ELSEVIER

Analytically defined surfaces to analyze molecular interaction properties

R.R. Gabdoulline and R.C. Wade

European Molecular Biology Laboratory, Heidelberg, Germany

Molecular surfaces are widely used for characterizing molecules and displaying and quantifying their interaction properties. Here we consider molecular surfaces defined as isocontours of a function (a sum of exponential functions centered on each atom) that approximately represents electron density. The smoothness is advantageous for surface mapping of molecular properties (e.g., electrostatic potential). By varying parameters, these surfaces can be constructed to represent the van der Waals or solvent-accessible surface of a molecular with any accuracy. We describe numerical algorithms to operate on the analytically defined surfaces. Two applications are considered: (1) We define and locate extremal points of molecular properties on the surfaces. The extremal points provide a compact representation of a property on a surface, obviating the necessity to compute values of the property on an array of surface points as is usually done; (2) a molecular surface patch or interface is projected onto a flat surface (by introducing curvilinear coordinates) with approximate conservation of area for analysis purposes. Applications to studies of protein–protein interactions are described. © *1996 by Elsevier Science, Inc.*

INTRODUCTION

The surface of a molecule bears information about how it interacts with other molecules and its solvent. As the surface of a molecule is not a quantity for which a unique physicochemical definition exists, several definitions have been introduced. Those most commonly used are as follows:

- The *probe* or *solvent-accessible surface*, i.e., the boundary of the region where the centers of the atoms of other molecules with the radius of the probe are allowed to be placed.[1] The solvent-accessible areas and volumes of molecules may be used to quantify their interaction with the solvent.[2]
- The *probe-excluded molecular surface*, which is drawn on the surface of the probe as it is rolled over the molecule.[3] It highlights the region belonging to the molecule alone; namely, no part of the probe is allowed to be inside the molecular surface. It may be calculated by construction of a dot surface[3] or analytically.[4]
- The *van der Waals surface* of the molecule, dependent only on the atomic radii and coordinates of the atoms in the molecule. It is equivalent to a molecular surface computed with a probe of zero radius

The analytical molecular surface is smoother than the van der Waals surface and this facilitates its analysis. However, its primary importance is not its smoothness but rather the fact that it gives a description of the molecular interior that differs from that of the van der Waals surface, since the cavities inside the molecule that are not accessible to the solvent probe appear to belong to the interior of the molecule.

The properties of a molecule are usually displayed on its surface by assigning values to points on the surface—all but a minority of which will be on the surface of a single atom and thus possess the properties of that atom. The points can be displayed on a graphical device or used for computational analysis. The analysis of surface properties is usually based on visual inspection for which points are colored or assigned sizes corresponding to the value of a property. Solid rendering, such as exemplified by the GRASP program of Nicholls,[5] can be used to enhance visual quality. There are several algorithms allowing automation of surface analysis. We mention two by way of example: (1) the fully automated detection of clusters of surface points with like properties in order to specify hydrophobic patches on protein surfaces by Lijnzaad et al.,[6] and (2) the location of knobs and holes on a protein surface, using a geometric hashing algorithm applied to coordinates derived from a dot representation of the molecular surface, by Fischer et al.[7] A

Color Plates for this article are on page 374–375.

Address reprint requests to: R.C. Wade, European Molecular Biology Laboratory, Postfach 10.2209, Meyerhofstrasse 1, 69069012 Heidelberg, Germany.

Paper submitted as part of the Electronic Conference of the Molecular Graphics and Modelling Society, October 1996.

