

Rapid atomic density methods for molecular shape characterization

Julie C. Mitchell,* Rex Kerr,† and Lynn F. Ten Eyck*

*San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA

†Department of Biology, University of California, San Diego, 9500 Gilman Dr., La Jolla, CA

Two methods for rapid characterization of molecular shape are presented. Both techniques are based on the density of atoms near the molecular surface. The Fast Atomic Density Evaluation (FADE) algorithm uses fast Fourier transforms to quickly estimate densities. The Pairwise Atomic Density Reverse Engineering (PADRE) method derives modified density measures from the relationship between atomic density and total potentials. While many shape-characterization techniques define shape relative to a surface, the descriptors returned by FADE and PADRE can measure local geometry from points within the three-dimensional space surrounding a molecule. The methods can be used to find crevices and protrusions near the surface of a molecule and to test shape complementarity at the interface between docking molecules. © 2001 by Elsevier Science Inc.

Keywords: molecular shape, molecular surface, molecular recognition, shape complementarity, protein cavities, protein binding site, atomic density, fast Fourier transform

INTRODUCTION

Molecular shape is one of many factors that determine how a protein interacts with its surroundings. The molecular surface describes shape at an atomic level but does not explicitly provide larger-scale structural information. Although humans are adept at recognizing and comparing large-scale features of molecular surfaces, automated methods for analyzing protein shape are desirable. Analytical characterization of shape at a variety of scales is therefore of interest. Moreover, rapid methods have the advantage of facilitating interactive testing of hypotheses and analysis of large datasets.

A variety of methods have been developed to describe molecular shape. These techniques can be used to highlight

Color Plates for this article are on pages 388–390.

Corresponding author: Dr. J.C. Mitchell, San Diego Supercomputer Center, Univ. of California, San Diego, 9500 Gilman Dr., La Jolla, CA 92093-0527, USA

E-mail address: mitchell@sdsc.edu

features of interest such as pockets, grooves, and protrusions. Small-scale features of a molecule are evident in the molecular and solvent-accessible surfaces.^{1–4} Although the curvature of these surfaces is locally constant, surface-based solid angle methods can elucidate shape on a coarser level.^{5–7} Spherical harmonic approximation and Fourier analysis are likewise able to derive coarser measures of molecular shape and curvature,^{8–10} and convolution of electron density with derivatives of Gaussians may be used to derive curvature measures on a variety of scales.¹¹ Electron density libraries for molecular fragments¹² can be used to approximate density and shape for model systems. In addition to analytic methods for shape characterization, Voronoi polyhedra^{13,14} and alpha shapes^{15–17} have been used to find surface pockets and inaccessible cavities. Another class of shape analysis methods perform direct comparison and global matching of molecules, and we point the reader toward more comprehensive volumes on molecular shape^{18,19} for an overview.

Information about local shape can be derived from the density of atoms near a point in space. Atomic density has been proposed as a measure of intermediate-scale molecular shape and been used to detect water-binding grooves in proteins.²⁰ However, existing methods for the computation of atomic density are time-consuming. We present two methods for rapid density computations. The first uses fast convolution methods to approximate atomic density. The second exploits the relationship between total potential at a point and the density of atoms contributing to the potential. Lennard-Jones potentials are used to define a smoothly-varying volumetric quantity that measures the local shape of a molecule from points outside the molecular surface.

We first review atomic density measures and discuss how they have been implemented using fast Fourier transforms in the Fast Atomic Density Evaluation (FADE) program. Density measures computed with FADE are used to find crevices near the surface of trypsin, and simple functions for scoring shape complementarity are used to analyze candidate complexes for docking fasciculin with acetylcholinesterase and for docking a pair of hemoglobin $\alpha-\beta$ dimers. We follow this with an outline of how the Pairwise Atomic Density Reverse Engineering (PADRE) method uses total potential values to produce

surface-independent measures of molecular shape. Three-dimensional density measures are given for bovine pancreatic trypsin inhibitor, and the density dimensions computed using FADE and PADRE are compared for Uracil DNA Glycosylase.

RAPID APPROXIMATION OF ATOMIC DENSITY

Atomic Density as a Measure of Shape

Consider a point $x = (x_1, x_2, x_3)$ near the surface of a molecule. If x lies near a protrusion, there will be relatively few atoms nearby, and the number of atoms will rise slowly as the size of the neighborhood increases. Conversely, one expects to find many atoms in the neighborhood of a crevice (Color Plate 1). This intuitive analysis can be quantified using atomic density measures.²⁰

The radial counting function, $N(x, r)$, counts the number of atoms that lie inside a ball of radius r centered about the point x . That is, if M is a collection of atomic centers,

$$N(x, r) = |\{a \in M : |x - a| \leq r\}|$$

gives the number of atoms within distance r of the point x .

The “fractal” atomic density dimension²⁰ $\lambda = \lambda(x)$ has been defined as the slope of the line determined by least squares fit of $\log(N)$ versus $\log(r)$ over the range $0 < r < 10 \text{ \AA}$. This gives an approximation

$$N(x, r) \approx r^\lambda \quad (1)$$

The value of λ determines the rate of increase in the number of neighboring atoms as the size of the neighborhood increases. This value will be higher in a crevice than near a protrusion.

