BUTTERWORTH HEINEMANN

Quantification and visualization of molecular surface flexibility

Carl-Dieter Zachmann, Stefan Michael Kast, and Jürgen Brickmann

Institut für Physikalische Chemie, Technische Hochschule Darmstadt, Darmstadt, Germany

Two new methods for the quantification and visualization of the flexibility of molecular surfaces are presented. Both methods rely on results of molecular dynamics (MD) simulations. Whereas method I is based on a simple but fast grid-counting algorithm, method II uses a mapping function that allows for a sharp and clear visualization of atomic RMS fluctuations on a molecular surface. To demonstrate the scope of the methods, MD simulations of two proteins, PTI and ubiquitin, were performed. The flexibility data are mapped onto the molecular surfaces of the proteins and visualized using texture mapping technology available on modern workstations.

Keywords: surface flexibility, molecular surface, molecular dynamics, proteins, computer graphics, molecular recognition, texture mapping

INTRODUCTION

The concept of a "molecular surface" is frequently used in many fields of computational chemistry.[1–13] Despite the great usefulness of this concept, almost all approaches suffer from a severe limitation: the flexibility of molecular surfaces is not taken into account. There is no doubt, however, that a rigid surface model gives only a rough impression of the scenario faced by one molecule approaching the surface of another (e.g., a protein). Particularly for the selectivity and specificity of enzymatic reactions, the flexibility of the compounds is extremely important.[2] Molecular recognition is not the pairwise identification of rigid surfaces. Speaking in terms of the lock-and-key principle, neither the lock nor the key are rigid, but may accommodate in such a manner that optimal interaction is ensured.

Up to now, there has been no unique way to quantify the matching of two arbitrarily shaped surfaces. The matching procedure should take into account the "softness" of molecular surfaces. If one knows that a certain part of a surface is flexible, one should not worry about local disagreement of the surfaces to be matched and take this flexibility into consideration when designing the matching strategy.

Sophisticated docking experiments must be designed in order to examine, for example, an enzymatic reaction in full detail. Our interest in the present work, however, is focused on a more general a priori characterization of molecular surface flexibility, suited for the interactive treatment of molecular recognition processes with the aid of high-performance computer graphics techniques as well as further nonvisual algorithmic processing.

In this work, two methodologically new approaches (termed method I and method II) for the quantification and visualization of surface flexibility are presented. The basic data for both approaches are supplied by molecular dynamics (MD) simulations, and the methods are applied to two proteins (pancreatic trypsin inhibitor [PTI] and ubiquitin).[14–18] The calculation and visualization of the local flexibility of the protein surfaces are based on the solvent accessible surface (SAS) concept introduced by Lee and Richards[3–5] and, in a modified version, by Connolly.[6,7] To improve the computer graphical representation, a triangulated version of the SAS is used.[8] The visualization of the numerical data (which are color coded on the SAS) is accomplished by techniques using the advanced facilities of high-performance graphics workstations, in particular texture mapping technology.[9,10]

Method I is based on a statistical analysis of the surface fluctuations of a protein during the MD run, determined by taking periodic "snapshots" of the protein. Although designed for proteins, the technique is not restricted to molecular systems. Any flexible surface may be analyzed if its position in space is well defined as a function of time.

Method II, which is conceptually simple and needs only a few hardware resources, relies on atomic root mean square (RMS) fluctuations, which can easily be calculated from the results of the MD simulations. By combining the

Color Plates for this article are on pages 110 and 111.

Address reprint requests to Dr. Brickmann at the Institut für Physikalische Chemie, Technische Hochschule Darmstadt, Petersenstr. 20, D-64287 Darmstadt, Germany.

This article is part of the Ph.D. thesis of Carl-Dieter Zachmann, Technische Hochschule Darmstadt (D17).

Received 15 September 1994; revised 1 November 1994; accepted 8 November 1994

655 Avenue of the Americas, New York, NY 10010

0263-7855/95/$10.00
SSDI 0263-7855(94)00015-K

texture-mapping technique with a simple but effective mapping function, method II provides a sharp and clear visualization of the RMS fluctuations of the atoms forming the surfaces of the proteins.

This article is organized as follows: In Methods, both approaches to the quantification of molecular surface flexibility are described. The procedure for generating the input data (i.e., molecular dynamics simulations and surface generation) and some technical details concerning the computer algorithms calculating local surface flexibilities from MD simulations are reported in the third section (see Procedure, below). The results obtained for PTI and ubiquitin are presented and discussed in the fourth section (see Results and Discussion below). The last section (Conclusion and Outlook) provides some concluding remarks and a short preview of future developments.

METHODS

Method I

Method I consists of two major parts. In the first part, a general procedure for the statistical description of a distribution of dots in three-dimensional (3D) space is proposed. On the basis of some specific assumptions concerning the surfaces involved in the analysis, in the second part a measure of molecular surface flexibility is introduced.

To ensure a unique procedure, some basic assumptions are made:

1. It is assumed that the density of surface dots (i.e., the number of dots per surface area) is identical on all surfaces S_i.
2. It is assumed that, considering one surface, the number of dots per volume is proportional to the surface area per volume.

