

Wavelets and molecular structure*

Mike Carson

*Center for Macromolecular Crystallography, University of Alabama at Birmingham,
274 BHS, 79 THT University Station, Birmingham, AL 35294, U.S.A.*

Received 25 September 1995

Accepted 10 February 1996

Keywords: B-splines; NURBS; Ribbons; Protein editing; Protein fold; Low-resolution models

Summary

The wavelet method offers possibilities for display, editing, and topological comparison of proteins at a user-specified level of detail. Wavelets are a mathematical tool that first found application in signal processing. The multiresolution analysis of a signal via wavelets provides a hierarchical series of ‘best’ lower-resolution approximations. B-spline ribbons model the protein fold, with one control point per residue. Wavelet analysis sets limits on the information required to define the winding of the backbone through space, suggesting a recognizable fold is generated from a number of points equal to 1/4 or less the number of residues. Wavelets applied to surfaces and volumes show promise in structure-based drug design.

Introduction

Wavelet multiresolution analysis produces a hierarchical series of best lower-resolution approximations in a least-squares sense. Wavelets can decompose any function: a two-dimensional image, a curve, a surface, or a volume. This provides an elegant representation of the level of detail present.

A multiresolution representation for end-point interpolation of B-spline curves was developed using wavelets by Finkelstein and Salesin [1]. This multiresolution curve (MRC) formulation allows continuous levels of smoothing and approximating the curve with a given error tolerance. It also allows editing of the overall sweep of the curve at a lower resolution, while maintaining the fine details at high resolution.

The protein folding process, where the linear sequence of amino acids determines the unique three-dimensional path of the backbone, remains a major unsolved problem in biology and chemistry. Ribbon splines are used to model the backbone. The wavelet analysis should establish limits on the minimum information required to specify the winding of the backbone through space.

The manipulation and comparison of protein backbones using computers and graphics is vital to protein

crystallography and molecular modeling. Multiresolution comparison and editing are powerful new tools. Techniques to interactively make large changes in a protein backbone with a minimal amount of user manipulation should improve user interfaces for molecular modeling.

Molecular surface complementarity is critical in the binding of ligand to receptor. The ‘docking problem’ is central to computer-aided structure-based drug design. The problem has been addressed in many ways [2], all computationally intensive. Wavelets extended to surfaces and texture mapping apply to any molecule and should yield simplified representations. To assess the potential of wavelets applied to molecular structure, the ideas suggested by Finkelstein and Salesin [1] were implemented as interactive computer graphics programs.

Methods

Wavelets

Wavelets were discovered as a mathematical tool in approximation theory for the hierarchical decomposition of functions [3] and first found application in signal processing [4]. The discrete wavelet transform is a fast linear operation, like the fast Fourier transform. Whereas the latter is described in terms of sines and cosines, the basis

*This paper is based on a presentation given at the 14th Molecular Graphics and Modelling Society Conference, held in Cairns, Australia, August 27–September 1, 1995.

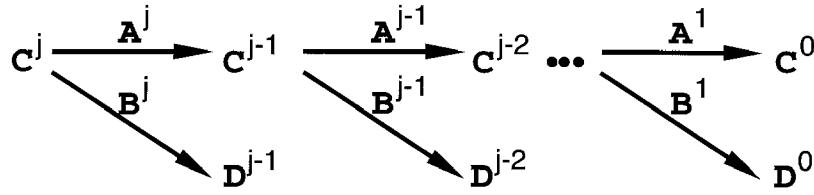


Fig. 1. The filter bank.

functions for the former are ‘wavelets’. Individual wavelet functions have a characteristic frequency like sines. Unlike sines, individual wavelet functions are localized in space. The discrete wavelet transform operates on a data vector of length equal to an integer power of two, returning a numerically different vector of the same length [5].

B-splines and wavelets

B-splines are a standard curve-fitting primitive, provided by most computer graphics systems. Given coordinates of $(n+3)$ ‘control points’ and the default cubic polynomial basis functions, a curve with n smoothly joined segments is drawn. Only the four control points bracketing each segment have any influence on its shape. The curve generally does not pass through the control points, except by special construction at the end points.

Following Finkelstein and Salesin [1], our discrete signal is the set of control points \mathbf{C}^j (for example, the set of n C^α coordinates) defining the B-spline curve. A down-sampling filter to create a lower-resolution version \mathbf{C}^{j-1} with n' coordinates is an n by n' matrix \mathbf{A}^j :

$$\mathbf{C}^{j-1} = \mathbf{A}^j \mathbf{C}^j \quad (1)$$

The coordinate detail lost in the downsampling process above can be captured by another filter, the $(n - n')$ by n matrix \mathbf{B}^j :

$$\mathbf{D}^{j-1} = \mathbf{B}^j \mathbf{C}^j \quad (2)$$

The correctly chosen \mathbf{A}^j and \mathbf{B}^j matrices are called analysis filters; splitting the input \mathbf{C}^j into the lower-resolution \mathbf{C}^{j-1} and detail \mathbf{D}^{j-1} is called decomposition. The original input can be recovered by another pair of matrices, the synthesis filters \mathbf{P}^j and \mathbf{Q}^j , through the reconstruction:

$$\mathbf{C}^j = \mathbf{P}^j \mathbf{C}^{j-1} + \mathbf{Q}^j \mathbf{D}^{j-1} \quad (3)$$

The decomposition applied recursively is known as the filter bank (Fig. 1).

The original signal \mathbf{C}^j can then be reconstructed from the sequence $\mathbf{C}^0, \mathbf{D}^0, \mathbf{D}^1, \dots, \mathbf{D}^{j-1}$. This sequence has the same size as the original signal and is known as its wavelet transform.

The B-spline curves are generated from the control points and scaling functions (or blending functions or basis functions):

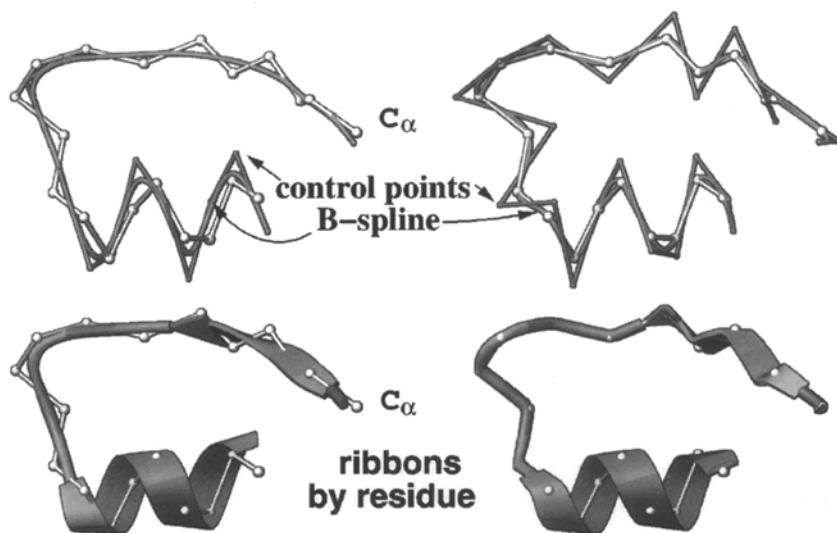


Fig. 2. Ribbon model construction. Peptide plane ribbon (left) versus C^α ribbon (right). The former uses one control point per peptide plane, the latter uses one control point per α -carbon. For each, two additional points are required at the termini to start and end the B-spline curve. Each four consecutive points define one segment of the curve. A sheet-turn-helix section of the protein ubiquitin [47] is shown. The top portion shows the C^α trace as white balls-and-sticks, the B-spline control points as smaller dark-gray balls-and-sticks, and the B-spline curve as a light-gray tube. The lower portion shows the C^α trace along with the ribbon drawing, colored by residue type. Each B-spline segment corresponds to one residue in the peptide-plane-based curve. Each segment is split in half, with adjacent halves flanking the C^α for that style.

$$f^j(u) = \Phi^j(u) C^j \quad (4)$$

The basis functions, Φ^j , are cubic polynomials that interpolate the control points to create the parametric curve, $f^j(u)$. The relationship between the number of control points, n , and the integer resolution level, j , of C^j is required to be: $n = 3 + 2^j$.