FAST ATOMIC DENSITY EVALUATION (FADE) METHODOLOGY

The problem of computing $N(x, r)$ can be recast in terms of convolution integrals. In particular,

$$N(x, r) = \int_{R^3} B_r(x - y) \cdot \sigma(y) dy$$

where σ is the sum of delta functions taken at the atomic centers and B_r is a function that equals one inside a ball of radius r and zero elsewhere. Note that the convolution is with respect to the variable $x \in R^3$ and assumes r is fixed.

For a fixed value of r , the value of $N(x, r)$ is determined for all $x \in R^3$ by a convolution integral. The Convolution Theorem for Fourier transforms indicates that $\hat{N} = \hat{B}_r \cdot \hat{\sigma}$. That is, the Fourier transform of a convolution product is the scalar product of Fourier transforms.

The fast Fourier transform (FFT) can be used to speed the computation of $N(x, r)$ considerably. This type of scheme has previously been applied to the computation of potential functions and energies.^{21,22} Use of FFTs in FADE produces similar results to the original methods²⁰ and reduces computation time from hours to seconds. Implementation of the rapid technique is straightforward. To discretize σ , atoms in the molecule are rounded onto grid points, and a “discrete ball” B_r occupies all grid points within a fixed distance r of the origin. Convolution of the occupancy grid σ with B_r is performed for radii $r_k = 1, 2, \dots, 10$ to produce grids $N(x, r_k)$. For a grid point x , the radial counting function is approximated by the sequence $\{N(x, r_1), N(x, r_2), \dots, N(x, r_{10})\}$, and a log-log fit is used to determine the density dimension.

RESULTS

Atomic density has previously been used to locate grooves and crevices along the surface of a molecule using the program Surfractal.²⁰ Our FADE program quickly locates pockets and gives a global characterization of the molecular geometry. Color Plate 2 illustrates the effect of plotting points of high density dimension along the surface of trypsin. Points inside small crevices have been identified, and the resulting object has only subtle, large-scale concavities. Coordinates for trypsin were obtained from the file 1SGT.pdb.

To test the utility of atomic density methods for scoring shape complementarity in docked complexes, we have devised a simple method for using FADE to analyze docked configurations returned by the molecular docking program DOT.²¹ DOT uses fast Fourier transforms to compute potential energy values for rigid body docking in a way that is similar to FADE’s use of FFTs to compute the radial counting function. DOT returns a list of favorable docked configurations, which typically includes many answers close to the crystal structure along with some “false positive” results. To score the docked complexes returned by DOT, FADE was used to produce density dimension values at all points between 0 and 3 Å of the “stationary” molecule. For each orientation of the “moving” molecule, the same computation was performed. Grid points found to be in both, that is, all points that were within 3 Å of both molecules, were saved. For each interface point x_i the value of

$$C_i = (s_i - \lambda_0) \cdot (m_i - \lambda_0) \quad (2)$$

gives a simple measure of shape complementarity near this point. Here, s_i denotes the density dimension relative to the stationary molecule at the i th interface point, and m_i is the density dimension relative to the moving molecule. The value of λ_0 is chosen in the range [2.8, 3.0] and gives the value of the density dimension for which local geometry is expected to be flat. In theory, one would expect a value of $\lambda_0 = 3.0$ for three-dimensional objects. In practice, this value tends to be slightly less for atomic distributions found in proteins.²⁰ Our test results use the value $\lambda_0 = 2.9$. The value of C_i is positive in a crevice–crevice or protrusion–protrusion docking and is negative in cases where a protrusion has been docked into a crevice. Values near zero suggest the docking of flat interfaces. The following formula was used to score the total shape complementarity

$$C_{abs} = \sum_{i=1}^n (s_i - \lambda_0) \cdot (m_i - \lambda_0) \quad (3)$$

We also examined the average level shape complementarity using the formula $C_{avg} = C_{abs}/n$, where n denotes the number of interface points.

Color Plate 3 shows acetylcholinesterase and fasciculin as docked in the crystal structure 1MAH.pdb. While this interaction is driven primarily by electrostatics, there is a great deal of surface contact between the molecules. Grid points in the interface have been colored according to the value of the

complementarity measure C_i , with negative values colored red. Places where a crevice has been docked against a protrusion are highlighted by orange and red spheres.

Figure 1 shows the C_{abs} complementarity scores for docking fasciculin with acetylcholinesterase, and we also examined the docking two (deoxy) hemoglobin $\alpha-\beta$ dimers to form the tetramer using coordinates from the file 1BZO.pdb. In each case, we used the top 20 solutions returned by DOT as well as the known crystal structures. For both systems, the crystal structure produced the lowest values of C_{abs} and C_{avg} . The C_{avg} scores (not shown) have a similar distribution to those shown for C_{abs} . For candidate fasciculin-acetylcholinesterase complexes, there is a clear correlation between complementarity scores and RMSD. For hemoglobin, all the lowest scoring structures had small RMSD values and the crystal structure produced by far the best score. Note that some low RMSD scores produced relatively high complementarity scores, which is likely due to mismatches between the protrusions and crevices. While very preliminary, these results suggest that FADE complementarity scores will prove valuable in determining correctly-docked complexes. A more detailed investigation is underway.