The surface flexibility analysis is done in this work according to the following strategy. Let us consider a surface S_1, which may be defined by a set of surface points given by their Cartesian coordinates in 3D space. This surface can be considered to be embedded in a 3D grid of size $V_G = l_x \times l_y \times l_z$, which may be defined with respect to an arbitrarily fixed origin. This grid is then subdivided into $N_G = N_x \times N_y \times N_z$ cubic subunits of equal volume V_c (Figure 1). By counting the number of surface dots in each cube, Z_c, the relative number of dots, z_c, can be determined as

$$z_c = Z_c/V_G \tag{1}$$

Now let us consider a sequence of surfaces S_i that are related to the surface S_1. The surfaces $S_2, S_3, \ldots, S_M$ (M is the number of surfaces) that may be generated, for example, for different conformations of a molecule can now be characterized in the same manner as S_1. The relative number of surface points from $S_1, S_2, \ldots, S_M$ within a cube can be calculated as an average:

$$z_M = \frac{1}{MV_G} \sum_{i=1}^{M} Z_c(i) \tag{2}$$

The procedure results in a 3D data field representing the relative distribution of surface points in the grid.

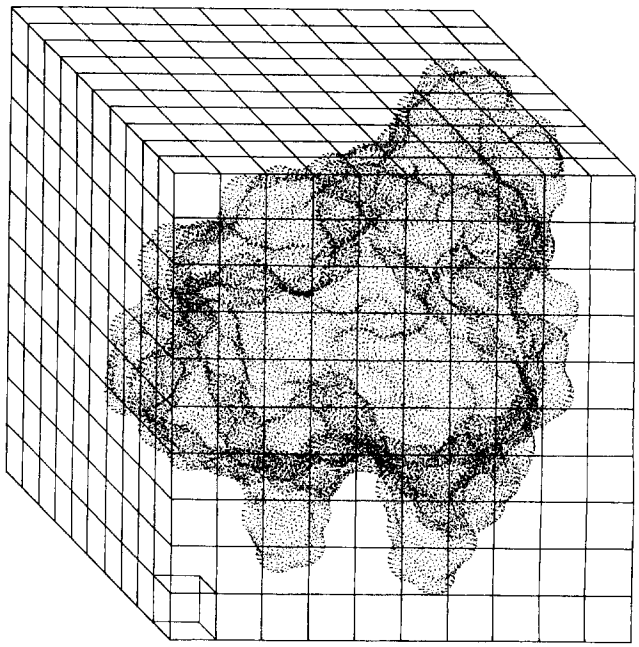

Figure 1. Dotted molecular surface embedded in a 3D grid.

Let now a surface S_R (reference surface) be defined by a set of points (reference dots, p_R) that are interconnected by a triangular mesh.[8] The explicit construction of the reference surface is not important at this stage, but it can be postulated that S_R is some kind of average of all the surfaces $S_1, \ldots, S_M$.

As a consequence of the triangulation, a normal vector $\mathbf{v}_n$ can be assigned to each reference dot p_R[8] (see Figure 2). In

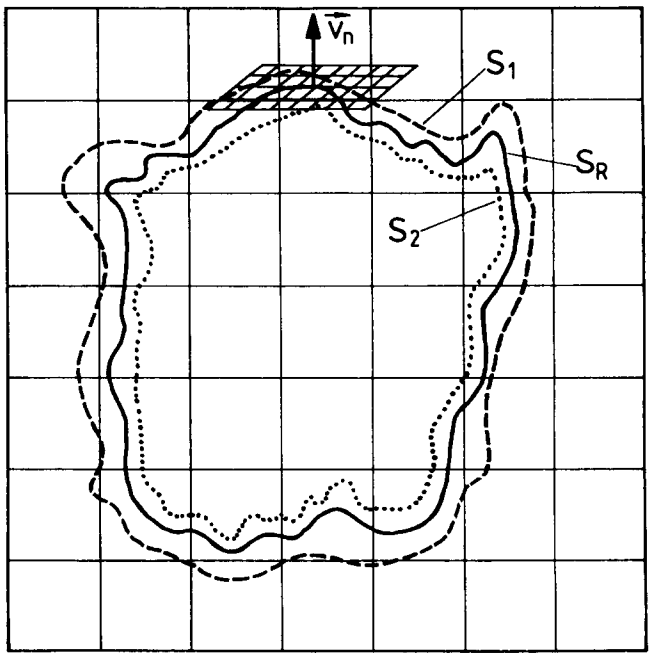

Figure 2. Schematic cut through three surfaces embedded in the 3D grid. The reference surface, S_R, is shown with a normal vector $\mathbf{v}_n$ standing orthogonal to the corresponding tangent plane which may be constructed for each surface dot.

the following considerations we restrict the discussion to one of these surface points. The calculations must be repeated for every dot of the reference surface.

The variance of the different surfaces at the position of the reference point can now be monitored by a histogram reflecting the distribution of dots in the grid with respect to the reference point. For this histogram, at set of "measuring points" p_{mi} along a predetermined reference direction (which can, in principle, be chosen arbitrarily) is constructed. For our purposes, the set of measuring points was defined to be an equidistant set in the direction of the normal vector $\mathbf{v}_n$ starting from the reference point

$$\mathbf{p}_{mi} = \mathbf{p}_R + \lambda_i \mathbf{v}_n \tag{3}$$

The scalars λ_i may take positive and negative values.

For all p_{mi}, the corresponding relative number of dots, that is, a quantity calculated according to Eq. (2) for the cube containing p_{mi}, can be determined by a simple index calculation. The construction of the resulting histogram is illustrated in Figure 3.