655 Avenue of the Americas, New York, NY 10010

0263-7855/97/$15.00
PII S0263-7855(97)00008-8

reasonably accurate representation of a molecular surface requires 10–30 points per Å^2, generating 100 000–300 000 points per surface for medium-sized proteins. While the scanning of all surface points is fast, considerable computational effort is necessary to establish the connectivity (neighborhood) of the points. Operations with dot surfaces can, however, be made extremely efficient by avoiding time-consuming distance checks.[8]

An alternative description of a molecular surface, which is analogous to the van der Waals surface, may be derived from a Gaussian description of the molecule.[9] It is defined using the approximate electron density distribution:

$$|\psi(\mathbf{r})|^2 = \sum_i w_i \cdot e^{-|\mathbf{r} - \mathbf{r}_i|^2/\sigma_i^2} \tag{1}$$

representing the contribution from each atomi i of the molecule with different weighting factors w_i, and taking into account the differnt sizes σ_i of the atoms. This form of representation is rather arbitrary and exponential functions such as:

$$|\psi(\mathbf{r})|^2 = \sum_i w_i \cdot e^{-|\mathbf{r} - \mathbf{r}_i|/\sigma_i} \tag{2}$$

should give a more correct description of the asymptotic behavior of electron density.

The volumetric properties of the Gaussian molecular representation have been compared to those of other surfaces by Duncan and Olson[9] and Grant and Pickup.[10] They found that a Gaussian representation could reproduce area and volume quantities computed using a hard-sphere model. The Gaussian representation is exploited to locate specific regions on molecular surfaces in the SURFNET program of Laskowski.[11] Physically, the Gaussian representation provides a much more natural and realistic description of the shape of molecules than the atomic hard-sphere representation, e.g., it can describe the fuzziness of molecular surfaces owing to high-frequency atomic vibrations.[12] Another advantage is its continuity, which allows straightforward analytical estimation of the derivatives of surface-dependent properties, in contrast to the computational difficulties of dealing with the discontinuous hard-sphere representation.

In this study, we use a related, exponential density function to construct molecular surfaces. It differs from previous uses of Gaussian functions in the following respects:

1. The function is constructed to define molecular surfaces rather than the molecular interior. We intentionally study the surfaces directly, since quantization and numerical operations on the two-dimension (2D) surface require considerably fewer operations and memory than operations on the 3D density function itself
2. Surfaces similar to the solvent-accessible surface of the molecule can be derived as well as those corresponding to its van der Waals surface. The solvent-accessible surface is easier to treat analytically because of its relative simplicity
3. The parameters of the density function are assigned to define the degree of smoothing of a hard-sphere surface. These parameters are related to electron density parameters to the extent that the hard-sphere model reflects the electron density

The starting point for this study is the analytical definition of the surface approximating the solvent-accessible or van der Waals surface. The surface is defined implicitly, using the functional of the distance to it (simply, distance = 0). Using this functional, one can make a numerically fast and stable projection to the surface on which one can place a set of approximately equally spaced points with obvious connectivity. The latter may be used to build a grid of quasi-curvilinear coordinates on the surface. We describe two important applications of these surfaces.

- The first application is to couple a potential function to the surface so that the potential function appears to be defined on the molecular surface. It then has meaningful extrema separated from singularities at sources of the potential or asymptotic values at infinity. These extrema are the most concise representation of the potential function
- The second application is the mapping of part of the surface or a molecular interface to a flat rectangle. The mapping allows the projection of the surface from 3D to 2D with conservation of neighborhood and reasonably small distortions. Although the mapping algorithm used may give rise to overlaps when the distance conservation requirement is strong, overlaps decrease as this requirement is weakened. The mapping algorithm is particularly suitable for analyzing the solvent-accessible surface of a molecule, and can be used for the van der Waals surface although interpretation is more difficult as overlap is more of a problem. Interfaces between molecules can be mapped reasonably in the majority of cases

DISTANCE TO THE SURFACE

Let us define an exponential parametric function g(**r**, A;d) associated with atom A, whose center is positioned at $\mathbf{r}_0$ and that has a radius a_0. The parametric function depends on coordinate **r** and adjustable parameter d:

$$g(\mathbf{r}, A;d) = e^{-(|\mathbf{r}-\mathbf{r}_0| - a_0)/d} \tag{3}$$

While other functional forms may be equally appropriate, we consider only the preceding exponential function. A simple manipulation gives the exact value of the distance to the surface of the atom A, ρ(**r**, A;d):

$$\rho(\mathbf{r}, A;d) \equiv |\mathbf{r} - \mathbf{r}_0| - a_0 = -d \ln[g(\mathbf{r}, A;d)] \tag{4}$$

which is independent of the parameter d. Assignment of zero atom radius a_0 would give the distance to the atom center.