ATOMIC DENSITY FROM TOTAL POTENTIALS

Pairwise Atomic Density Reverse Engineering (PADRE) Methodology

We now examine the relationship between the radial counting function $N(x, r)$ and total potentials given as the sum of pairwise interactions. Total potential values depend on the distribution of atoms within three-dimensional space and thus contain information on molecular shape. Let $P(r)$ be a pairwise potential dependent only on the distance between points. For a point x outside of the molecule, the total potential $T(x)$ due to pairwise interactions is

$$T(x) = \sum_{i=1}^m P(r_i)$$

where the sum is taken over all atoms in the molecule and $r_i = |x - a_i|$ is the distance between x and an atomic center a_i .

The radial counting function encodes information on the distribution of atoms based on their distances to a point in space. If x is fixed, then $N(x, r)$ is a step function having jump discontinuities whenever $r = r_i$ for some i , that is, whenever x accumulates a new neighbor. Then $\partial N/\partial r$ is the sum of delta functions taken at these discontinuities, which implies

$$T(x) = \sum_{i=1}^m P(r_i) = \int_0^\infty P(r) \cdot \frac{\partial N}{\partial r}(x, r) dr$$

Using an approximation of the form $N(x, r) \approx r^\lambda$ would lead to integrals that do not converge, due to the singularity at $r = 0$. Therefore, PADRE uses an approximation of the form

$$N(x, r) \approx \begin{cases} 0 & r \leq d \\ (r - d)^\lambda & r > d \end{cases} \quad (4)$$

where d is the distance between the point x and its closest atomic neighbor.

Suppose $\tilde{N}(x, r) = (r - d)^\lambda$ is an atomic density approximation to the radial counting function $N(x, r)$, and consider the approximation

$$T(x) \approx \int_0^\infty P(r) \cdot \frac{\partial \tilde{N}}{\partial r}(x, r) dr$$

While this approximation does not give a high level of numerical accuracy in general, the relationship can be “reverse engineered” to derive modified density dimensions. Given the correct value for the total potential, we will solve this equation for the appropriate value of $\lambda = \lambda(x)$. Note that the total potential $T(x)$ and nearest neighbor distance $d(x)$ can be com-

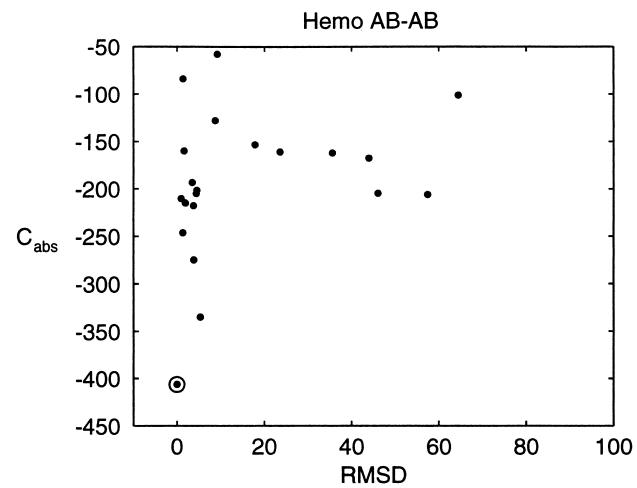
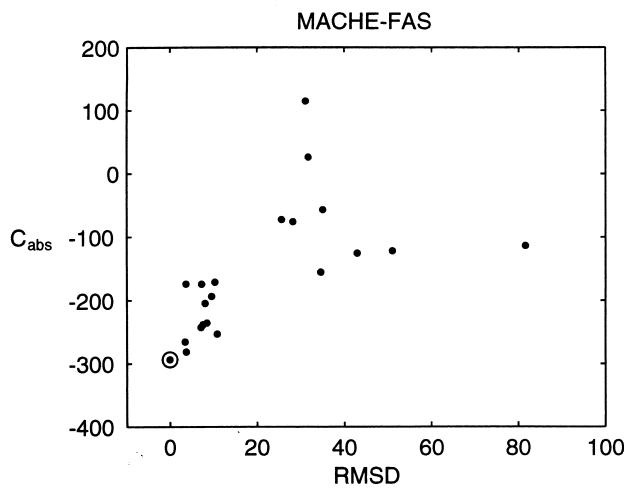


Figure 1. Shape complementarity scores for docking acetylcholinesterase with fasciculin and for docking a pair of (deoxy) hemoglobin $\alpha-\beta$ dimers. Candidate structures returned by the molecular docking program DOT are scored according to their shape complementarity. In each case, the crystal structures give the lowest absolute complementarity scores. Structures that have a low root mean standard deviation from the crystal structures typically return lower values than mislocated complexes.

puted directly from the pairwise definition. Also note that density dimensions computed using this type of method will differ from those discussed in the previous section. The original definition of atomic density is independent of any potential function, while those outlined in this section will depend somewhat on the choice of $P(r)$.

