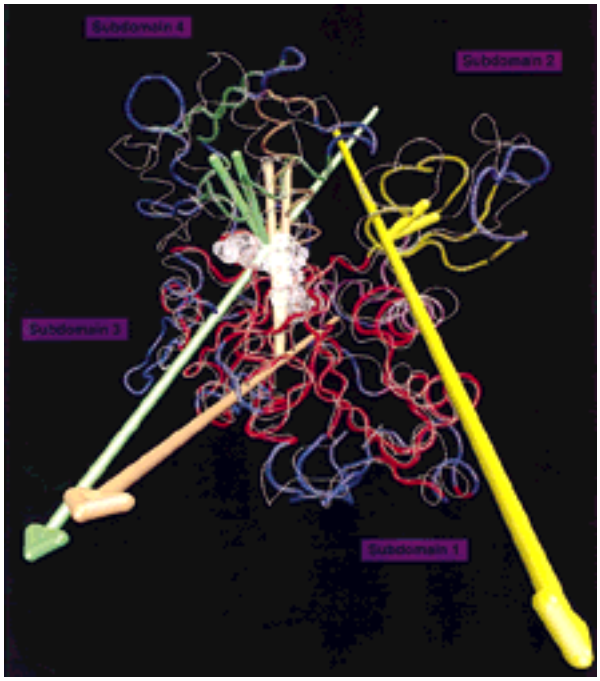


Fig. 12

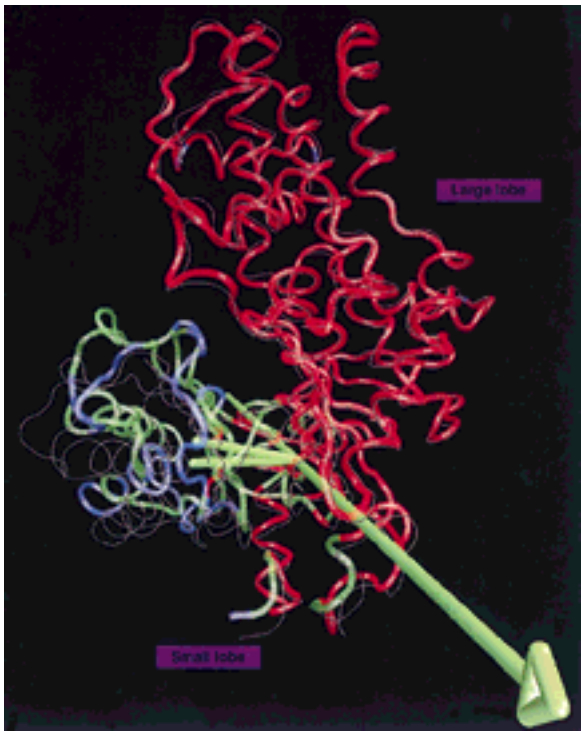
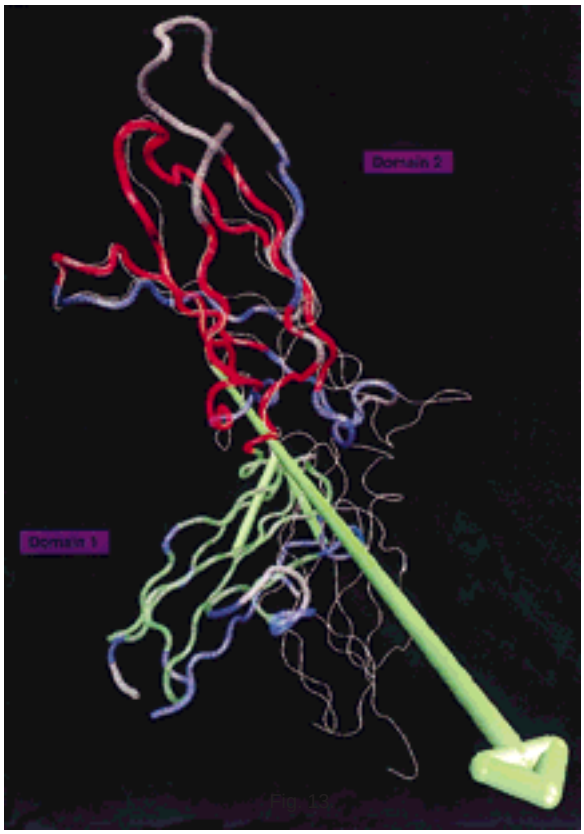