The basis functions, Ψ^j , that are orthogonal to every Φ^j under the chosen inner product, are called wavelets. Orthogonality implies that no redundant information exists in the coefficients of the wavelet transform. The linear algebra and derivation of the orthogonal wavelets for B-splines have been given and the scaling functions for the B-splines and wavelets have been plotted [1]. The subject is also presented in a tutorial [6,7]. The P^j and Q^j matrices given provide sufficient information to implement the method.

Ribbon models and the protein fold

The ribbon diagram is an excellent way to represent a protein fold by approximating the path of the backbone and revealing hydrogen-bonding patterns. The protein backbone may be modeled as a B-spline ribbon, defined by one point per peptide plane [8]. A 'ribbon space curve' spline is the basis of a variety of styles of ribbon drawings [9]. Two points on each terminus are used to start and stop the curve and N peptide planes imply $(N+1)$ residues, thus the number of control points is $n = (3+N)$ residues. This conveniently sets the number of curve segments equal to the number of residues. Complete recipes, including relevant information gleaned from the Evans & Sutherland and Silicon Graphics manuals, may be found in the description of the program *ribbons 2.0* [10].

The 'best' curve to specify a backbone is an open question. At issue are the path to be traced and the underlying order and basis of the defining polynomial. I contend that the cubic polynomial B-spline formulation is the minimum needed for computer graphics to capture the feel of the drawings in Richardson's monograph [11]. There are many possible formulations*.

MRC analysis of protein backbones

The construction of a ribbon is shown in Fig. 2. The spline curves rarely pass through the C^α positions in the standard ribbon drawings. Sheets are not pleated and coils are smoothed out as the spline is fit to the peptide planes. The spline can easily be fit through the C^α atoms, but the results are not as visually appealing. C^r , the center of the curve segment corresponding to a residue, is the same point as C^α for the C^α -based ribbons of Fig. 2. For the peptide-plane-based ribbons of Fig. 2, C^r is systemati-

cally displaced, as explained above. The C^α - C^r distance is used as a measure of fit during multiresolution analysis.

The multiresolution curve analysis consists of taking the control points defining the ribbon curve and successively calculating and displaying lower-resolution versions. The relationship between the number of control points, n , and the integer resolution level, j , of C^j is $n = 3 + 2^j$. The number of segments in the curve is 2^j . The number of segments equals the number of residues in the peptide-plane-based ribbons, but the number of residues in a protein is generally not an integer power of 2. Experimentation showed that padding of the signal with multiple copies of the chain termini had no effect on the shape of the ribbon curve. An example: $j = 7$ for a 128-residue protein, as $2^7 = 128$. For a 123-residue protein, five redundant points must be added to pad the signal to integer level 7. A maximum fractional level is taken as $j = \log_2(N_{\text{res}}) = 123 = 6.93$.

Co-workers were asked to display their favorite protein as a tube ribbon and to interactively adjust the fractional level. The adjustment was stopped at the lowest level that still looked like the structure with which they were so familiar. The ribbon was displayed with additional structural cues, e.g., coloring by secondary structure or making the width of the tube a function of secondary structure.

For a less subjective test, a representative sample of 230 proteins [12] was analyzed numerically. The distances between the C^α and the C^r position on the original ribbon and low-resolution versions were monitored as a function of the wavelet resolution level and the secondary structure type.

Topological comparison

Utilitarian uses of the protein backbone are found in the comparison, classification, and construction of proteins. Functional binding domains can be described, aligned, and compared [13]. Proteins can be grouped according to the topology of their folds [11]. The optimal superposition of protein backbones is critical in homology modeling [14]. Complete backbones may be fit piecewise to electron-density maps using a database of unrelated backbone conformations [15].

The wavelet formalism is useful for objective comparisons at the appropriate adjustable scale, allowing one to focus on supersecondary structures, motifs, domains, or complete proteins as need be. Visual comparisons were made. Manipulation involved only rigid-body superposition of structures. The root-mean-square (rms) deviations were calculated after a least-squares superposition of two coordinate sets. Superpositions are based on various sets of C^α and C^r at different resolutions.

Multiresolution editing

Crystallographers have created software packages that greatly facilitate the interpretation of crystallographic

*See references in the Methods of Enzymology preprint at URL: <http://www.cmc.uab.edu:8001/www/ribbons/menz.html>.

Multiresolution curve analysis of Calmodulin

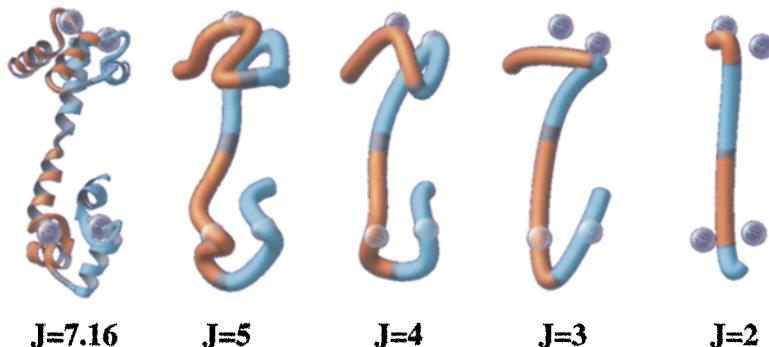


Fig. 3. MRC analysis of calmodulin. White spheres represent the position of the calcium ions in the structure. The 143 residues ($J = 7.16$) visible in the electron-density map define the ribbon drawing on the far left. The four nearly identical 'EF-hand' domains are given alternate colors. In the lower-resolution versions ($J = 5, 4, 3, 2$), the curve defined by the 32, 16, 8 or 4 segments is rendered as a thick tube.

maps (O [16]; XtalView [17]). Commercial packages have integrated the needs of the crystallographer within a more general molecular-modeling environment, e.g., INSIGHT II (Biosym Technologies, San Diego, CA), QUANTA (Molecular Simulations, Inc., Waltham, MA) and SYBYL (Tripos Associates, St. Louis, MO). One significant feature missing from current modeling programs is the ability to easily make very large changes to the protein backbone. The Sculpt [18] prototype tries to address this. The decomposition/reconstruction via wavelets makes this possible by allowing interactive selection of the level of approximation.