Now let us consider a molecule M composed of atoms A_i, i = 1,2, . . . ,N, with radii a_i and coordinates $\mathbf{r}_i$. One can define the exponential parametric function G(**r**, M;d) for a molecule as a sum of those for every atom:

$$G(\mathbf{r}, M; d) = \sum_i g(\mathbf{r}, A_i; d) \tag{5}$$

Then, the distance functional is

$$\rho(\mathbf{r}, M; d) = -d \ln\left[G(\mathbf{r}\ M; d) \right] = -d \ln\left[\sum_i e^{-(|\mathbf{r}-\mathbf{r}_i|-a_i)/d} \right] \tag{6}$$

This definition is formally equivalent to an isocontour

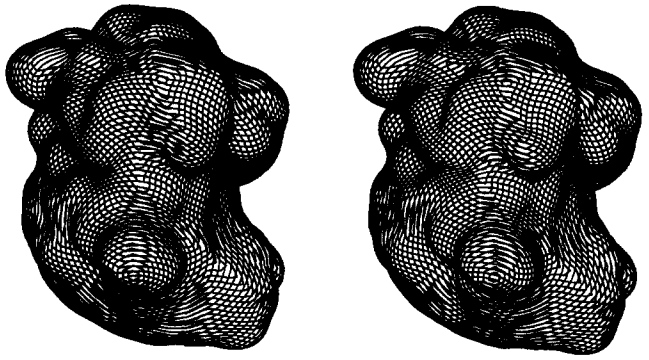

Figure 1. Contours on the solvent-accessible surface of human growth factor (hGH)[13] computed with d = *1.0 Å. There are 6 911 points at the surface. The solvent-accessible surface area, computed from the hard-sphere representation, is 9 785 Å*^{2}.

value of the sum in Eq. (5), but is easier to handle numerically.

At any point **r**, the major contribution to the sum in Eq. (5) will be from the closest atom. Thus, the distance, ρ(**r**, M;d), derived from G(**r**, M;d), will largely reflect the distance to the closest atom (as the distance to a collection of atoms should do according to molecular surface ideology). Variation of the parameter d changes the relative contribution of the close atoms to the sum. As d is decreased, the number of contributing atoms decreases and, when d = 0, only the closest atom contributes.

Depending on which radii a_i are ascribed to the atoms, one obtains the distances to different surfaces: (1) If all a_i are atomic van der Waals radii, then ρ(**r**, M;d) = 0 defines an approximation to the van der Waals surface; (2) if all a_i are van der Waals radii of atoms incremented by a solvent probe radius, then the solvent-accessible surface is approximated.

The choice of the value of the parameter d should be related to the characteristic interatomic distances. At any given point on the surface, the relative contributions to the sum in Eq. (6) arising from the closest atom and the next closest atom can be compared. Considering a typical interatomic distance of 1.5 Å between these atoms, the contribution to the sum from the second atom will be 0.05 and

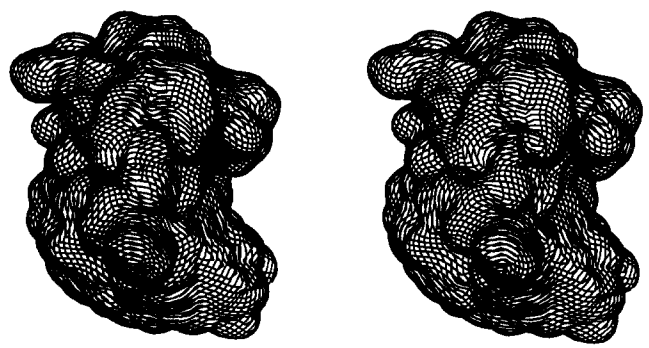

Figure 2. Contours on the solvent-accessible surface of human growth factor (hGH)[13] computed with d = *0.5 Å. There are 7 497 points at the surface.*