Integration Formulas for r^{-k} Potentials

We will now examine the case when $P(r) = r^{-k}$ for integer $k > 1$. Suppose \tilde{N} is defined as in Equation 4, and consider the integral

$$\int_0^\infty P(r) \cdot \frac{\partial \tilde{N}}{\partial r}(x, r) dr = \int_d^\infty r^{-k} \cdot [\lambda \cdot (r - d)^{\lambda-1}] dr$$

For integer $0 < \lambda < k$, the following formula can be derived

$$\int_d^\infty r^{-k} \cdot [\lambda \cdot (r - d)^{\lambda-1}] dr = \frac{\lambda! \cdot (k-1-\lambda)!}{(k-1)!} \cdot d^{\lambda-k} \quad (5)$$

Although the factorial function is only defined for integers, the Gamma function will allow us to construct more general formulas. The Gamma function is formally defined as

$$\Gamma(z) = \int_0^\infty t^{z-1} \cdot \exp(-t) dt$$

and has the property that $\Gamma(z+1) = z!$ for integer z . Equation 5 extends readily to noninteger λ using the equation

$$\int_d^\infty r^{-k} \cdot [\lambda \cdot (r - d)^{\lambda-1}] dr = c_k(\lambda) \cdot d^{\lambda-k} \quad (6)$$

where

$$c_k(\lambda) = \frac{\Gamma(k-\lambda) \cdot \Gamma(\lambda+1)}{\Gamma(k)}. \quad (7)$$

This expression for $c_k(\lambda)$ is most compact, but not numerically efficient. In particular, the Gamma function is slow to compute, and the given formula involves ratios of large numbers. Fortunately, Equation 7 can be rewritten in a way that requires less computing time. When λ is an integer, well-known combinatorial formulas can be used to compute the value of $c_k(\lambda)$ most efficiently. To find a simplified formula for noninteger λ , we first use the identity

$$\Gamma(z+1) = z \cdot \Gamma(z)$$

By “pulling out terms” one arrives at

$$c_k(\lambda) = \frac{(k-1-\lambda) \cdot (k-2-\lambda) \cdots (1-\lambda) \cdot \lambda}{(k-1) \cdot (k-2) \cdots 1} \cdot \frac{\pi}{\Gamma(1-\lambda) \cdot \Gamma(\lambda)}$$

For noninteger z the formula

$$\Gamma(1-z) = \frac{\pi}{\Gamma(z) \cdot \sin(\pi \cdot z)}$$

reflects the Gamma function about $z = 1$. This allows us to rewrite Equation 7 as

$$c_k(\lambda) = \frac{(k-1-\lambda) \cdot (k-2-\lambda) \cdots (1-\lambda)}{(k-1) \cdot (k-2) \cdots 1} \cdot \frac{\pi \cdot \lambda}{\sin(\pi \cdot \lambda)} \quad (8)$$

This equation can be evaluated much faster than Equation 7. Polynomial approximations to the Gamma function could likewise be used to speed computation, although Equation 8 is exact and nearly as fast to evaluate in practice.

Lennard-Jones Density Dimensions

Lennard-Jones “6–12” potentials have the form

$$P(r) = a \cdot \left(\left(\frac{r_0}{r} \right)^{12} - 2 \cdot \left(\frac{r_0}{r} \right)^6 \right)$$

Since the value of a has no effect on the density dimensions recovered, we will assume $a = 1$. These potentials model van der Waals interactions along with steric repulsion. The results given in the previous section imply that

$$\int_0^\infty P(r) \cdot \frac{\partial \tilde{N}}{\partial r}(x, r) dr = d^\lambda \cdot \left(c_{12}(\lambda) \cdot \left(\frac{r_0}{d} \right)^{12} - 2 \cdot c_6(\lambda) \cdot \left(\frac{r_0}{d} \right)^6 \right)$$

This effectively rewrites the sum of pairwise potentials in terms of a single potential that is similar in form. Our goal is to solve the equation

$$T = d^\lambda \cdot \left(c_{12}(\lambda) \cdot \left(\frac{r_0}{d} \right)^{12} - 2 \cdot c_6(\lambda) \cdot \left(\frac{r_0}{d} \right)^6 \right)$$

for the correct value of $\lambda = \lambda(x)$ given the values of $T = T(x)$ and $d = d(x)$. Because inverting the right-hand side of this equality as a function of λ does not seem possible, a bisection method has been employed to compute the inverse numerically.

RESULTS

Color Plate 4 plots the nearest neighbor distance, Lennard-Jones potential, and atomic density dimension for a simple 2-D molecule containing 10 atoms. A value of $r_0 = 2$ was used in computing the potential, and the surface of constant distance 3 Å from the atomic centers is outlined in white. Crevices, such as the low-energy regions shown in blue, are evident in the potential values, but the effect appears very localized. PADRE effectively rescales the total potential according to distance in a way that elucidates features of interest. While total potential values are rapidly vanishing, density dimensions are less sensitive to variations in the distance between a point and its closest atomic neighbor. This simplifies the problem of locating crevices and protrusions, since molecular features are well-characterized from points near the surface as well as those which are more distant.

Color Plate 5 shows Lennard-Jones density dimensions along a shell 2 Å outside the molecular surface of bovine pancreatic trypsin inhibitor (BPTI), and a slicing plane indi-

cates how dimensions vary beyond this surface. The surface was computed using MSMS⁴ by adding 2 Å to the radius of each atom. The high density “flares” of red highlight regions near a crevice, while low-density blue regions can be seen in the neighborhood of each protrusion. Regions of high and low density dimension extend beyond the protrusions and grooves themselves. Coordinates for BPTI were obtained from the file 5PTI.pdb.

Color Plate 6 compares the density dimensions computed by FADE and PADRE for Uracil DNA Glycosylase (UDG). The shell of points generated by FADE consists of grid positions that are roughly 2–3 Å from an atomic center. Each point has been colored according to its atomic density dimension, with low density regions in blue and high density regions in red. The results are comparable to those generated by PADRE on a surface ~3 Å outside the molecular surface. This set of coordinates was taken from the docked complex of UDG with its inhibitor (1UGH.pdb).