Fig. 11.



work in situations with several differently oriented rigid domains. Distance matrix methods, however, suffer from high storage and processing demands. The Nichols routine hinges on the rather unpermissive requirement that all interatomic distances in a comparison of two domain conformations must satisfy an ϵ-criterion. An exhaustive search to satisfy this requirement becomes unfeasible for structures comprising >50 α-carbons.[22] Nichols et al. perform, therefore, a nonexhaustive search, which does not result in unique domains but in families of intersecting domains.[22]

The adaptive selection method presented in this work similarly exhibits, for very few residues, an uncertainty of their membership to a rigid domain, but we have demonstrated that a found domain can be considered unique in practical applications. The sole requirement of the algorithm is that the distances between corresponding atoms in the superimposed domains must satisfy the ϵ-criterion. The method therefore has low storage and processing demands. There is no limit for the system size other than given by the configuration of the program used in the analysis. The algorithm converges rapidly within five to ten cycles of the convergence loop (Fig. 1). Despite the differences in the method, the observed domain size scaling of the trial structures with changing tolerance (Fig. 9) agrees qualitatively with the domain sizes obtained for hemoglobin by Nichols et al. (Fig. 6 in ref. 22).

We have been able to reproduce and visualize aspects of domain movements exhibited by the trial proteins as discussed in the literature.[33,34,38,43,45,48] Not every conceivable conformational change in a protein constitutes an effective rotation of rigid domains. As seen in the case of hexokinase and actin, tertiary structure rearrangements may accompany the movement of domains. Gerstein et al. have observed that in the case of shear motions, as exhibited by hexokinase, main chain atoms are often constrained by close packing, and small torsion angle changes are spread over many angles.[2,30] Even in this case, the effective rotation axis determined by *Hingefind* permits a coarse representation of the conformational change (Fig. 11). The comparison of human tissue factor and human growth hormone receptor has demonstrated that homologous proteins can be compared in regard to domain movements if their sequences are reasonably well aligned.

Methods for automated biomolecular docking of proteins in rational drug design and discovery require a priori knowledge of suitable hinges for ligand binding and protein assembly.[3,4] One possible application of *Hingefind* is the prediction of feasible binding modes for induced fit-docking of proteins. At present, the method requires two sets of atomic coordinates to find the hinges. As the case of actin shows, structures from MD simulations can sometimes substitute to identify rigid domains and their flexible joints if only a single structure is known. However, MD simulations rarely involve sufficient sampling to reproduce the types of rigid-body motions found in proteins.[51] It therefore remains a challenge for computational biologists to predict rigid-body movements based on a single structure.

ACKNOWLEDGMENTS

We thank M. Balsera, Ch. Martin, A. Dalke, and M. Seto for discussions and suggestions. This work was supported by NIH grant PHS 5 P41 RR05969-04, NSF grant BIR-9318159, Roy J. Carver Charitable Trust, and MCA93S028P computertime grant at Pittsburgh Supercomputing Center.