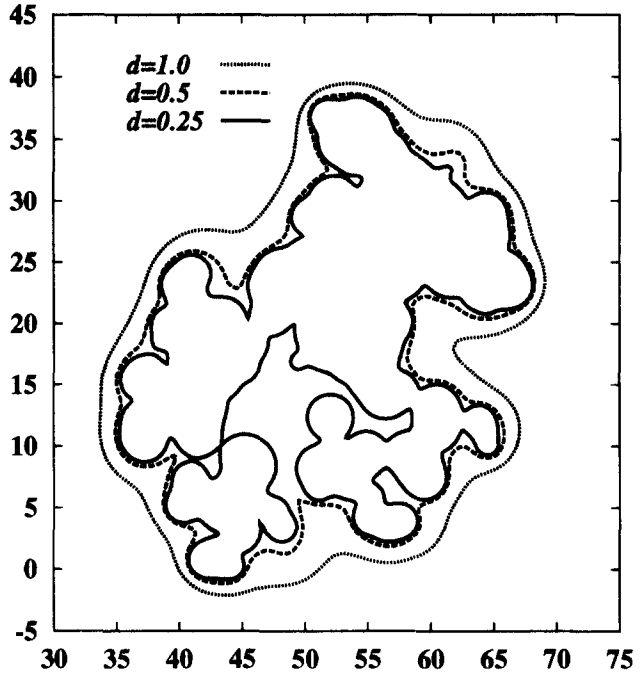

Figure 3. Planar section (axes labelled in Å) through the same molecule, hGH,[13] showing contours corresponding to van der Waals surfaces at d = *1.0, 0.50, and 0.25 Å. The contour with* d = *0.5 Å essentially shows all the details of the molecular surface except the cavities within the molecule. At* d < *0.25 Å the contours are not of closed form, making continuous tracing of the surface impossible.*

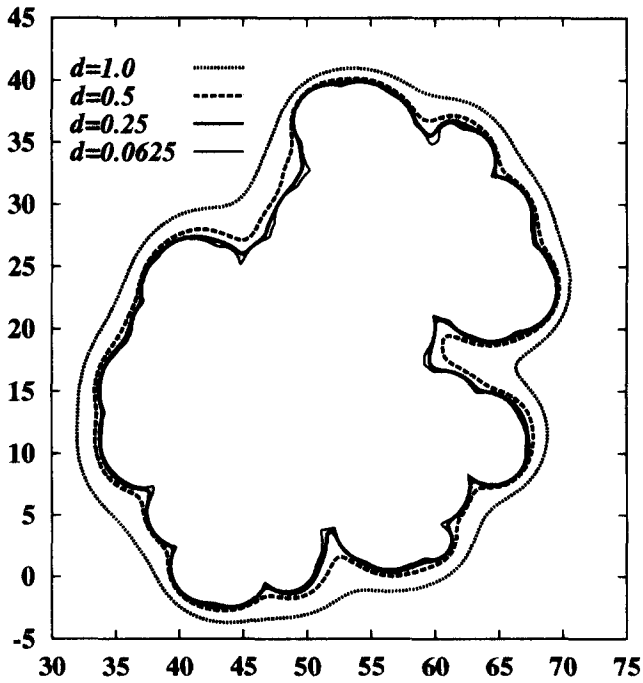

Figure 4. Contours on the same cross-section plane as in Figure 3 for solvent-accessible surfaces computed with four values of d. *Note that surfaces computed with* d < *0.25 Å are almost indistinguishable.*

0.22 times that of the closest atom for values of d of 0.5 and 1.0 Å, respectively. Consequently, considering only the two closest atoms, the $\rho(\mathbf{r}, M;d) = 0$ surface will follow the van der Waals or solvent-accessible surface with a deviation of only 0.025 Å at d = 0.5 Å or 0.2 Å at d = 1.0 Å. If the geometric positions of the atoms are different from those described above, or more atoms are considered, the distortion may be much greater. Figures 1–4 illustrate the dependence of the computed surfaces on the parameter d for the human growth factor hormone.[13]