DISCUSSION

FADE allows for extremely rapid characterization of molecular shape. Although FADE requires atoms to be rounded onto grid points, this does not appear to significantly distort larger-scale features highlighted by the atomic density dimension. In a matter of seconds, FADE can produce a shell of points and density dimensions such as that shown in Color Plates 1 or 6. Running time depends primarily on the speed of convolution, which in turn depends on the grid size. A 64^3 grid is sufficient for most molecules of interest, and for a problem of this size, running times of 8–12 s are typical on SGI O2 machines. On a DEC DS20, computations for 64^3 grids terminate in 1–3 s, while larger problems (80^3 – 128^3) require 5–30 s. We are unaware of any faster methods for molecular shape characterization. In addition, we have demonstrated that density dimensions computed by FADE give useful measures of the quality of docked complexes.

Pairwise density dimensions computed using PADRE are smoothly varying with a well-defined gradient except at atom positions and points along the Voronoi diagram. Lennard-Jones density dimensions are “softer” and have a longer range than the potential itself, as may be seen in Color Plate 4. PADRE appears to give good results for both large and small molecules, and we are currently investigating how PADRE might best be used for shape-based optimization of docked complexes. PADRE’s methods are more computationally expensive than those used by FADE, but a detailed analysis can nonetheless be produced in seconds to minutes. The dominant term in the running time is computation of total potential. PADRE can compute density dimensions over a 41^3 grid (68,921 grid points) at a rate of 1 s per 100 atoms on the DEC DS20 and 4 s per 100 atoms on the SGI O2. Computing density dimensions over a surface typically requires less time than is needed to produce a grid.

FADE and PADRE have both similarities and differences with existing shape characterization methods. The major difference between these methods and existing techniques^{5,6,9–11} is that density dimensions are defined independently of any surface and give measures of local shape throughout three-dimensional space. Aside from the original work,²⁰ FADE is most similar to the solid angle method,⁶ which measures shape based on the intersection of the molecule interior with a shell

centered at a point on the molecular surface. The methods used by PADRE most resemble the Gaussian summation technique.¹¹ While the work¹¹ does not explicitly discuss atomic density, the idea that shape influences total potential values is implicit in their methods.

Both FADE and PADRE will be distributed by the Computational Center for Macromolecular Structures (<http://www.sdsc.edu/CCMS/>).

CONCLUSION

Automated analysis of molecular shape is a key step in the visualization, prediction, and characterization of protein–protein interactions and the interactions of a protein with a ligand or solvent molecule. Atomic density provides a volumetric measure of local protein shape, and previous studies have shown the utility of atomic density measures in identifying protein surface features such as water-binding grooves.²⁰ Here, we have introduced two novel methods that combine the idea of atomic density with efficient numerical techniques. Our atomic density methods have been used to locate potential binding pockets and perform detailed shape characterization of single molecules. In addition, FADE complementarity scores were able to distinguish the best structures from a list of candidate docking configurations for selected protein–protein systems.

The fast transform method used in FADE greatly accelerates computation of atomic density in a region surrounding the molecule. FADE is able to identify protrusions and grooves along protein surfaces, and the results of FADE can be used to illustrate shape complementarity in docked complexes. The reverse potential method used in PADRE allows for shape analysis at arbitrary points in space. PADRE’s shape measure avoids discontinuities and retains rapid computation time. Together, the two methods allow rapid identification of coarse surface features as well as shape characterization to arbitrary resolution. Both methods should aid in protein surface visualization and the analysis of shape complementarity in molecular docking.

ACKNOWLEDGMENTS

The authors thank the La Jolla Interfaces in Science Program, Burroughs-Wellcome Fund, the National Science Foundation (NSF DBI 9616115/9911196), and the Department of Energy (DOE/OER DE-FG03-096ER62262) for helping to support our research.

Graphics for this manuscript were created using Rasmol, MATLAB, and the Cortona VRML Client. The surfaces shown in Color Plates 5 and 6 were computed using MSMS⁴.

Finally, we would like to thank Jeff Mandell and Igor Tsigelny, whose DOT results were used to test FADE’s potential for scoring docked complexes.