To avoid two adjacent measuring points being located within the same cube, the spacing of the points along the reference direction must be chosen appropriately. Here the scalars λ_i are chosen such that the distance between two points p_{mi} is larger than the space diagonal of a grid cube.

The distribution represented by the histogram can be characterized by mean and standard deviation [cf. Eqs. (4) and (5), respectively],

$$\bar{z}_M(p_R) = \frac{1}{N_B} \sum_{i=1}^{N_B} z_M(i) \tag{4}$$

$$\sigma(p_R) = \sqrt{\frac{1}{N_B - 1} \sum_{i=1}^{N_B} [z_M(i) - \bar{z}_M(p_R)]^2} \tag{5}$$

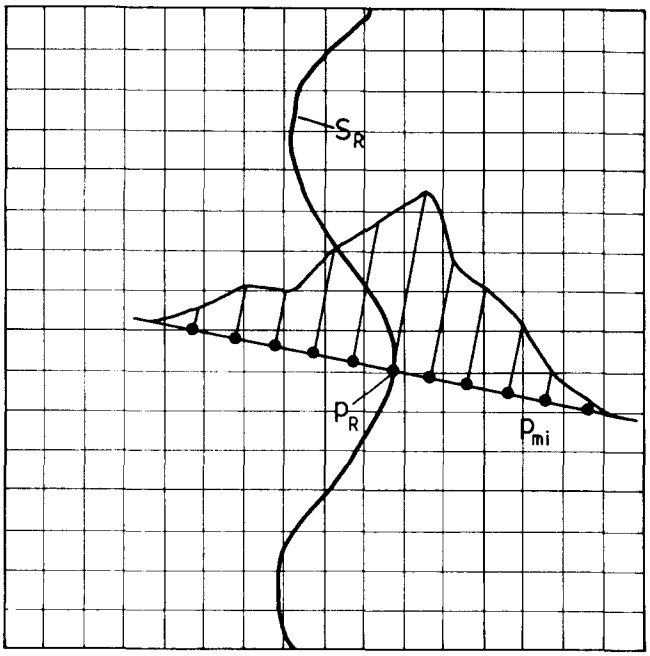

Figure 3. Discrete distribution of the averaged relative number of surface dots around reference dot p_R. The measuring points p_{mi} are used to set up the histogram.

where N_B is the number of bins of the histogram, which is in our case identical with the number of measuring points.

Up to now, there has been no need for the specification of the reference surface. At this point, some further assumptions concerning the reference surface S_R and the set of surfaces S_i are made in order to derive a physically meaningful measure of molecular surface flexibility.

It is postulated that the set of surfaces S_i provides an appropriate representation of the structural fluctuations of the reference surface, which is defined to be the surface of the average structure of the protein calculated from the results of the MD simulation by averaging the positions of the atoms. Moreover, the reference direction for setting up the histogram is defined to be the direction of the normal vector $\mathbf{v}_n$ assigned to each dot of the reference surface during the triangulation procedure,[8] as pointed out earlier in this section.

Local surface flexibility is now defined as the standard deviation $\sigma(p_R)$ of the local dot density distribution [cf. Eq. (5)], thus describing flexibility orthogonal to the tangent plane of the reference surface. This quantity $\sigma(p_R)$, which may be calculated for each surface point of the reference surface (see Figure 2),[8,11] strongly depends on the choice of both S_R and the set of S_i. Consequently, these surfaces must be chosen adequately.

Method II

The second method is based on atomic RMS fluctuations, which can be easily calculated from the results of an MD simulation. Method II takes advantage of a mapping strategy that has been developed by our group.[19] The numerical value assigned to each surface dot (which is defined by a position vector $\mathbf{r}_S$) is calculated as a superposition of atomic contributions that are "projected" onto the surface by a special distance weighting incrementation procedure.[19] This procedure guarantees that only those atoms that are in the close neighborhood of the surface dot contribute significantly to the surface value $F(\mathbf{r}_S)$.[19]

$$F(\mathbf{r}_S) = \sum_i f_i(\mathbf{r}_S) \tag{6}$$

where $f_i(\mathbf{r}_S)$ is the contribution of the ith atom. This contribution is given by the following expression:

$$f_i(\mathbf{r}_S) = \mathrm{RMS}_i \mu_i(\mathbf{r}_S) \tag{7}$$

with an atomic fluctuations RMS_i and a normalized distance function $\mu(\mathbf{r}_S)$ that depends on the distance of the surface dot from the center of the atomic increment i:

$$\mu_i(\mathbf{r}_S) = N\, g(|\mathbf{r}_i - \mathbf{r}_S|;\, c,\, \delta) \tag{8}$$

For simplicity, we set

$$r = |\mathbf{r}_i - \mathbf{r}_S| \tag{9}$$

The "normalization" function N is then given by

$$N = \left[\sum_i g(r;\, c,\, \delta) \right]^{-1} \tag{10}$$

c and δ are characteristic parameters (termed proximity pa-

rameters[20]) that determine how the increments influence the overall value at the surface point given by $\mathbf{r}_S$.