The surface of molecule M is thus defined by $\rho(\mathbf{r}, M;d) = 0$. To define the interface between two molecules, M_1 and M_2, the definition of the distance to the molecular surface given in Eq. (6) can be used in

$$\rho(\mathbf{r}, M_1;d) = \rho(\mathbf{r}, M_2;d) \quad (7)$$

where the distances $\rho(\mathbf{r}, M_1;d)$ and $\rho(\mathbf{r}, M_2;d)$ are the distances to the surfaces of the first and second molecules, respectively. The parameters a_i in the definition of the distance Eq. (6) can be modified to obtain different interfaces defined as equidistant to molecular surface using atomic radii or defined as equidistant to the closest atoms by setting all radii a_i to zero.

OPERATIONS ON THE SURFACE

The performance of some of the operations to be described below depends on the value assigned to the parameter d, defining the closeness of the computed surface to a hard-sphere surface. Except for the first two, which are performable at any d, mapping of the surface can usually (for proteins) be done only at $d > 0.5$ Å when the surface is smooth enough to keep the mapping operations stable. Improvement of the algorithms may, however, enable mapping of the more complicated surfaces obtained at smaller d.

Placing a point on the surface

The surface is defined implicitly and is smooth. The basic operation for moving from any given point to the point on the surface is projection to the surface along the gradient of the distance Eq. (6). Projection is performed iteratively with a step size equal to the distance to the surface.

Motions along the surface

Motions along the surface from any point can be done using the surface tangents. The choice of the functional given by Eq. (6) makes the higher order derivatives small, so the tangent movements can be corrected by a few iterations (typically 1 or 2, when the deviations from the surface are less than 1 Å), when the step size of the tangent motion is within 1–3 Å. Consequently, the move along the surface costs several computations of sums like that in Eq. (6) (the functional itself and its first three derivatives are computed at each iteration). Computation of the sum in Eq. (6) scales as the number of atoms: The same number of motions will cost only twice as much when the molecule size is doubled.

Mapping part of a surface

Locally, the molecular surface resembles a Euclidean rectangle (being topologically equivalent to it). One can define pseudo-Euclidean coordinates on an interesting part of the surface. It is natural to introduce polar coordinates, starting from some specific point on the surface. This can be achieved by generating rings of points starting from a given center. A variety of algorithms can be used. We introduce the basal distance D to define the distance between the points on the surface. Each ring of points is constructed at this distance from the previous ring and then projected onto the surface. The points on each newly constructed ring are then checked, one after the other, to see if the angle formed by the lines joining the point with its neighbors is less than a threshold (77°), in which case the point is eliminated. Then a check of the distance to the next point is performed. If this is less than 0.75D, the next point is eliminated; whereas if it is greater than 1.7D, one more point is added. An example of a set of points resulting from applying this procedure is shown in Figure 5. This method is similar to that used by Bacon and Moult,[14] who introduced web coordinates on molecular surface patches by fitting the precomputed surface points by B-splines and constructing a self-growing web. We use the ring construction algorithm to map the interfaces between two molecules. If D = 1 Å, the algorithm places ca. 1 point per 1.225 $Å^2$ of the surface.

Scanning the whole molecular surface

To represent the molecular surface as a whole, one needs a representative set of points on the surface that covers the entire surface as uniformly as possible. The simplest solution is to project every atom center in the molecule onto the closest point on the surface. A drawback of this procedure is that the set of points on the surface cannot be made more dense unless a check over all surface point pairs is done, because the points are not ordered according to proximity. It is just such computationally intensive searching that we wish to avoid. An alternative procedure is to use the algorithm described above in the previous section to map the entire molecular surface. For relatively spherical molecular surfaces, the growing rings may eventually contract to a point on the side of the molecular surface opposite to that at which growing commenced, thus defining the spherical coordinates. These mapped coordinates may be used to locate