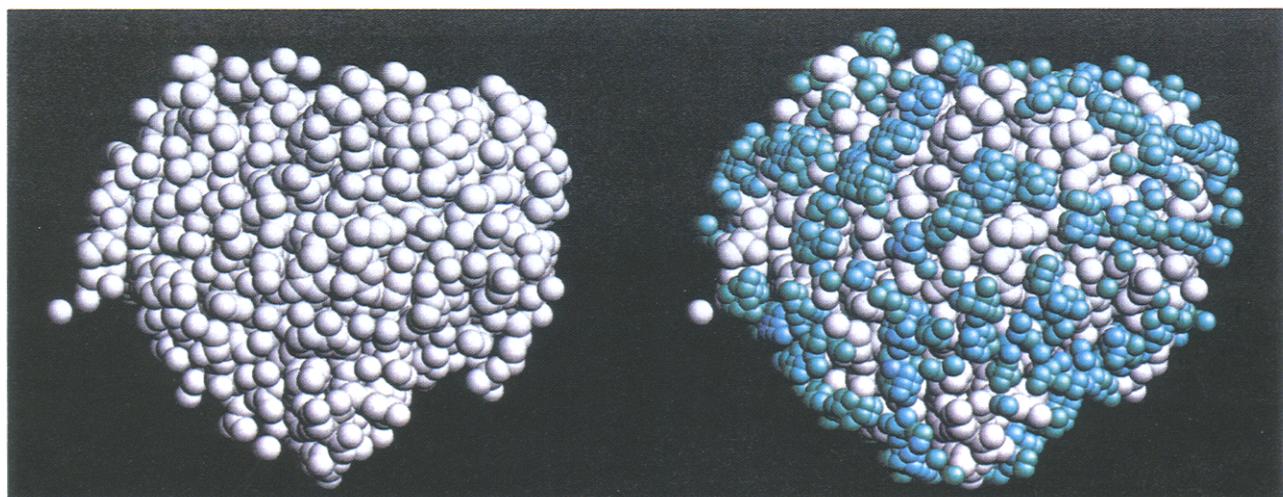
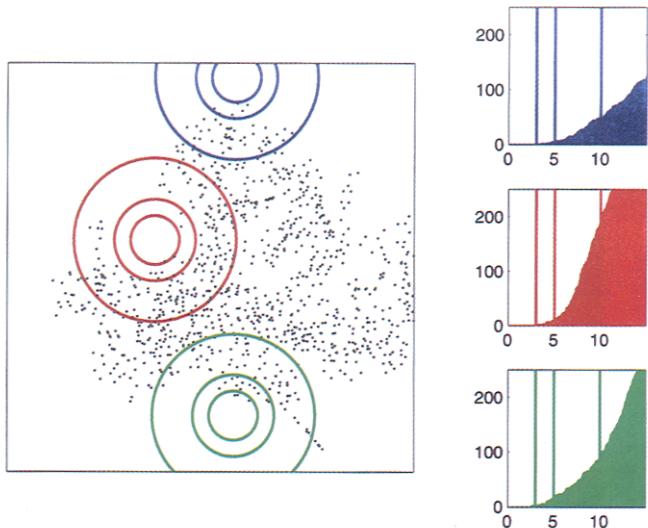
REFERENCES

- 1 Richards, F.M. Areas, volumes, packing and protein structure. *Ann. Rev. Biophys. Bioeng.* 1977, **6**, 151–176
- 2 Lee, B., and Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 1971, **55**, 379–400

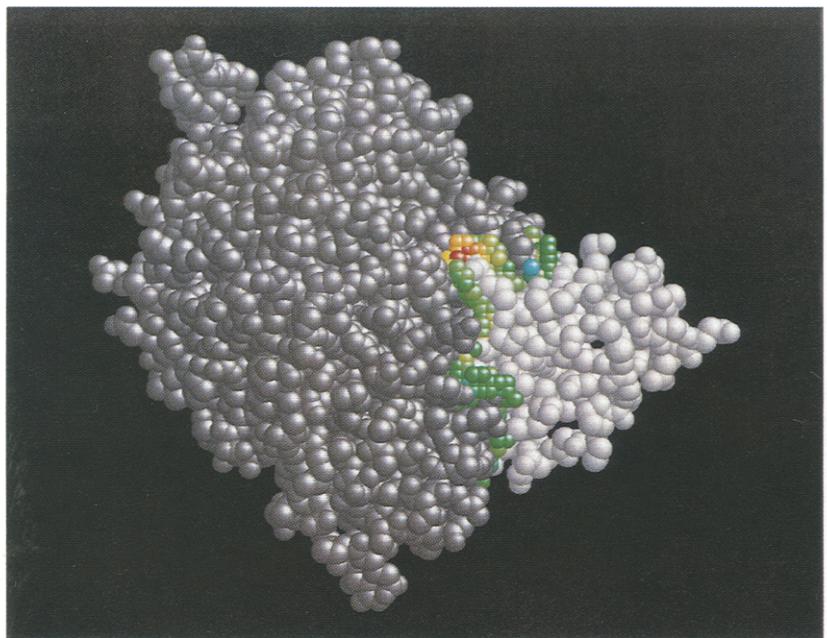
- 3 Connolly, M.L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* 1983, **221**, 709–713
- 4 Sanner, M., Olson, A.J., and Spehner, J.C., Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* 1996, **38**, 305–320
- 5 Nicholls, A., Sharp, K.A., and Honig, B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Gen.* 1991, **11**, 282
- 6 Connolly, M.L. Measurement of protein surface shape by solid angles. *J. Mol. Graphics* 1986, **4**, 3–6
- 7 Norel, R., Lin, S.L., Wolfson, H.J. and Nussinov, R. Shape complementarity at protein–protein interfaces. *Biopolymers* 1994, **34**, 933–940
- 8 Leicester, S.E., Finney, J.L. and Bywater, R.P. Description of molecular surface shape using Fourier descriptors. *J. Mol. Graphics* 1988, **6**, 104–108
- 9 Max, N.L., and Getzoff, E.D. Spherical harmonic representation of molecular surfaces. *IEEE Computer Graphics Appl.* 1988, **8**, 42–50
- 10 Duncan, B., and Olson, A.J. Approximation and characterization of molecular surfaces. *Biopolymers* 1993, **33**, 219–229
- 11 Duncan, B., and Olson, A.J. Shape analysis of molecular surfaces. *Biopolymers* 1993, **33**, 231–238
- 12 Walker, D.W., and Mezey, P.G. Ab initio quality densities for proteins: A MEDLA approach. *J. Am. Chem. Soc.* 1994, **116**, 12022–12032
- 13 David, C.W. Voronoi polyhedra and cleft recognition in aquated macromolecules. *Comput. Chem.* 1984, **8**, 225–226
- 14 Bradley, M., Richardson, W., and Crippen, G.M. Deducing molecular similarity using Voronoi binding sites. *J. Chem. Inform. Comp. Sci.* 1993, **33**, 750–755
- 15 Edelsbrunner, H., and Mücke, E. Three-dimensional al-pha shapes. *ACM Transactions on Graphics* 1994, **13**, 43–72
- 16 Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P.V., and Subramanian, S. Analytical shape computation of macromolecules: I. Molecular area and volume through alpha shape. *Proteins Struct. Funct. Gen.* 1998, **33**, 1–17
- 17 Liang, J., Edelsbrunner, H., Fu, P., Sudhakar, P.V., and Subramanian, S. Analytical shape computation of macromolecules: II. Inaccessible cavities in proteins. *Proteins Struct. Funct. Gen.* 1998, **33**, 18–29
- 18 Mezey, P.G. *Shape in Chemistry: An Introduction to Molecular Shape and Topology*. VCH Publishers Inc., Saskatoon, Canada. 1993
- 19 Arteca, G.A. Molecular shape descriptors. *Rev. Comp. Chem.* 1994, **9**, 191–254
- 20 Kuhn, L.A., Siani, M.A., Pique, M.E., Fisher, C.L., Getzoff, E.D., and Tainer, J.A. The interdependence of protein surface topography and bound water molecules revealed by surface accessibility and fractal density measures. *J. Mol. Biol.* 1992, **228**, 13–22
- 21 Ten Eyck, L.F., Mandell, J., Roberts, V.A., and Pique, M.E. Surveying molecular interactions with DOT. In: *Proceedings of the 1995 ACM/IEEE Supercomputing Conference*, Hayes, A., and Simmons, M., Eds., ACM Press, New York, 1995
- 22 Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A.A., Aflalo, C., and Vakser, I.A. Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. *Proc. Nat. Acad. Sci. USA* 1992, **89**, 2195–2199
- 23 Marchot, P., Prowse, C.N., Kanter, J., Shelley Camp, S., Ackermann, E.J., Radić, Z., Bougis, P.E., and Taylor, P. Expression and activity of mutants of fasciculin, a peptidic acetylcholinesterase inhibitor from mamba venom. *J. Biol. Chem.* 1997, **272**, 3502–3510