Fig. 10. Domain movements of lactoferrin. Shown are the backbone traces of iron-bound[33] (color) and iron-free lactoferrin[34] (white). Iron ions are shown in orange. Effective rotation axes and perpendicular centroid-connecting lines are rendered as tubes in the color of the corresponding domain. The arrows indicate a left-hand rotation, which shifts the center of mass of the domain in the iron-free structure onto the center of mass in the iron-bound structure. Three domains >15 residues have been found at 1.2 Å tolerance: domain 1 (red, 325 residues) is the reference domain that has been superimposed with the iron-free structure; domain 2 (green, 171 residues) rotates by 8° (relative error $\Delta = 13\%$); domain 3 (yellow, 155 residues) rotates by 54° (relative error $\Delta = 4\%$). Disordered regions (26 residues) are shown in blue.

Fig. 11. Domain movements of hexokinase. Rendered are the backbone traces of liganded[37,38] (color) and unliganded hexokinase[39] (white). Glucose is not shown. Effective rotation axes and perpendicular centroid-connecting lines are depicted as tubes in the color of the corresponding domain. The arrow indicates a left-hand rotation, which shifts the center of mass of the domain in the unliganded structure onto the center of mass in the liganded structure. Two domains >15 residues have been found at 1.8 Å tolerance: domain 1 (red, 317 residues) is the reference domain that has been superimposed with the unliganded structure; domain 2 (green, 98 residues) rotates by 13° (relative error $\Delta = 3\%$). Disordered regions (42 residues) are shown in blue.

Fig. 12. Domain movements of actin. The backbone traces of the simulated structure[35] (color) and crystal structure[36] (white) are shown. ADP nucleotide atoms are rendered as transparent spheres. Effective rotation axes and perpendicular centroid-connecting lines are shown as tubes in the color of the corresponding domain. The arrows indicate a left-hand rotation, which shifts the center of mass of the domain in the crystal structure onto the center of mass in the simulated structure. Five domains >15 residues have been found at 1.8 Å tolerance: domain 1 (red, 166 residues) is the reference domain that has been superimposed with the crystal structure; domain 2 (green, 26 residues) rotates by 14° (relative error $\Delta = 5\%$); domain 3 (yellow, 24 residues) rotates by 21° (relative error $\Delta = 13\%$); domain 4 (purple, 22 residues, hinge axis not shown); domain 5 (orange, 17 residues) rotates by 5° (relative error $\Delta = 6\%$). Disordered regions are shown in blue.

Fig. 13. Comparison of domain orientations of human growth hormone receptor (HGHR) and human tissue factor (HTF). The backbone traces of HGHR[45] (color) and HTF[43,44] (white) are shown. The arrow indicates a left-hand rotation, which shifts the center of mass of the domain in HTF onto the center of mass in the HGHR structure. Two domains >15 residues have been found at 3.5 Å tolerance: domain 1 (red, 54 residues) is the reference domain that has been superimposed with HTF; domain 2 (green, 53 residues) rotates by 46° (relative error $\Delta = 4\%$). Disordered regions (61 residues) are shown in blue. Regions where sequence alignment could not identify pairs of corresponding residues are rendered gray.