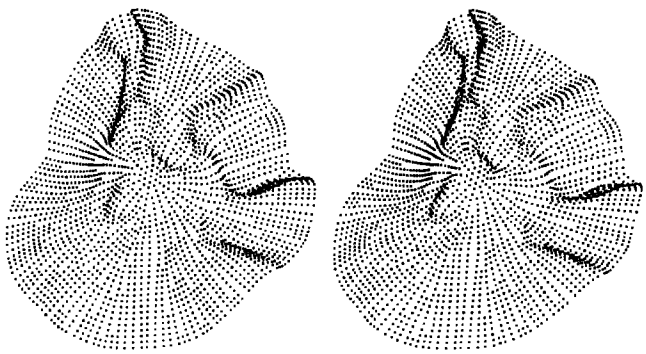

Figure 5. Example of a set of points resulting from the mapping procedure described in text.

every point on the surface. Making a more dense representation is straightforward since the set of points is ordered. Examples are presented in Figures 1 and 2, where this algorithm was used to cover the solvent-accessible surface of the molecule by a sequence of rings.

FINDING POTENTIAL EXTREMA

Potentials representing diverse properties, for example, electrostatic potential, lipophilicity, hydrophobicity, or the curvature of the surface itself, can be studied. To locate the extrema of the potential on the surface, the formalism of Lagrange multipliers is used and the equations are solved by the method of Newton (see, for example, Ref. 15). For this, it is necessary to compute the first and second derivatives of the potential and the distance functional. These can be computed analytically if the potential and its gradient are smooth.

Figure 6 shows an example of the characterization of a potential on the surface of a molecule by its extremal points. The circles are drawn to show the sizes (derived from the second derivative) of the minima and maxima along the surface tangent plane. Saddle points are shown as lines along the two main directions, defining the steepest increase and decrease of the potential. Surprisingly, the potential at the (smoothed with d = 1 Å) 1.4-Å probe-accessible surface of the molecule is not extremely complicated, having only 50–100 extremal points in all. These points can be used as a minimal representation to restore the potential and to compare molecules. For instance, many molecular properties correlate with the properties of its electrostatic potential as described, for example, by Murray et al.[16] and Richard.[17]

PROJECTIONS OF MOLECULAR SURFACES AND INTERFACES

Surfaces are usually defined as a set of points that is visualized with the aid of graphical programs and analyzed by eye. The automation of the procedure would not only save time spent in analysis, but would also avoid possible errors resulting from the subjective nature of the manual analysis procedure. The projection of two different surfaces onto one simple surface is necessary for the comparison of these surfaces (see review by Masek[18]). A natural solution is to project the surface onto a plane, where it can be quantified in a straightforward manner. However, simple projection of the entire molecular surface encounters a topological problem (met already in mapping the surface of the earth), since the connectivities between points on a closed surface can be better preserved by projecting onto a sphere rather than a plane. A beautiful solution to this problem for small molecules was suggested by Gasteiger et al.,[19,20] who used Kohonen maps to project the surface onto a torus conserving neighborhood.

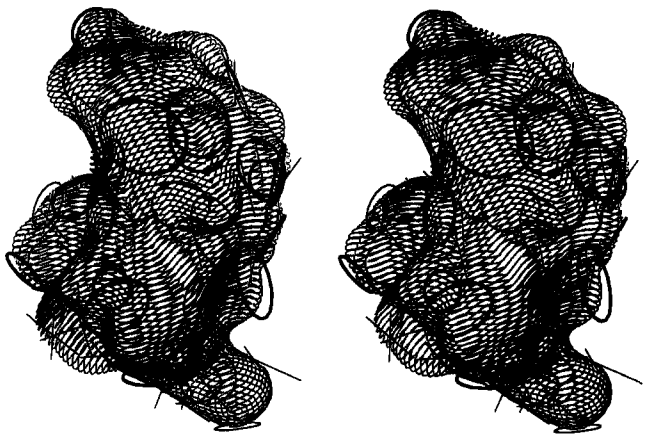

Figure 6. Minima, maxima, and saddle points of the electrostatic potential on the smoothed, solvent-accessible surface (shown by gray contours) of hGH.[13] Thin-line circles show the position and extension of minima (which, in this example, all have a negative value of the potential), thick-line circles show the maxima, and crosses show the saddle points of the potential on the surface.