In accordance with Heiden et al.,[19] for the distance function g an expression of the Fermi type[19] was chosen:

$$g(r;\, c,\, \delta) = \frac{\exp(-2c/\delta) + 1}{\exp[2(r - c)/\delta] + 1} \qquad (11)$$

The distance function $g(r)$ has no a priori physical meaning. It fulfills two conditions: It is smooth and has finite values for $r < c$, where c is a cutoff value termed the *proximity distance* of an atom. The proximity distance should be larger than any atomic van der Waals radius in the molecule under consideration. For distances $r > c$, the function values of $g(r)$ rapidly tend towards zero, that is, the corresponding atoms do not significantly contribute to the overall value mapped onto the surface. The proximity parameter δ controls the steepness of the decay of the function.[19,20]

The application of this mapping strategy leads to a sharp and clear display of atomic flexibility, as is demonstrated below.

PROCEDURE

Molecular surface flexibility may be derived from RMS fluctuations of the atoms of the molecule that contribute to the molecular surface. The required data can, in principle, be derived from temperature factors supplied by X-ray crystallography. Unfortunately, no information on hydrogen atoms is available from X-ray scattering. On the other hand, hydrogen atoms often play an important role for the rigidity of a molecular surface. The forming of effective hydrogen bonds, for example, is an essential pattern in protein chemistry.

Information on the positions of hydrogen atoms can be obtained from neutron scattering and 2D nuclear magnetic resonance (NMR) studies. However, only a few entries in the Brookhaven protein databank (PDB)[21,22] provide hydrogen positions. Simulations can fill this gap in experimental data. More recently, molecular dynamics simulations of proteins have become widely feasible.[23–32] Starting from structural data deposited in a PDB file, hydrogen atoms can be added according to standard distances and angles. Molecular dynamics simulations make it possible to observe the structural fluctuations of any protein with known crystal structure, at least in principle. From a practical point of view, the computational effort to simulate large proteins in a realistic environment (such as an aqueous solution) is still extremely high.

In the present study, we report on MD simulations of two small proteins, pancreatic trypsin inhibitor (PTI) and ubiquitin, serving as test cases for checking the scope of our measure of flexibility applied to complex molecular surfaces. Both proteins play an important role in biochemical processes. Pancreatic trypsin inhibitor is an extraordinarily well-examined protein with a highly refined crystal structure available.[14,15,23–30] It is part of the digestive system, regulating the activity of the proteolytic enzyme trypsin by competitive inhibition. Inhibition is achieved by forming a complex with trypsin, in which the substrate recognition site of trypsin (i.e., the "specificity pocket") is part of the large interface between the proteins.[18] Residue Lys-15 of PTI fits exactly into this specificity pocket.

Ubiquitin is probably present in all eukaryotic cells. It has been identified in the nucleus, in the cytoplasm, and on the cell surface membrane, but it appears that its primary role is in intracellular ATP-dependent protein degradation. Protein breakdown in this pathway requires the formation of covalent conjugates in which carboxyl terminals of ubiquitin molecules become attached to the target protein via amide linkages.[16,17] The overall structure of ubiquitin is extremely compact and tightly hydrogen bonded. The only portion of the molecule without significant intramolecular hydrogen bonding and close packing contacts is the COOH terminus, which protrudes from the structure, and thus is accessible by enzymes involved in formation or cleavage of peptide bonds.[16,17]

The primary purpose of the investigation presented in this article was to estimate the plausibility and reliability of the calculated local flexibilities. Therefore, the simulations were carried out in a crystalline environment including a sufficient number of water molecules. Under these circumstances, no large-scale conformational changes (e.g., rotations of side chains), rendering the evaluation of the results more difficult, were expected to occur.

Molecular dynamics simulations

Structural data to build up the initial configurations for the MD runs were taken from the PDB files 4PTI and 1UBQ for PTI and ubiquitin, respectively. In the case of PTI, the positions of deuterium and hydrogen atoms obtained by neutron scattering are included in the PDB file. For ubiquitin, hydrogen atoms were added according to standard distances and standard angles. All simulations were run with the program package CHARMM22, using the standard force field for proteins[31,32] according to the following scheme:

1. Building up an initial structure by modifying and extending the original PDB file
2. Energy minimization (EM) of the initial structure using the Adopted Basis Newton—Raphson (ABNR) algorithm implemented in CHARMM22
3. Equilibration
4. Sampling run yielding the coordinates of the conformations used for the calculation of surface flexibility

Pancreatic trypsin inhibitor We followed in principle the procedure described by Berendsen et al.,[30] but with some minor modifications. Starting from the structure deposited in the PDB entry 4PTI, which includes 892 protein atoms (58 residues) and 57 water molecules, a unit cell (space group $P2_12_12_1$, four asymmetric units) was constructed using the CRYSTAL facility of CHARMM. All deuterium atoms were replaced by hydrogens. Three hundred and twenty-four water molecules were added, yielding a density of 3.31×10^{-3} Å^{-3}. To neutralize the unit cell having a net charge of $+24e$, 24 water molecules were substituted by chloride ions. The water molecules to be substituted were determined by calculating the electric potential at all water oxygens and replacing the 1 at the highest potential, which was repeated 24 times.[30] The initial configuration

contained 3 568 protein atoms, 552 water molecules, and 24 chloride ions, that is, a total of 5 248 atoms.

The high potential energy of this configuration, which is due to a large number of close contacts within the unit cell, was relaxed by ABNR energy minimization.