The matrix equations for reconstructing the maximum resolution version of a curve defined by \mathbf{C}^m after editing the control points of a lower-resolution version \mathbf{C}^j is derived [1] as:

$$\mathbf{C}_{\text{new}}^m = \mathbf{C}_{\text{old}}^m + \mathbf{P}^m \mathbf{P}^{m-1} \dots \mathbf{P}^{j+1} \Delta \mathbf{C}^j \quad (5)$$

Editing at fractional levels and making changes through direct manipulation can be done by the user, directly tugging the curve instead of the control point.

The mapping from the protein's atomic structure to the ribbon curve is well-defined. An inverse mapping is needed to go from an edited curve, as described above, back to the atomic structure. Each residue is treated as a rigid body. The tangent-normal-binormal (TNB) coordinate frame of the original ribbon curve is saved at each \mathbf{C}^j . The new frame, TNB', centered at the new \mathbf{C}^j , is used to determine the rotation/translation required to transform the coordinates. No other constraints are currently applied.

Molecular surfaces

The accessible surface of a macromolecule is a significant determinant of its action. The definition [19] and computer implementation [20] of such surfaces have had a profound effect on subsequent calculations and visualizations of molecular form and function. Usually such

surfaces are displayed as dots or triangles and may be color-coded to highlight chemical properties [21]. Various types of splines [22–24] and spherical harmonics [25] have also been used to model molecular surfaces.

Wavelet analysis of B-splines can be extended to tensor product surfaces [1]. NURBS (Non-Uniform Rational B-spline Surfaces) are now common in graphics libraries. A parametric surface, $S(u,v)$, is defined by B-spline functions in the u- and v- directions. Modeling of the surface of DNA with textured NURBS and the implementation of the program *DNurbs* have been described [24]. We also discuss how a 'globe' can be collapsed onto a globular molecule to create a set of NURBS (topologically a sphere, with singularities at the poles) to crudely approximate the molecular surface. For exact molecular surfaces, Connolly's Molecular Surface Program (MSP) [26] is used to create a triangulation for display with *ribbons*.

Texture maps for the *DNurbs* were created by a 64×64 sampling of the MSP triangular surface, colored by either electrostatics or curvature. The texture map for globe surfaces was created by a 64×64 sampling of the set of atomic spheres of the protein.

The multiresolution analyses of the spline surfaces above require special surface topologies. A recent paper describes multiresolution analysis of arbitrary meshes [27]. Orthogonal wavelet transforms are proposed [28] for obtaining a unique hierarchical shape description of volumetric data in the field of medical imaging.

Software

All code was developed in C++ under UNIX on an SGI Indy. A set of C++ classes implemented the multiresolution curve (MRC) analysis of B-splines, using the published formulas [1]. These MRC classes were used with the *ribbons 3.0* software to create the example images.*

*More information on the *ribbons* package can be found through the URL: <http://www.cmc.uab.edu:8001/www/ribbons/ribbons.html>.

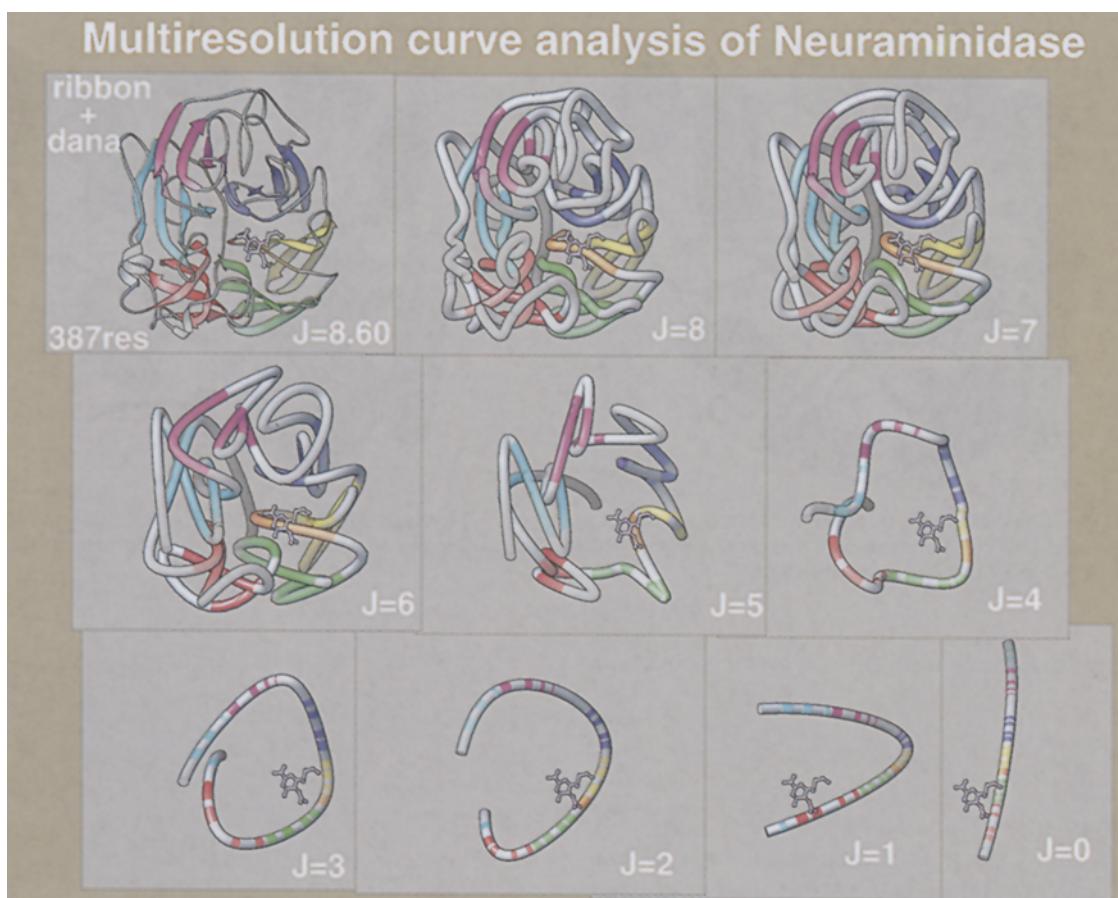


Fig. 4. MRC analysis of neuraminidase. A ball-and-stick drawing marks the inhibitor DANA (2-deoxy-2,3-dehydro-N-acetyl neuraminic acid) in the active site. The sheet residues are given different colors. The 389 residues ($J = 8.60$) define the ribbon drawing in the upper left. In the lower-resolution versions ($J = 8, 7, \dots, 0$), the curve defined by the 256, 128, ..., 1 segments is rendered as a tube.

The ‘ribbon dimension control panel’ has a set of Motif widgets for the MRC analysis. The user turns on the analysis to initialize, then uses a slider to adjust the MRC level. A toggle is used to restrict the MRC level to integer values. This all takes place in real time for a small protein on a low-end SGI Indy. The user can also smoothly interpolate to any fractional level in real time.