REFERENCES

1. Lesk, A.M. "Protein Architecture." New York: Oxford University Press, 1991.
2. Gerstein, M., Lesk, A.M., Chothia, C. Structural mechanisms for domain movements in proteins. Biochemistry 33:6739–6749, 1994.
3. Sandak, B., Nussinov, R., Wolfson, H.J. An automated computer vision and robotics-based technique for 3-D flexible biomolecular docking and matching. Comp. Appl. Biosci. 11:87–99, 1995.
4. Sandak, B., Nussinov, R., Wolfson, H.J. Docking of conformationally flexible proteins. In: "Proceedings of the 1996 Symposium on Combinatorial Pattern Matching." Dan Hirschberg, Gene Myers, eds. Lecture Notes in Computer Science. Vol. 1075. Springer Verlag, 1996.
5. Brünger, A.T. "X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR." The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 1992.
6. Rossman, M.G., Argos, P. Exploring structural homology of proteins. J. Mol. Biol. 105:75–95, 1976.
7. Rao, S.T., Rossman, M.G. Comparison of super-secondary structures in proteins. J. Mol. Biol. 76:241–256, 1973.
8. Remington, S.J., Mathews, B.W. A general method to assess similarity of protein structures, with applications to T4 bacteriophage lysozyme. Proc. Natl. Acad. Sci. U.S.A. 75:2180–2184, 1978.
9. Levine, M., Stuart, D., Williams, J. A method for the systematic comparison of three-dimensional structures of proteins and some results. Acta Cryst. A 40:600–610, 1984.
10. Barton, G.J., Sternberg, M.J.E. LOPAL and SCAMP: Techniques for the comparison and display of protein structures. J. Mol. Graphics 6:190–196, 1988.
11. Zuker, M., Somorjai, R.L. The alignment of protein structures in three dimensions. Bull. Math. Biol. 51:55–78, 1989.
12. Taylor, W.R., Orango, C.A. Protein structure alignment. J. Mol. Biol. 208:1–22, 1989.
13. Šali, A., Blundell, T.L. Definition of general topological equivalence in protein structures. J. Mol. Biol. 212:403–428, 1990.
14. Brint, A.T., Davies, H.M., Mitchell, E.M., Willett, P. Rapid geometric searching in protein structures. J. Mol. Graphics 7:48–53, 1989.
15. Holm, L., Sander, C. Structure comparison by alignment of distance matrices. J. Mol. Biol. 233:123–138, 1993.
16. Abagyan, R.A., Maiorov, V.N. A simple qualitative representation of polypeptide chain folds: Comparison of protein tertiary structures. J. Biomed. Struct. Dyn. 5:1267–1279, 1988.
17. Abagyan, R.A., Maiorov, V.N. An automatic search for similar spatial arrangements of α-helices and β-strands in globular proteins. J. Biomed. Struct. Dyn. 6:1045–1060, 1989.
18. Lesk, A.M. Detection of 3D patterns of atoms in chemical structures. Communications A.C.M. 22:219–224, 1979.
19. Vriend, G., Sander, C. Detection of common three-dimensional substructures in proteins. Proteins 11:52–58, 1991.
20. Freer, S.T., Kraut, J., Robertus, J.D., Wright, H.T., Xuong, N.H. Chymotrypsinogen: 2.5-Å crystal structure, comparison with α-chymotrypsin, and implications for zymogen activation. Biochemistry 9:1997–2009, 1970.
21. Chothia, C., Lesk, A.M. The relation between the divergence of sequence and structure in proteins. EMBO J 5:823–826, 1986.
22. Nichols, W.L., Rose, G., Eyck, L.F.T., Zimm, B.H. Rigid domains in proteins: An algorithmic approach to their identification. Proteins 23:38–48, 1995.
23. Kabsch, W. A solution for the best rotation to relate two sets of vectors. Acta Cryst. A 32:922–923, 1976.
24. Kabsch, W. A discussion of the solution for the best rotation to relate two sets of vectors. Acta Cryst. A 34:827–828, 1978.
25. Browner, M.F., Fauman, E.B., Fletterick, R.J. Tracking conformational states in allosteric transitions of phosphorylase. Biochemistry 31:11297–11304, 1992.
26. Gerstein, M., Altman, R.B. Average core structures and variability measures for protein families: Application to the immunoglobulins. J. Mol. Biol. 251:161–175, 1995.