Throughout the MD simulation, periodic boundary conditions corresponding to the crystal translational symmetry were applied. A cutoff radius of 11 Å was used. This value is chosen slightly smaller than half the value of the smallest lattice parameter (b = 23.4 Å), ensuring that an atom cannot interact with both another atom and its periodic image.[30] A shifted potential and a switching function were applied to smooth cutoff effects when calculating electrostatic and van der Waals interactions, respectively. The time step was set to Δt = 2 fs. All bonds to hydrogen atoms were kept rigid by the SHAKE method, and the TIP3P water model was used.

After initializing the atomic velocities according to a Gaussian distribution at 300 K, the system was equilibrated. A window around the final temperature (300 K) was specified, and the temperature was brought into the desired range by appropriate scaling of the velocities. After 15 ps, the average temperature of the system had stabilized. The ratio of total energy to kinetic energy fluctuations had dropped to 0.04.[33]

After 10 ps of further equilibration without any scaling of velocities, the sampling run was performed in the microcanonical ensemble. For 40 ps, the momentary configuration of the complete unit cell was stored every 0.4 ps, that is, 100 configurations were accessible for further computational processing.

Ubiquitin The procedure for ubiquitin is nearly identical to the one described above for PTI. Owing to the missing data concerning the hydrogen positions, 629 H atoms were added to the 602 heavy atoms (76 residues) deposited in the original PDB file (PDB entry 1UBQ) using the HBUILD facility provided by CHARMM. Because the unit cell (space group $P2_12_12_1$, four asymmetric units) is electrically neutral, no counterions are required. Six hundred and ninety-six water molecules were added to the 232 crystal water molecules generated using the 58 oxygen positions provided by the PDB file, thus yielding a density of 3.31×10^{-3} Å^{-3}. Finally, the system to be simulated (i.e., the initial configuration of the complete unit cell) contained 4 924 protein atoms and 696 water molecules, for a total of 7 012 atoms. The technical details of the simulation were the same as given above for PTI.

Surface generation

The main advantage of the SAS model[3–7] is the adequate representation of protein surfaces. The MS algorithm proposed by Connolly,[6,7] which is based on the idea of rolling a test sphere along a CPK model of the protein and calculating the contact surface, smooths over grooves and cavities that are small relative to the solvent molecule. Roughly speaking, the MS program generates a surface as "seen" by a solvent molecule. We used a probe sphere with radius 1.4 Å (that is, the radius of a water molecule) throughout the calculations.

The reference surface S_R was defined to be the surface of the average structure, which is determined by averaging the coordinates of each atom. To obtain the average structures of PTI and ubiquitin, the coordinates of 100 conformations of the molecule concerned, covering 40 ps of the MD simulation, were used.

The reference surface was triangulated[8] for two reasons. First, a triangular mesh is well suited for the representation of surfaces. The raster graphical representation of surfaces by sequences of shaded polygons is much faster than any type of pixel-oriented technique as used in ray-tracing algorithms. Triangles have been found to show a number of advantages compared to other polygons.[34] Modern high-end workstations support fast imaging of triangular meshes. It has been shown that such meshes can be generated from dotted surface representations with no other information than the 3D coordinates of the surface points.[8] Second, the reference direction, which must be defined for each surface dot within method I, is provided by the normal vector generated during the triangulation procedure.

The technique for the visualization of the calculated flexibilities is described below (see Visualization Techniques).

Calculation of local flexibilities

The local flexibilities for the two proteins were calculated along the lines given in Methods (above). Here we discuss some aspects of the practical implementation of method I.

The numerical results are influenced by at least four parameters:

- The number of dots (generated by the MS program) representing a surface
- The number of surfaces (each of them generated for a single conformer) included in the calculation of the averaged relative number of dots in each cube according to Eq. (2)
- The grid spacing, given by the size of the grid, V_G, and the number of cubes, N_G
- The number of bins and their spacing in the histograms determining the discretization of the dot distribution

In general, a compromise must be found between accuracy and computational effort (CPU time and memory requirements are listed in Table 1).

The surfaces of PTI and ubiquitin are represented by approximately 40 000 and 50 000 dots, respectively, yielding constant dot densities of 12 Å^{-2} for both proteins. This number of dots is sufficiently large to ensure valid statistical analysis and adequate consideration of finer topographical details of the surface.

Similar arguments are valid regarding the second parameter. One hundred conformers covering a time span of 40 ps provide an accurate picture of the structural fluctuations of protein surfaces.

The third and fourth parameter are crucial for the spatial resolution of the analysis. Because fluctuations within a cube cannot be recognized, the cube size (i.e., V_c) should not be too large. On the other hand, computational effort rises rapidly with an increasing number of cubes, and extremely fine spatial resolution is not necessary. A cube size

Table 1. Typical CPU times and memory requirements for molecular dynamics simulations, surface generation, and calculation of local flexibilities according to methods I and II

	Surface generation[a] (one surface)			Calculation of local flexibilities[a]	
	MS program	Triangulation	MD simulations[b] sampling run (40 ps)	Method I (100 surfaces)	Method II
PTI					
CPU time	1.0 min	20 min	85 h	3 min	6 min
Memory (MB)	3	20	40	50	0.5
Ubiquitin					
CPU time	1.5 min	30 min	100 h	4 min	8 min
Memory (MB)	3	24	60	64	0.5

[a]Performance on a Silicon Graphics INDIGO R4000.
[b]Performance on an IBM RS/6000.

of 0.2 × 0.2 × 0.2 Å turned out to be an applicable compromise.