Several utility and prototype programs were developed with Inventor, SGI’s object-oriented 3D graphics toolkit. The program *bs-edit* uses the direct-manipulation features of the toolkit to edit the curves/protein structure in a multiresolution analysis; *mr-nurbs* uses the texture-mapping and NURBS features to create the multiresolution parametric surfaces. Texture maps were created by a $64 \times$

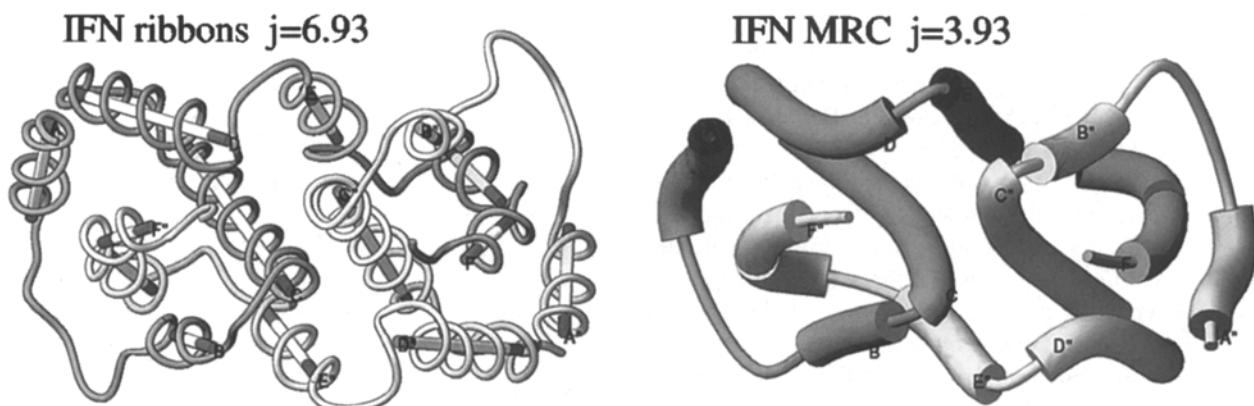


Fig. 5. MRC visualization of interferon. Left: the interferon- γ dimer is rendered as a tube ribbon ($j = 6.93$). Cylinders fit to the helical residues are shown. Right: the MRC version with $j = 3.93$. The tube has a larger diameter for helices.

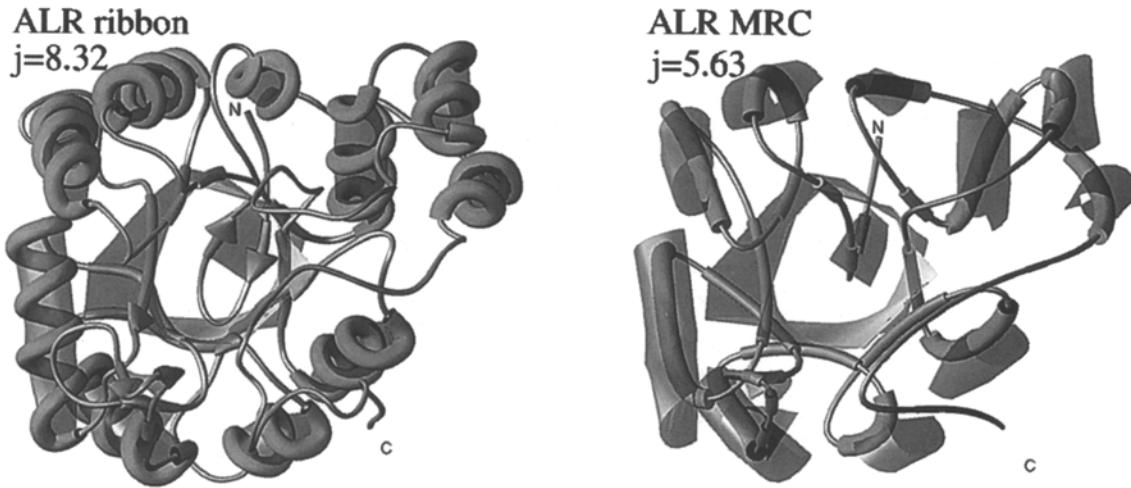


Fig. 6. MRC visualization of aldehyde reductase. Left: aldehyde reductase ($j = 8.32$) is shown with transparent polygons connecting the hydrogen-bonding regions of this TIM barrel protein. The protein is rendered as a tube colored by secondary structure, with diameter as follows: sheet > helix > other. Right: the MRC version with $j = 5.63$.

64 sampling of surfaces using the ray-casting objects in Inventor.

Results

MRC analysis of protein backbones

Figure 3 shows the multiresolution curves of calmodulin [29], and Fig. 4 depicts the MRC analysis of neuraminidase [30]. Both figures show only the discrete integer resolution versions of the curves.

Interferon- γ [31] is shown in Fig. 5. Each monomer has 123 residues, implying $j = 6.93$. The B-spline curve defined by 32 segments ($j = 5$) generally passes through the cylinders. The fractional level was lowered to visually assess the exact level where the topology is still recognizable. This subjective judgment is influenced by the style of the drawing presented. Thin tubes of a single color were contrasted with tubes colored by secondary structure. The tube radius was also varied as a function of secondary structure, affecting a ‘tube-and-cylinder’-style diagram.

Aldehyde reductase [32] is shown in Fig. 6. The chain has 320 residues, implying $j = 8.32$. The B-spline curve defined by 64 segments ($j = 6$) goes through the helices and hugs the central barrel. Lowering the level further caused degradation of the image.

About half a dozen co-workers examined the protein of their choice. One was disconcerted by the disappearance of the helices and thought reduction to half the number of points was the limit. Others felt reduction by $1/8$ or greater to be possible, especially if there is additional color/textured information (e.g., fat helices).

The deviation between the C^α and the C^r was monitored for database analysis on the low-resolution curves for the two types of ribbon described in Fig. 2. Except at $j = 0$, where the C^α ribbon is by construction C^r , the plots

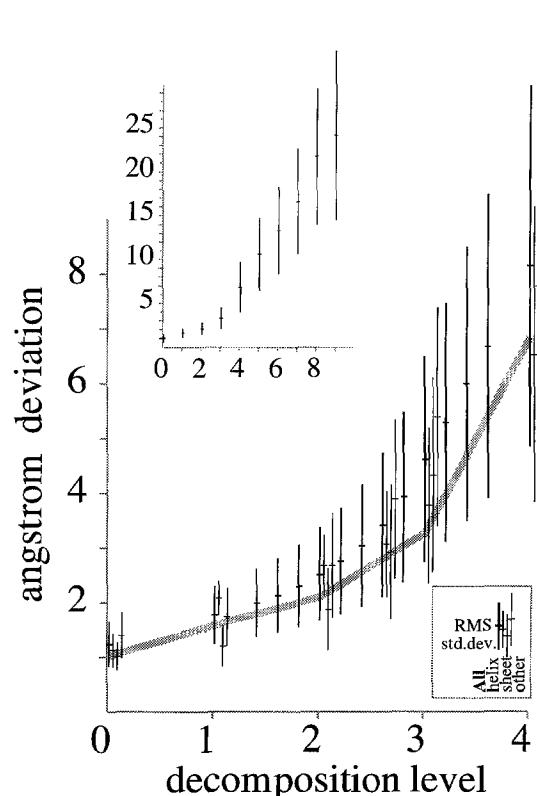


Fig. 7. Deviation from C^α as a function of resolution. The y-axis is the distance between C^α and C^r ; the x-axis is the decomposition level; 0 is the original (padded) curve, at maximum j . For integer level plots, 1, 2, 3,... are the curves with $1/2$, $1/4$, $1/8$,... of the segments. The inset at the upper left gives the results, as does the thick gray line in the main graph. For the fractional level plots, 0 is also the original curve, but 1, 2,... represent the initial fractional level minus 1, 2,... (for example, for neuraminidase 0, 1, 2,... is $j = 8.6$, 7.6 , 6.6 ,...) The rms value over all residues in the database is indicated with a tick mark at the center of error bars representing one standard deviation. Separate marks are shown for all residues and for the separate secondary structure classes: helix, sheet and other.