27. Lesk, A.M., Fordham, W.D. Conservation and variability in the structures of serine proteinases of the chymotrypsin family. J. Mol. Biol. 258:501–537, 1996.
28. Goldstein, H. "Classical Mechanics." Reading, MA: Addison-Wesley Pub. Co., 1980.
29. Diamond, R. On the factorization of rotations with examples in diffractometry. Proc. R. Soc. Lond. A (Math. Phys. Sci.) 428:451–472, 1990.
30. 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.
31. Chasles, M. Note sur les propriétés générales du système de deux corps semblables entre'eus et placés d'une manière quelconque dans l'espace; et sur la deplacement fini ou infinement petit d'un corps solide libre. Férusac, Bull. Sci. Math. 14:321–326, 1830.
32. Babcock, M.S., Pednault, E.P.D., Olson, W.K. Nucleic acid structure analysis. J. Mol. Biol. 237:125–156, 1994.
33. Anderson, B.F., Baker, H.M., Norris, G.E., Rice, D.W., Baker, E.N. Structure of human lactoferrin. J. Mol. Biol. 209:711–734, 1989.
34. Anderson, B.F., Baker, H.M., Norris, G.E., Rumball, S.V., Baker, E.N. Apolactoferrin structure demonstrates ligand-induced conformational change in transferrins. Nature 344:784–787, 1990.
35. Wriggers, W., Schulten, K. Stability and dynamics of G-actin: Back door water diffusion and behavior of a subdomain 3/4 loop. Biophys. J., in press.
36. Kabsch, W., Mannherz, H., Suck, D., Pai, E., Holmes, K. Atomic structure of the actin: DNase I complex. Nature 347:37–44, 1990.
37. Bennett, W.S., Steitz, T.A. Structure of a complex between yeast hexokinase A and glucose II. J. Mol. Biol. 140:211–230, 1980.
38. Bennett, W.S., Steitz, T.A. Structure of a complex between yeast hexokinase A and glucose I. J. Mol. Biol. 140:183–209, 1980.
39. Anderson, C.M., Stenkamp, R.E., Steitz, T.A. Sequencing a protein by X-ray crystallography. J. Mol. Biol. 123:1–13, 1978.
40. Schutt, C.E., Myslik, J.C., Rozycki, M.D., Goonesekere, N.C.W., Lindberg, U. The structure of crystalline profilin-β-actin. Nature 365:810–816, 1993.
41. McLaughlin, P.J., Gooch, J.T., Mannherz, H.-G., Weeds, A.G. Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. Nature 364:685–692, 1993.
42. Chik, J., Lindberg, U., Schutt, C. The structure of an open state of beta actin at 2.65 Å resolution. J. Mol. Biol. 263:607–623, 1996.
43. Muller, Y.A., Ultsch, M.H., Kelley, R.F., de Vos, A.M. Structure of the extracellular domain of human tissue factor: Location of the factor VIIa binding site. Biochemistry 33:10864–10870, 1994.
44. Harlos, K., Martin, D.M.A., O'Brien, D.P., Jones, E.Y., Stuart, D.I., Polikarpov, I., Miller, A., Tuddenham, E.G.D., Boys, C.W.G. Crystal structure of the extracellular region of human tissue factor. Nature 370:662–666, 1994.
45. de Vos, A.M., Ultsch, M., Kossiakoff, A.A. Human growth hormone and extra-cellular domain of its receptor: Crystal structure of the complex. Science 255:306–312, 1992.
46. Bazan, J.F. Structural design and molecular evolution of a cytokine receptor superfamily. Proc. Natl. Acad. Sci. U.S.A. 87:6934–6938, 1990.
47. Humphrey, W.F., Dalke, A., Schulten, K. VMD—Visual molecular dynamics. J. Mol. Graphics 14(1):33–38, 1996.
48. Anderson, C.M., Stenkamp, R.E., McDonald, R.C., Steitz, T.A. A refined model of the sugar binding site of yeast hexokinase B. J. Mol. Biol. 123:207–219, 1978.
49. Bennett, W.S., Steitz, T.A. Glucose-induced conformational change in yeast hexokinase. Biochemistry 75:4848–4852, 1978.
50. Lesk, A.M., Chothia, C. Mechanisms of domain closure in proteins. J. Mol. Biol. 174:175–191, 1984.
51. Balsera, M.A., Wriggers, W., Oono, Y., Schulten, K. Principal component analysis and long time protein dynamics. J. Phys. Chem. 100:2567–2572, 1996.