As pointed out in Methods, the minimum distance between two bins is determined by the lengths of the space diagonals of the cubes forming the grid. The number of bins was chosen such that a range of 4 Å (±2 Å) was covered, and N_B was therefore set to 11.

Visualization techniques

Visualization technology supported by modern graphics workstations opens the possibility to "see" molecular scenarios from the point of view of a molecule. It is common practice to use molecular surfaces as screens for the representation of different properties with different color-coding procedures. Color coding is a popular means of displaying scalar information on a surface.[35]

In interactive molecular graphics, high-contrast color code variation on sparsely tessellated geometry has so far been a serious problem. In this work, a new technology—texture mapping—is used, which fully exploits the hardware capabilities of modern graphics workstations. While standard hardware implementations allow the specification of an RGB color triplet per surface vertex, color interpolation on pixels in between is performed in a linear way.[36] As a consequence, all entries in the color code lying outside the linear color ramp joining two RGB triplets are never taken into account. This artifact can be avoided by using texture-mapping techniques.

Texture mapping is a technique that applies an image to the surface of an object. The image exists in a parametric coordinate space called the texture space.[9,37–39] Mapping the calculated property into texture space instead of color space ensures that the coloring evaluated at every pixel is taken from information lying between the values of the relevant vertices. High contrast variation in the color code is then possible, even on sparsely tessellated surfaces.

It is important to note that, although the texture is one-dimensional, it is possible to tackle a three-dimensional problem because the dimensionality of the texture space does not affect the object space. The independence of texture and object coordinate space is well suited to accommodate immediate changes to the meaning of the color (i.e., by applying simple 3D transformations in texture space). Translation allows readjustment of the zero line of the color code, while scaling of the texture changes the range of the mapping. Such modifications may be performed in real time.

Similar to the one-dimensional texture used as a color code on a molecular surface, the texture space may be extended to two or even three dimensions, incorporating additional information with each additional dimension, such that a maximum of three independent properties can be simultaneously visualized.[9]

Performance

Typical CPU times and memory requirements for surface generation, MD simulations, and calculation of local flexibilities are listed in Table 1. The data clearly show that the generation of input data, that is, molecular dynamics and surface generation, is extremely time consuming compared to the actual calculation of local flexibilities. Within both methods, a triangulated SAS of the average structure derived from the MD run must be generated. Method I requires, in addition, the generation of 100 surfaces (1 for each conformer) with the Connolly MS program.[6,7] Consequently, the computational effort for method I is considerably higher. For method II, both the CPU time and the memory requirements are low.

RESULTS AND DISCUSSION

Color Plates 1 and 2 show the surface of PTI with color-coded local surface flexibilities calculated according to method I and RMS fluctuations of the surface atoms mapped onto the surface according to method II. The same is shown for ubiquitin in Color Plates 3 and 4.

It turned out that some "hot spots" originating from large fluctuations within small regions made the display of the local flexibilities provided by method I somewhat difficult to survey. To smooth the data without loss of resolution, we did not enlarge the cube size V_c. Instead, a distance-weighted average of the quantity z_M [see Eq. (2)]

corresponding to the measuring point p_{mi} [see Eq. (3)] itself and to its 26 neighbored cubes in the grid was calculated.

$$z_{27} = \frac{\left[\sum_{i=1}^{27} \omega(i) z_M(i)\right]}{\left[\sum_{i=1}^{27} \omega(i)\right]} \qquad (12)$$

with

$$\omega(i) = 1$$

for the "central" cube and the 6 neighbors connected to it by a shared face,

$$\omega(i) = \tfrac{1}{2}\sqrt{2}$$

for the 12 neighbors connected to the central cube by a shared edge, and

$$\omega(i) = \tfrac{1}{3}\sqrt{3}$$

for the 8 neighbors connected to it by a shared vertex. This averaging is done for all measuring points p_{mi} defining the positions of the bins in the histogram. The standard deviation [Eq. (5)] is then calculated by substituting z_M by z_{27}.

There are two major advantages of the local flexibilities determined according to method I. First, a representation of surface flexibility along a freely adjustable reference direction (in this study, orthogonal to the reference surface) is provided, while no direction dependency can be derived from atomic RMS fluctuations displayed with method II. Second, the local flexibilities provided by method I are calculated by definition with respect to the dotted contact surface itself. Consequently, fluctuations of the surface originating from the superposition of atomic movements are taken into account adequately. This is not the case for the data calculated within method II, leading to some striking differences between the graphical representations generated with the two methods. By combining high resolution with the benefits of texture mapping, method I leads to a detailed but clear representation of surface flexibility (see Color Plates 1–4).

As the RMS fluctuations used as input data are averaged quantities, method II provides a clear representation of the flexibility of even large and complex surfaces. Owing to the mapping function and the texture-mapping technique, there are gentle but high-contrast transitions between areas exhibiting different flexibility. Furthermore, the only input data required (i.e, RMS values) can be calculated directly from the results of the MD run. In contrast to method I, only one relatively time-consuming surface generation procedure is necessary, and the computational cost of the mapping process is low (see Table 1).

It seems to us that method I is clearly preferable if an accurate and detailed representation is demanded, and, if this is the case, the higher computational effort compared to method II should be accepted. Method II is well suited if one is interested in a general overview of surface flexibility.