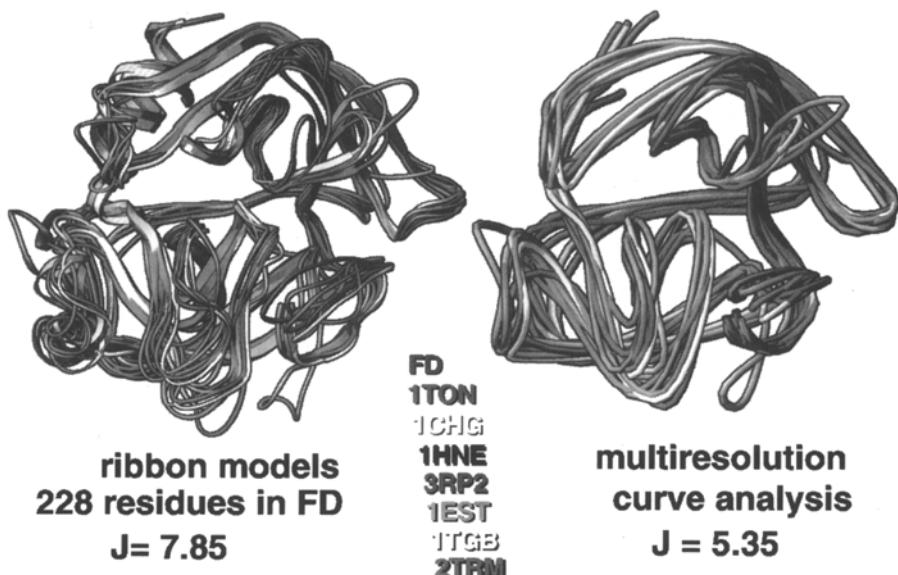


Fig. 8. Topological comparison with MRC representation. Left: standard ribbon drawings of seven serine proteases aligned to model factor D, labeled and colored by the PDB code. Factor D has 228 residues ($J = 7.85$). Right: MRC representations with $J = 5.35$.

are virtually identical. This shows an equivalence of the various ribbon formulations at lower resolution. The data for the standard ribbon drawing is plotted in Fig. 7. It is seen that helix and sheet residues are approximated slight-

ly better than those in coils. The plots are similar, with slightly higher values for the fractional mode. The main difference is the point at which the deviation trends sharply upward.

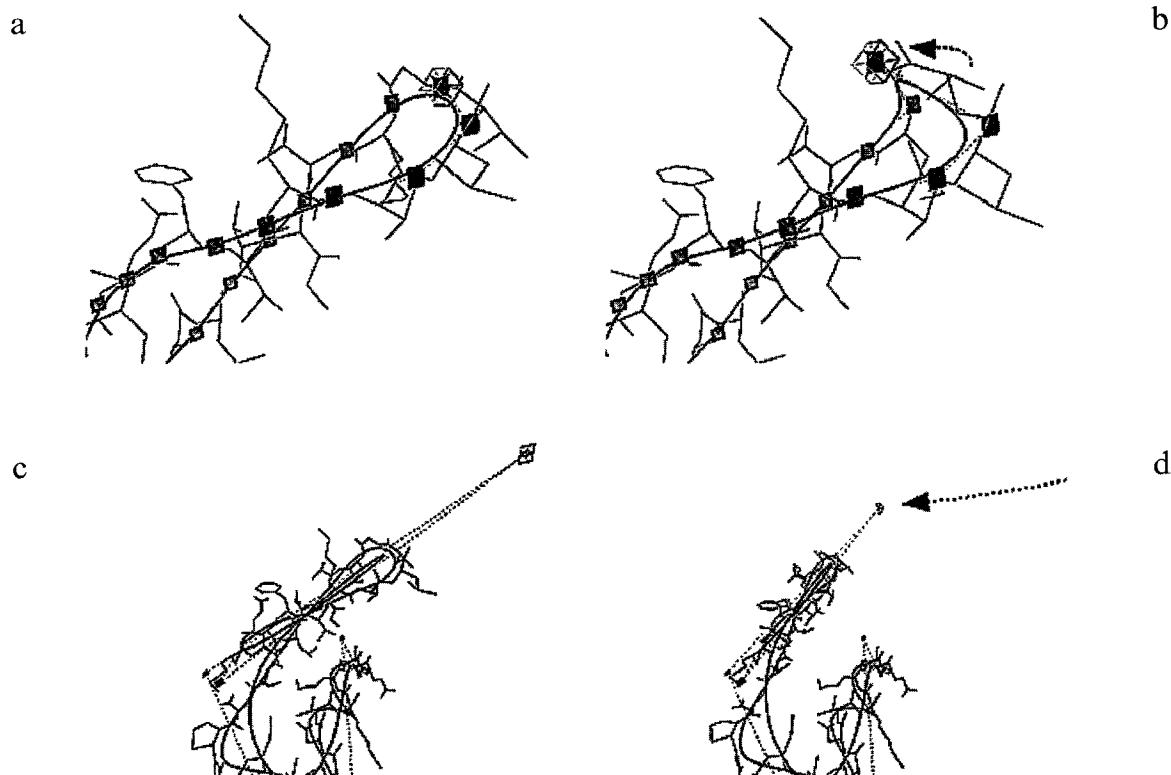


Fig. 9. Multiresolution editing. Thirty-two residues of ubiquitin ($j = 5$) were manipulated. Spheres represent the pickable control points. Upon picking, a 'handle-box' cube highlights the movable point. The thick curve is the high-resolution ribbon, the thinner curve the low-resolution version, and thin lines represent the atomic bonds. The dashed line/arrow sketches the motion of an edited control point in the panels on the right. The panels on the left show the picked point before editing. (a) picking on the original curve; (b) only four local residues are altered; (c) picking on the 1/4 resolution curve; (d) the entire 16-residue strand-loop-strand swings away from the helix.