For projecting the surface onto a sphere, the gnomonic projection[21] provides the simplest solution. The surface is placed within the sphere and then projected along the radials of the sphere. The feasibility of approximately conserving distances between points on the surface on projection is highly dependent on the relative orientations of the sphere and surface and distance conservation will not be uniform over the surface. Moreover, two points will often be projected onto one, resulting in loss of information. A projection that avoids overlaps is the spherical harmonics representation developed by Duncan and Olson,[22] in which a special procedure eliminates overlaps resulting from the gnomonic projection.

The solution proposed in this work is based on the analytical definition of molecular surfaces and the algorithm of growing rings, which builds up quasi-polar coordinates on the surface. This construction may overlap onto itself, causing the duplicate projection of some regions of the surface. The duplication problem is inevitable as can be appreciated from imagining covering an irregular surface with a piece of paper. The duplicate covering obviously can not be completely avoided except in the case of a planar surface. However, most of the duplications can be avoided by weakening the requirement of distance conservation, in an analogous way to covering the surface with a piece of stretch film rather than paper. Depending on the curvature of the surface, this will cause different degrees of distortion. On the other hand, overlaps caused by topological differences between the original and projection surfaces should be handled either by projecting onto an appropriate simple surface or by additional description, for example, specifying the equivalence of opposite borders of a rectangle onto which a torus has been projected.

Once the surface has been covered by a set of rings, the mth point of the nth ring can be projected onto a point (x,y) on the plane using the following formulas:

$$x = n \cdot \cos(2\pi m/M)$$
$$y = n \cdot \sin(2\pi m/M) \qquad (8)$$

where M is the number of points in the nth ring.

As a result, various properties on the surface or interface are transferred to an appropriate Euclidian rectangle.

Figures 1 and 2 show surfaces mapped while avoiding duplications. A correction of the procedure described in Scanning the Whole Molecular Surface (above) avoids local

sources of duplication, so that the growing rings eventually evolve to a point on the other side of the molecule. The ring numbering can serve as latitude and the length parameter along each ring as longitude. This provides a means by which to introduce spherical coordinates for a molecular surface automatically.

EXAMPLE OF INTERFACE ANALYSIS

By way of illustration, we provide a new view in Figures 7–12, and Color Plates 1 and 2) of the human growth hormone (hGH)–hGH receptor (hGHR) interfaces analyzed by Clackson and Wells.[23] The available crystal structure[13] consists of the ternary complex of one hGH molecule bound to two identical receptors: hGHR1 and hGHR2. Binding at the first interface, hGH–hGHR1, shown in Figure 7a, is high affinity in nature while that at the second interface, hGH–hGHR2 (see Figure 7b), is low affinity. The buried areas on the hGH–hGHR1 and hGH–hGHR2 interfaces are 1 300 and 900 Å^2, respectively. The two binding surfaces of hGH differ in sequence while the binding surfaces of the two receptors are similar (but not identical) both spatially and in terms of the residues involved.

For the mapped interfaces shown in Figures 8–11, and in Color Plates 1 and 2, the areas in the maps approximately equal the areas on the interface in three-dimensions. The scale is such that each map has dimensions of about 66 by 66 Å, and the tick marks are positioned at approximately every 1.1 Å. The contour plots of two-dimensional functions on a rectangle were generated using the program XFarbe.[24]