Both PTI and ubiquitin have a compact tertiary structure, and thus are well suited to demonstrate the scope of both methods, particularly with respect to sensitivity and resolution. Despite the limited mobility of the protein atoms in the crystal lattice, the flexibility of the surface is clearly recognized. The texture-mapping technique allows for a sharp differentiation between surface areas with various flexibility. Coherent domains of equal flexibility can also be identified unequivocally.

Pancreatic trypsin inhibitor

The compactness of PTI is reflected by the low mobility of large surface regions (colored in blue). Among the regions with a higher degree of flexibility are some of the exposed side chains. Residue Lys-15, however, is rigid. This rigidity may be due to the high specificity of this enzyme in binding to the active site of trypsin. Forming a stable complex with another protein either requires a high degree of complementarity of the surfaces (otherwise the rigid surface of PTI will not match with the surface of this protein) or the protein must be flexible enough to adapt to PTI.

Some interesting detail is revealed at the bottom of Color Plates 1c and 1d. As a consequence of the short distance between the carboxyl group of the C terminus and the ϵ-amino group of the N-terminal arginine residue, some surface dots between these residues (forming some kind of bridge) were generated by the MS program. As to be expected, this bridge is highly flexible.

Ubiquitin

The greater part of the surface of ubiquitin is relatively rigid. This is also true for the C terminus, which is of major importance for the biological function of this enzyme (see above). As revealed by the display of the local surface flexibility according to method I, some parts of the surface of the C terminus are flexible. A certain degree of flexibility of this part of the molecule is definitely required, because ubiquitin forms covalent conjugates with a variety of proteins,[16,17] but this process is probably accompanied by conformational changes of both ubiquitin and the target molecule (induced fit).

In general, it should be remembered that the discussion given above is somewhat speculative in nature, because the packing of the protein molecules in the crystal strongly influences the flexibility of the surface, and only small conformational changes can occur during the MD simulation of the crystal.

Let us briefly consider the strengths and weaknesses of our methods. Starting from structural data, information on surface flexibility with respect to an individual surface dot and, in the case of method I, a freely adjustable reference direction is available. The calculated flexibility values may be mapped onto an appropriate reference surface, which is supported by a flexible and powerful visualization technique, or can be used as input data for further nonvisual algorithmic processing. While it is still possible to display and manipulate the rigid reference surface, information on the dynamic behavior can be obtained whenever desired. Owing to the generality of both methods, there is a wide range of possible applications apart from quantifying the structural fluctuations of protein surfaces. Depending on the

kind of question to be answered, the suitable method may be chosen.

The main disadvantage of these concepts applied to molecular surfaces is clearly the need to run extensive molecular dynamics simulations to describe the time-dependent behavior of the surface, which is accompanied by large computational effort, particularly if macromolecules such as proteins are investigated. With the rapidly increasing computing power and memory capacities already provided by midpriced workstations, however, this problem can be handled with reasonable expense.

CONCLUSION AND OUTLOOK

We have presented two generally applicable, but quite simple and fast, procedures to quantify and visualize local surface flexibility. Both methods are based on a dotted surface representation and rely on "snapshots" of the flexible surface taken at different times. To supply the required data for two protein surfaces used to demonstrate the scope of the approach, extensive molecular dynamics simulations were carried out. The calculated local flexibilities were visualized by taking advantage of powerful graphic facilities, particularly texture mapping. The computer images allow for a fast and quantitative estimation of the flexibility even of complex molecular surfaces.

It should be kept in mind, however, that, up to now, our investigations are restricted to proteins in a crystalline environment. Careful studies of protein dynamics in aqueous solution with respect to the scope and limitation of our methodology are in preparation. In addition, we shall concentrate on detailed investigations concerning the influence of surface flexibility on chemical reactions and the usefulness of our concepts in this context.

ACKNOWLEDGMENTS

The authors thank S. Reiling and H. Vollhardt for their cooperation. This work was supported by the Fonds der Chemischen Industrie (Frankfurt am Main, Germany).

REFERENCES

1 Mezey, P. Molecular surfaces. In: *Reviews in Computational Chemistry,* (Lipkowitz, K. and Boyd, D., Eds.), VCH, Weinheim, 1990. pp. 265–294, and references given therein

2 Fisher, C.L., Tainer, J.A., Pique, M.E., and Getzoff, E.D. Visualization of molecular flexibility and its effect on electrostatic recognition. *J. Mol. Graphics* 1990, **8,** 125–132

3 Lee, B. and Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 1971, **55,** 379–400

4 Richards, F.M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 1977, **6,** 151–176

5 Richards, F.M. Calculation of molecular volumes and areas for structures of known geometry. *Methods Enzymol.* 1985, **115,** 440–464

6 Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Cryst.* 1983, **16,** 548–558

7 Connolly, M.L. Solvent accessible surfaces of proteins and nucleic acids. *Science* 1983, **221,** 709–713

8 Heiden, W., Schlenkrich, M., and Brickmann, J. Molecular surface triangulation. *J. Comp.-Aided Mol. Design* 1990, **4,** 255–269

9 Teschner, M., Henn, C., Vollhardt, H., Reiling, S., and Brickmann, J. Texture mapping: A new tool for molecular graphics. *J. Mol. Graphics* 1994, **12,** 98–105