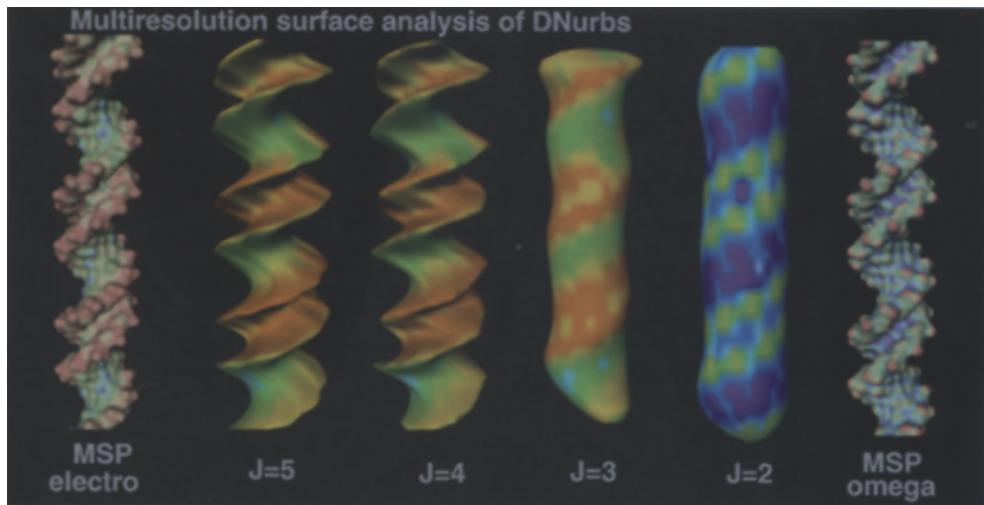


Fig. 10. Multiresolution surfaces of DNA. From left to right: MSP surface colored by electrostatics; *DNurbs* surface for $J = 5, 4, 3$ textured by electrostatics, for $J = 2$ textured by curvature; MSP surface colored by curvature.

Topological comparison

A visual comparison of proteins is shown in Fig. 8. The fractional level of eight superimposed serine proteases [33] was adjusted to the lowest value at which the fold was still deemed clearly recognizable. This value is 2.5 less than the maximum value, implying about 1/6 of the original information.

Human complement factor D is a serine protease with several flexible loops. The crystal structure [34] has two independent conformations, designated the A and B chains. Superposing the C^α atoms for all 228 residues gives an rms difference between C^α positions of 0.95 Å. For the 206 residues not in the flexible loops, the C^α atoms are fit to an rms difference of 0.35 Å. This latter transformation was taken as the best fit.

Least-squares fits were determined based on a reduced number of points. The results of the transformation were assessed by measuring the rms difference of the 206 non-loop residues described above. Every second, fourth, eighth, and sixteenth C^α position was used, starting on every possible initial point. For example, using every fourth point, fitting 1,5,9,..., 2,6,..., 3,7,..., and 4,8,..., gave rms fits of 0.874, 0.727, 0.917 and 1.176 Å, respectively.

Using the MRC representation and the maximum integer level $j - 2$, or 1/4 of the control points, and fitting based on the C^α points gave corresponding values of 0.633, 0.674, 0.715 and 0.674 Å, respectively. These latter results are slightly better and, more importantly, show less variation. However, little difference is seen between the C^α and MRC methods from visual assessment of the superpositions. Also, fits based on the control points of the MRCs, instead of the curve itself, gave poor results.

Multiresolution editing

The user is presented with pickable control point

spheres that define the B-spline ribbon. Dragging the picked point with the mouse changes the ribbon curve and demonstrates the local control property of B-splines: only the two curve segments (residues) on either side of the control point change.

Wavelets allow one to interactively choose the level of approximation. Each decrease in the multiresolution level doubles the number of residues that change when a control point is moved. Figure 9 is a representation. Picking a point in the center of a turn at the maximum level moves only the four residues in the turn. Changing the level to $j = 3$ allowed a point in the low-resolution version to be selected. Moving this point changed all the residues in a strand-loop-strand motif.

A model of calmodulin was presented as an editable MRC. Also displayed was a model of calcium-free calmodulin, with a large conformational change in two helices of the calcium-binding domains [35]. An unsuccessful attempt was made to interconvert the structures by moving only one control point of a very low resolution curve. By raising the level and moving only two control points, the helix-loop-helix backbones could nearly be overlaid, at the expense of considerable distortion of one of the helices.

Molecular surfaces

DNurbs representations were created with 32×32 surface patches modeling a 32-base-pair model of DNA. MSP triangle surfaces are shown for comparison in Fig. 10. The grooves disappeared by level $j = 3$ (8×8 patches), although the texturing makes the groove pattern clear.

The globular protein neuraminidase is shown as a space-filling model in Fig. 11. The textured globe NURBS representations are shown at various resolutions. The method does a decent job at modeling the surface, but the active site progressively loses definition.

For general surfaces, the progressive remeshing of the neuraminidase active site was taken as a test case. The MSP surface is approximated by 6, 24, 96,... triangles in Fig. 12. Note that this is just a remeshing, i.e., a precursor to actual wavelet analysis [27].

Discussion

MRC analysis of the protein fold

The wavelet approximation of the protein backbone produces interesting visuals. A clean view of the packing and interdigititation of the helices in interferon is generated with 1/8 of the geometric information. The lower-resolution 'ribbon' closely matches cylinders fit to the helices (Fig. 5). The visual experiment is complicated by subjectivity and the variety of graphic representations. The consensus of half a dozen colleagues was that 1/4 of the information suggested the same fold.

The distance difference versus resolution plot of Fig. 7 may suggest a best value. In terms of integer levels, the rms deviation is 2.1 Å at 1/4 of the segments and 3.3 Å at 1/8 of the segments. The rms is less than the ideal C^α-C^α distance (3.8 Å) at fractional levels between 2.0 and 3.0. At 2.6 (1/6 of the segments), the rms for helix/sheet residues is less than 3 Å.

At what reduced level of detail is a protein fold recognizable? Can this be expressed as a unique small integer fraction, say 1/4 or 1/6? This number should be optimal for lower-resolved comparisons. It should also have bearing on the information content of the protein fold. An investigation of the number of possible folds has been published recently [36], including a procedure to approximate the protein fold using the discrete cosine transform. Lower-resolution curves resembling wavelet curves suggest that 1/3 of the points always generate a recognizable fold.