Rapid atomic density methods for molecular shape characterization

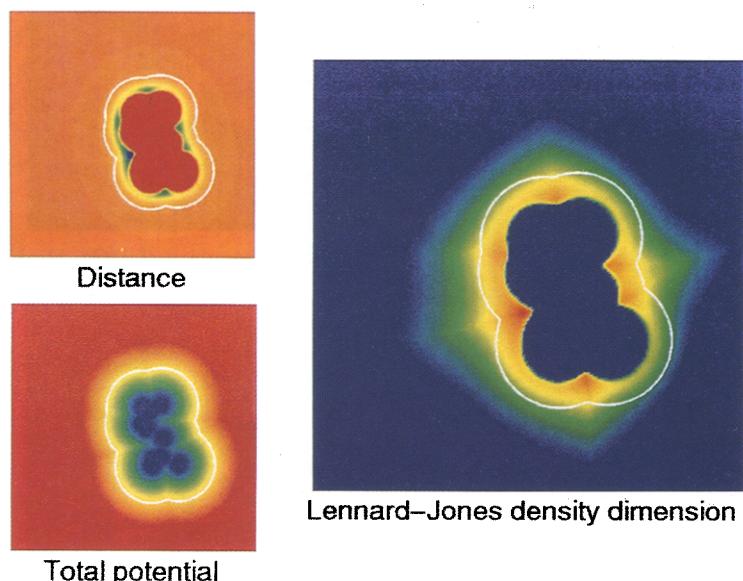
Color Plate 1. The relationship between the radial counting function and shape. The radial counting function has a rate of increase that varies according to the local environment. Points near a protrusion accumulate atomic neighbors slowly (blue), while points inside a crevice are apt to see a rapid increase in the number of atomic neighbors (red). Regions near a flat edge have behavior that lies between these two extremes (green).



Color Plate 2. Identifying grooves in trypsin. A monochrome image of trypsin (left) shows numerous crevices along the protein surface. Grid points with a high atomic density found by FADE effectively fill in crevices, leaving an overall structure that is more convex (right).