10 Heiden, W., Reiling, S., Vollhardt, H., Zachmann, C.-D., and Brickmann, J. Interactive visualization of molecular scenarios with the MOLCAD/SYBYL package. In: *Insight and Innovation in Data Visualization* (Bowie, J.E., Ed.), Manning Publications, Greenwich, 1995, pp. 83–97

11 Zachmann, C.D., Heiden, W., Schlenkrich, M., and Brickmann, J. Topological analysis of complex molecular surfaces. *J. Comp. Chem.* 1992, **13,** 76–84

12 Zachmann, C.D., Kast, S.M., Sariban, A., and Brickmann, J. Self-similarity of solvent accessible surfaces of biological and synthetical macromolecules. *J. Comp. Chem.* 1993, **14,** 1290–1300

13 Heiden, W., and Brickmann, J. Segmentation of protein surfaces using fuzzy logic. *J. Mol. Graphics* 1994, **12,** 106–115

14 Wlodawer, A., Walter, J., Huber, R., and Sjolin, L. Structure of bovine pancreatic trypsin inhibitor. Results of joined neutron and X-ray refinement of crystal form II. *J. Mol. Biol.* 1984, **180,** 301–329

15 Wlodawer, A., Deisenhofer, J., and Huber, R. Comparison of two highly-refined structures of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* 1987, **193,** 145–156

16 Vijay-Kumar, S., Bugg, C.E., and Cook, W.J. Structure of ubiquitin refined at 1.8 Å resolution. *J. Mol. Biol.* 1987, **194,** 531–544

17 Vijay-Kumar, S., Bugg, C.E., Wilkinson, K.D., Vierstra, R.D., Hatfield, P.M., and Cook, W.J. Comparison of the three-dimensional structures of human yeast, and oat ubiquitin. *J. Biol. Chem.* 1987, **262,** 6396–6399

18 Bode, W. and Schwager, P. The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. *J. Mol. Biol.* 1975, **98,** 693–717

19 Heiden, W., Moeckel, G., and Brickmann, J. A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *J. Comp-Aided Mol. Design* 1993, **7,** 503–514

20 Pixner, P., Heiden, W., Merx, H., Möller, A., Moeckel, G., and Brickmann, J. An empirical method for the quantification and localization of molecular hydrophobicity. *J. Chem. Inf. Comput. Sci.* 1994, **34,** 1309–1319

21 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., and Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977, **112,** 535–542

22 Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., and Weng, J. Protein data bank. In: *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen, F.H., Berger-

hoff, G., and Sievers, R., Eds.), Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, 1987, pp. 107–132
23 McCammon, J.A. and Harvey, S.C. *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge, 1987
24 Van Gunsteren, W.F., Berendsen, H.J.C., Hermans, J., Hol, W.G.J., and Postma, J.P.M. Computer simulation of the dynamics of hydrated protein crystals and its comparison with X-ray data. *Proc. Natl. Acad. Sci. U.S.A.* 1983, **80,** 4315–4319
25 Van Gunsteren, W.F. and Berendsen, H.J.C. Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. *J. Mol. Biol.* 1984, **176,** 559–564
26 Yu, H., Karplus, M., and Hendrickson, W.A. Restraints in temperature-factor refinement for macromolecules: An evaluation by molecular dynamics. *Acta Cryst.* 1985, **B41,** 191–201
27 Levitt, M and Sharon, R. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. U.S.A.* 1988, **85,** 7557–7561
28 Ichiye, T. and Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Struct. Funct. Genet.* 1991, **11,** 205–217
29 Kossiakoff, A.A., Randal, M. Guenot, J., and Eigenbrot, C. Variability of conformations at crystal contacts in BPTI represent true low-energy structures: Correspondence among lattice packing and molecular dynamics structures. *Proteins: Struct. Funct. Genet.* 1992, **14,** 65–74
30 Berendsen, H.J.C., van Gunsteren, W.F., Zwinderman, H.R.J., and Geurtsen, R.G. Simulations of proteins in water. *Ann. N.Y. Acad. Sci.* 1986, **482,** 269–286
31 Brooks, B.R., Bruccoleri, R.E., Olafsen, B.D., States, D.J., Swaminathan, S., and Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 1983, **4,** 187–217
32 Nilsson, L. and Karplus, M. Empirical energy functions for energy minimization and dynamics of nucleic acids. *J. Comp. Chem.* 1986, **7,** 591–616
33 Kitchen, D.B., Hirata, F., Westbrook, J.D., Levy, R., Kofke, D., and Yarmush, M. Conserving energy during molecular dynamics simulations of water, proteins, and proteins in water. *J. Comp. Chem.* 1990, **11,** 1169–1180
34 Zauhar, R.J. and Morgan, R.S. The rigorous computation of the molecular electric potential. *J. Comp. Chem.* 1988, **9,** 171–187
35 Dill, J.C. An application of color graphics to the display of surface curvature. *Computer Graphics* 1981, **15,** 153–161
36 Gouraud, H. Continuous shading of curved surfaces. *IEEE Trans. Computers* 1971, **20,** 623–628
37 Catmull, E. A Subdivision Algorithm for Computer Display of Curved Surfaces. Ph.D. thesis, University of Utah, Salt Lake City, Utah, 1974
38 Crow, F.C. Summed area tables for texture mapping. *Computer Graphics* 1984, **18,** 207–212
39 Williams, L. Pyramidal parametrics. *Computer Graphics* 1983, **17,** 1–11