Do each three, four, or six residues establish the local

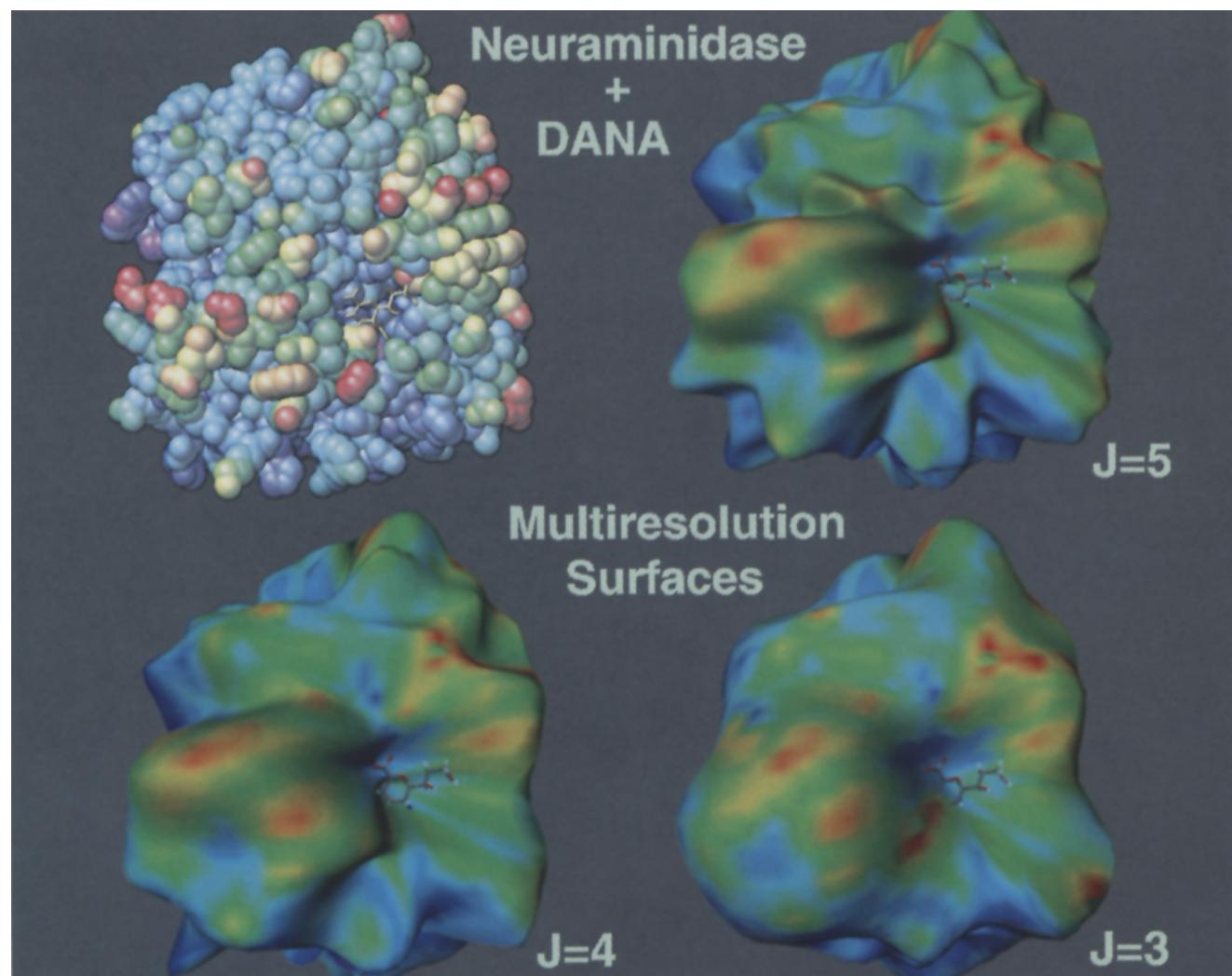


Fig. 11. Multiresolution surfaces of neuraminidase. The inhibitor DANA (2-deoxy-2,3-dehydro-N-acetyl neuraminic acid) is shown as a ball-and-stick model. From upper left: atomic spheres colored by temperature factor; NURBS globe for J = 5, 4, 3 textured by temperature factor.

folding path in space? Einstein said everything should be made as simple as possible, but no simpler. Three nucleotides translate to one amino acid. Three amino acids could code a 3D coordinate. Speculation is based on recent work in theoretical conchology [37]. The serpentine meandering of certain shells may be described by a geodesic (straight line) in a generalized 6D isoEuclidean space. It was noted that this seashell mathematics may be applied to proteins (their pictures might pass for the low-resolution ribbon tubes shown herein).

Topological comparison

The usual technique of substructure detection is a distance-geometry approach, followed by the computationally more expensive 3D superposition [15,38]. Sophisticated alignment techniques [39] assign multiple properties (e.g., hydrophobicity) to each C^α position to aid in the superposition.

Improved methods would be welcome due to the exponential increase in the number of structures being deposited in the Protein Data Bank. How can a structural motif be recognized, and will a lower-resolution model be useful? Searches conducted using the standard techniques will be compared to searches using the wavelet approximation. The wavelet decomposition of the hydrophobicity or other properties should provide a ‘texturing’, useful both for alignment and for interesting graphics. This research is in progress.

Multiresolution editing

Currently no structural constraints are imposed between the residues, each of which moves as an independent rigid body tied to the space curve. It is easy to convert the spline curve from right-handed helix to left- by editing a very low resolution version, as none of the well-known constraints on protein structure are used. A geometric minimizer [40] could easily be added in real time. More work is needed to handle nonbonded interactions. The reconstruction of the underlying atoms from an edited spline is a question for research.

Considerable trial-and-error experimentation will be required to discover the best interactive and numerical techniques. Prototyping is being done in the SGI Inventor programming environment. The ‘virtual reality’ (direct manipulation) interface of the environment has great potential as a general tool to construct and edit proteins through interactive computer graphics.

Molecular surfaces and volumes

The multiresolution NURBS surface visuals look much like the molecular surfaces generated by Max [25] with spherical harmonics. A procedure to texture map spherical harmonic surfaces has been published recently [41]. Methods for more general topologies need to be explored.

Wavelet techniques can be applied to the three-dimensional grid of an electron-density map or various potential maps used by structure-based design. The wavelet-

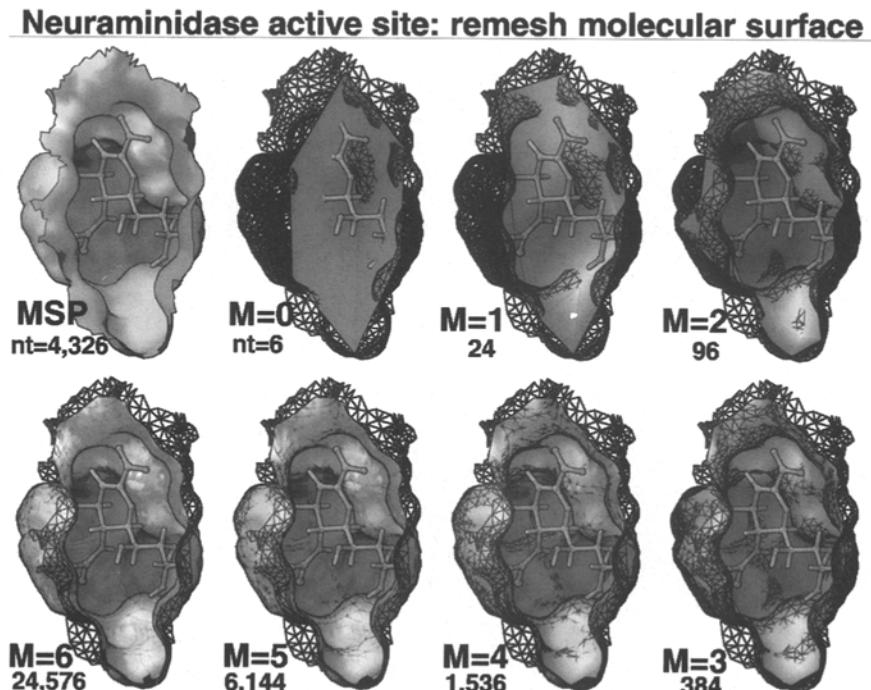


Fig. 12. Multiresolution surface remeshing. The MSP triangular surface of the neuraminidase active site colored by electrostatics along with the inhibitor DANA (2-deoxy-2,3-dehydro-*N*-acetyl neuraminic acid) is shown at the upper left. In each successive panel moving clockwise, the original MSP is shown as a mesh along with the approximating surface. The lowest-resolution level, $M = 0$, has only six triangles. Each successive level of M has four times the number of triangles.

approximation techniques could then be combined with docking methods; either the brute-force surface comparison [42] or searching of an energy grid [43].