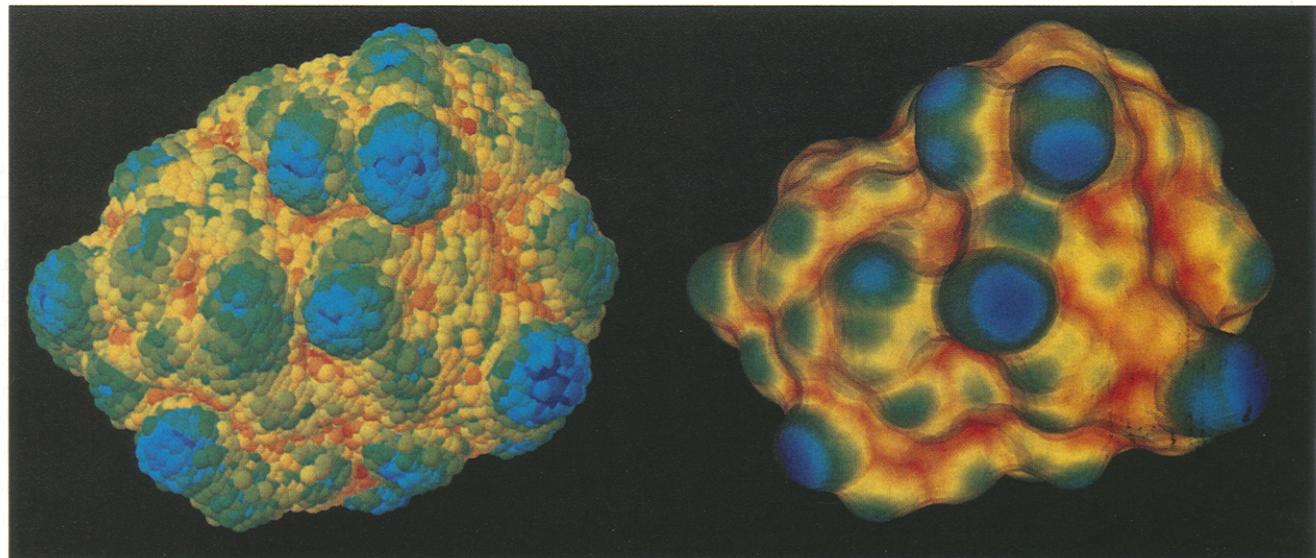
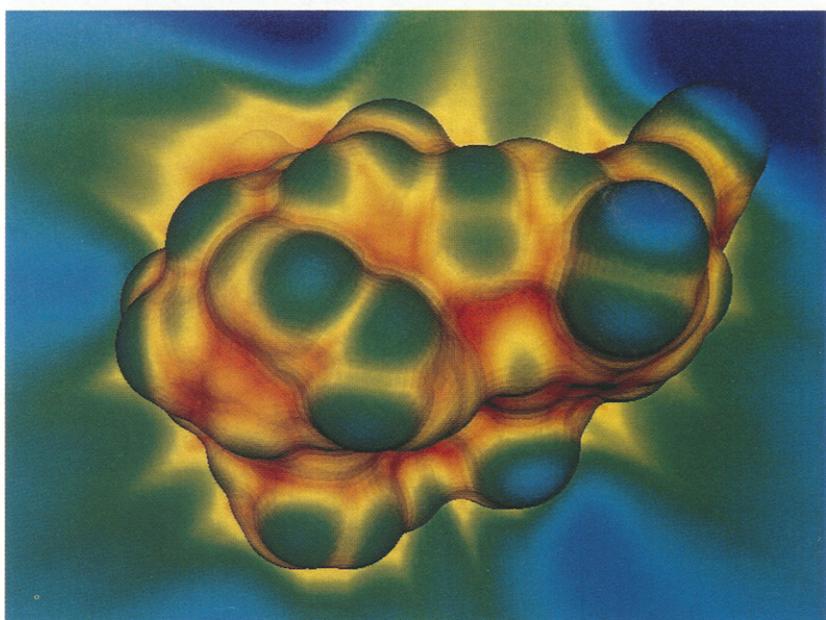


Color Plate 3. Complementarity of density dimensions at the interface of fasciculin (light grey) and acetylcholinesterase (dark grey). Points at the docking interface have been colored according to their individual complementarity scores, with red indicating a negative score. Red spheres denote points along the interface having high density dimension relative to one subunit and low density dimension relative to the other, which suggests the docking of a protrusion into a crevice. Grid points having median atomic density relative to both subunits are shown in yellow or green and indicate the docking of flat interfaces. The highest level of shape complementarity is seen near residues ARG11, ARG27, PRO31, and TYR61 of fasciculin. Mutations to any of the first three of these are known to produce significant changes in the inhibitory activity of fasciculin.²³



Color Plate 4. PADRE provides shape information in an extended region surrounding a molecule. A surface 3 Å from the atomic centers (white outline) provides limited information on molecular shape. The Lennard-Jones potential provides shape information immediately outside the molecule but falls off rapidly with distance. Density dimensions computed with PADRE provide local shape information in a broad region surrounding the molecule. Red indicates regions of high atomic density, while low density regions are shown in blue. Crevices in the 3-Å surface are identified by PADRE as having high density, whereas the entire surface has Lennard-Jones potential values that are very close to zero.

Color Plate 5. Shape characterization using PADRE along a surface and intersecting plane about bovine pancreatic trypsin inhibitor (BPTI). Regions of low atomic density (blue) and high atomic density (red) are derived from Lennard-Jones potential values at each point. The density measure is smoothly varying both over the molecular surface and along the plane intersecting the molecule.



Color Plate 6. Comparison of shape characterization with FADE and PADRE for Uracil DNA Glycosylase (UDG). High density regions in crevices are colored red, while low density regions surrounding protrusions are shown in blue. The shell of points and density dimensions computed by FADE are capable of distinguishing crevices and protrusions (left). Smaller-scale features are visible in the high-resolution surface used by PADRE (right).