Comparison of wavelet transforms

Molecular shapes can be compared by generating molecular envelope volumes, superposing the envelopes, computing their Fourier transforms, and comparing the transforms at various resolutions [44]. This process was used to automatically classify antigen-binding sites.

How can wavelet transforms be compared? Are there any advantages over Fourier techniques [45]? Wavelet analysis should allow much faster comparisons at low resolution. The localization properties of the wavelet transform show promise [46].

Conclusions

Wavelet decomposition of protein backbones and molecular surfaces has been demonstrated and the hierarchy of resolution levels visualized. The multiresolution-editing capability introduces a new interactive tool for protein modeling. The minimum specification of the protein fold is of philosophic interest.

Comparisons of lower-resolution versions of backbones, molecular surfaces, and density volumes should be possible through their wavelet transforms, yielding substantial computational speedup for database and brute-force searches. Wavelet analysis may allow much faster comparisons at low resolution, but at what cost? Important detail may be lost. This is the focus of future research.

Acknowledgements

Thanks to Lily Yang and all, past and present, at the UAB CMC. Thanks to Maxine Rice for preparing and to Susan Baum for editing the manuscript.

References

- 1 Finkelstein, A. and Salesin, D.H., SIGGRAPH Proc., 28 (1994) 261.
- 2 Shoichet, B.K. and Kuntz, I.D., Protein Eng., 6 (1993) 723.
- 3 Daubechies, I., Commun. Pure Appl. Math., 1 (1988) 909.
- 4 Mallat, S., IEEE Trans. Pattern Anal. Machine Intell., 11 (1989) 674.
- 5 Press, W.H., Teukolsky, S.A., Vetterling, W.T. and Flannery, B.P., *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, New York, NY, 1992.
- 6 Stollnitz, E.J., DeRose, T.D. and Salesin, D.H., IEEE Comput. Graph. Appl., 15 (1995) 76.
- 7 Stollnitz, E.J., DeRose, T.D. and Salesin, D.H., IEEE Comput. Graph. Appl., 15 (1995) 75.
- 8 Carson, M. and Bugg, C.E., J. Mol. Graph., 4 (1986) 121.
- 9 Carson, M., J. Mol. Graph., 5 (1987) 103.
- 10 Carson, M., J. Appl. Crystallogr., 24 (1991) 958.
- 11 Richardson, J.S., Adv. Protein Chem., 34 (1981) 167.
- 12 Hobohm, U., Scharf, M., Schneider, R. and Sander, C., Protein Sci., 1 (1992) 409.
- 13 Rao, S.T. and Rossmann, M.G., J. Mol. Biol., 76 (1973) 241.
- 14 Greer, J., Proteins, 7 (1990) 317.
- 15 Jones, T.A. and Thirup, S., EMBO J., 5 (1986) 819.
- 16 Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M., Acta Crystallogr., A47 (1991) 110.
- 17 McRee, D.E., *Practical Protein Crystallography*, Academic Press, San Diego, CA, 1993.
- 18 Surles, M.C., Richardson, J.S., Richardson, D.C. and Brooks Jr., F.P., Protein Sci., 3 (1994) 198.
- 19 Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
- 20 Greer, J. and Bush, B.L., Proc. Natl. Acad. Sci. USA, 75 (1978) 303.
- 21 Connolly, M.L., Science, 221 (1983) 709.
- 22 Klein, T.E., Huang, C.C., Pettersen, E.F., Cough, G.S., Ferrin, T.E. and Langridge, R., J. Mol. Graph., 8 (1990) 16.
- 23 Colloc'h, N. and Mornon, J.-P., J. Mol. Graph., 8 (1990) 133.
- 24 Carson, M. and Yang, Z., J. Mol. Graph., 12 (1994) 116.
- 25 Max, N.L. and Getzoff, E.D., IEEE Comput. Graph. Appl., 8 (1988) 42.
- 26 Connolly, M.L., J. Mol. Graph., 11 (1993) 139.
- 27 Eck, M., Duchamp, T., Lounsbury, M., DeRose, T., Hoppe, H. and Stuetzle, W., SIGGRAPH Proc., 29 (1995) 173.
- 28 Muraki, S., IEEE Comput. Graph. Appl., 13 (1993) 50.
- 29 Babu, Y.S., Sack, J.S., Greenhough, T.J., Bugg, C.E., Means, A.R. and Cook, W.J., Nature, 315 (1985) 37.
- 30 Bossart-Whitaker, P., Carson, M., Babu, Y.S., Smith, C.D., Laver, W.G. and Air, G.M., J. Mol. Biol., 232 (1993) 1069.
- 31 Ealick, S.E., Cook, W.J., Vijay-Kumar, S., Carson, M., Nagabhushan, T.L., Trotta, P.P. and Bugg, C.E., Science, 252 (1991) 698.
- 32 El-Kabbani, O., Green, N.C., Lin, G., Carson, M., Narayana, S.V.L., Moore, K.M., Flynn, R.G. and DeLucas, L.J., Acta Crystallogr., D50 (1994) 859.
- 33 Carson, M., Bugg, C.E., DeLucas, L.J. and Narayana, S.V.L., Acta Crystallogr., D50 (1994) 889.
- 34 Narayana, S.V.L., Carson, M., El-Kabbani, O., Kilpatrick, J.M., Moore, D., Chen, X., Bugg, C.E., Volanakis, J.E. and DeLucas, L.J., J. Mol. Biol., 235 (1994) 695.
- 35 Stynadka, N.C.J. and James, M.N.G., Proteins Struct. Funct. Genet., 3 (1988) 1.
- 36 Crippen, G.M. and Maiorov, V.N., J. Mol. Biol., 252 (1995) 144.
- 37 Illert, C. and Pickover, C., IEEE Comput. Graph. Appl., 15 (1995) 89.
- 38 Vriend, G. and Sander, C., Proteins Struct. Funct. Genet., 11 (1991) 52.
- 39 Sali, A. and Blundell, T.L., J. Mol. Biol., 212 (1990) 403.
- 40 Hermans, J. and McQueen, J.E., Acta Crystallogr., A30 (1974) 730.
- 41 Duncan, B.S. and Olson, A.J., J. Mol. Graph., 13 (1995) 258.
- 42 Jiang, F. and Kim, S.-H., J. Mol. Biol., 219 (1991) 79.
- 43 Goodsell, D.S. and Olson, A.J., Proteins, 8 (1990) 195.
- 44 Gerstein, M., Acta Crystallogr., A48 (1992) 271.
- 45 Leicester, S.E., Finney, J.L. and Bywater, R.P., J. Mol. Graph., 6 (1988) 104.
- 46 Gross, M.H., Gatti, R. and Staadt, O., *IEEE Visualization Proceedings*, Institute of Electrical and Electronics Engineers Inc., New York, NY, 1996, pp. 135–142.
- 47 Vijay-Kumar, S., Bugg, C.E., Wilkinson, K.D. and Cook, W.J., Proc. Natl. Acad. Sci. USA, 82 (1985) 3582.