Interfacial cavities and crevices can be identified from Figure 8. Those for which the interprotein distance is greater than about 5 Å are large enough to accommodate water molecules. As can be seen from Color Plate 1, the interfaces consist of residues from discontiguous parts of the sequences of the proteins. However, there are some continuous stretches of residues at the interfaces, notably the participation of the D helix and C terminus of hGH (residues 155–190) in interactions with the hGHR1 is readily seen in Color Plate 1 (top right) as a more ordered region of residues within a smaller range of numbers. The interface to HGHR1 is, however, not centered on this helix but between two separate secondary structure patterns. On hGHR1, there are three residues that contribute large surface areas to the interface. Tryptophans 104 and 169 contact part of the hGH D helix (residues 155–184) and the AB loop (residues 35–71, especially the N-terminal part of minihelix 2, consisting of residues 64–70). Asparagine 218 interacts with part of the hGH A helix (residues 9–34). It has been shown experimentally that the high affinity of the hGH:hGHR1 complex is due to the contribution of a "hot spot" on the hGHR1 receptor consisting of a small number of hydrophobic residues, including W104 and W169.[23] The hydrophobic patch at the interface created by these residues is clearly visible in Figure 9. It is the largest continuous hydrophobic patch on the interface. It satisfies the description on the binding epitope as consisting of a hydrophobic center surrounded by polar moieties. An equivalent patch, due also to W104 and W169, is observed on the hGHR2 in the second (lower affinity) interface, but it differs in shape and in the arrangement of the polar residues around it. The second patch on

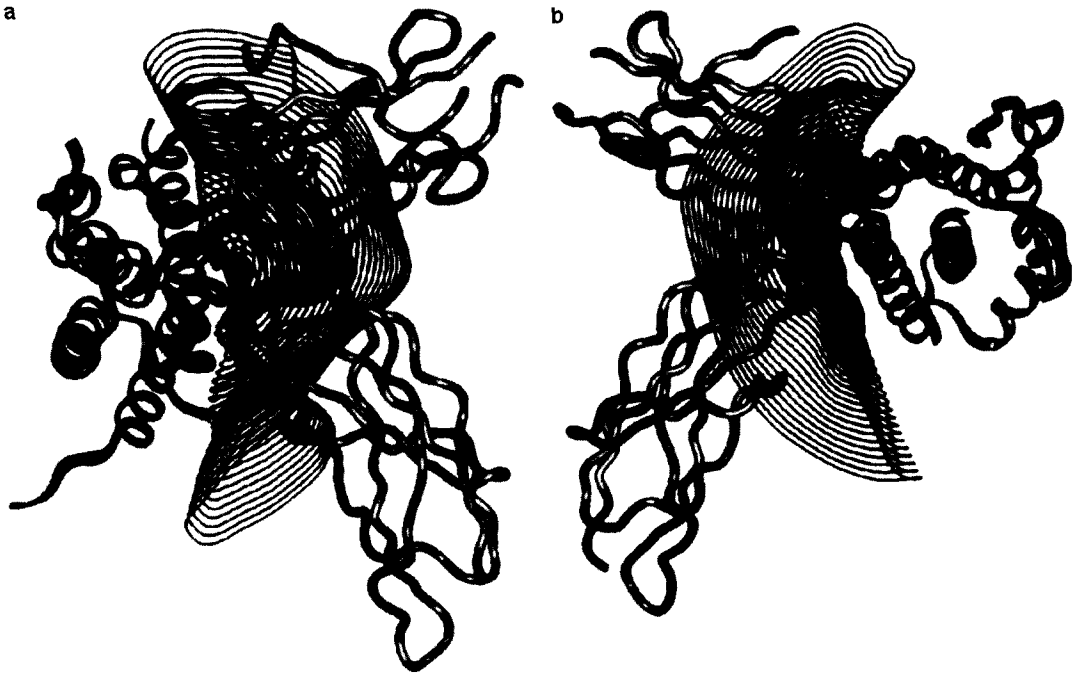

Figure 7. Ribbon representation of (a) hGH (dark ribbon) and the high-affinity hGHR1 (light ribbon) and (b) hGH (dark ribbon) and the low-affinity hGHR2 (light ribbon), as arranged in the crystal structure of the ternary complex of hGH and the extracellular part of the homodimeric receptor. Surfaces are shown for each of the hormone–receptor interfaces. These surfaces were projected onto two dimensions and their properties are displayed in Figures 8–11, and in Color Plates 1 and 2.

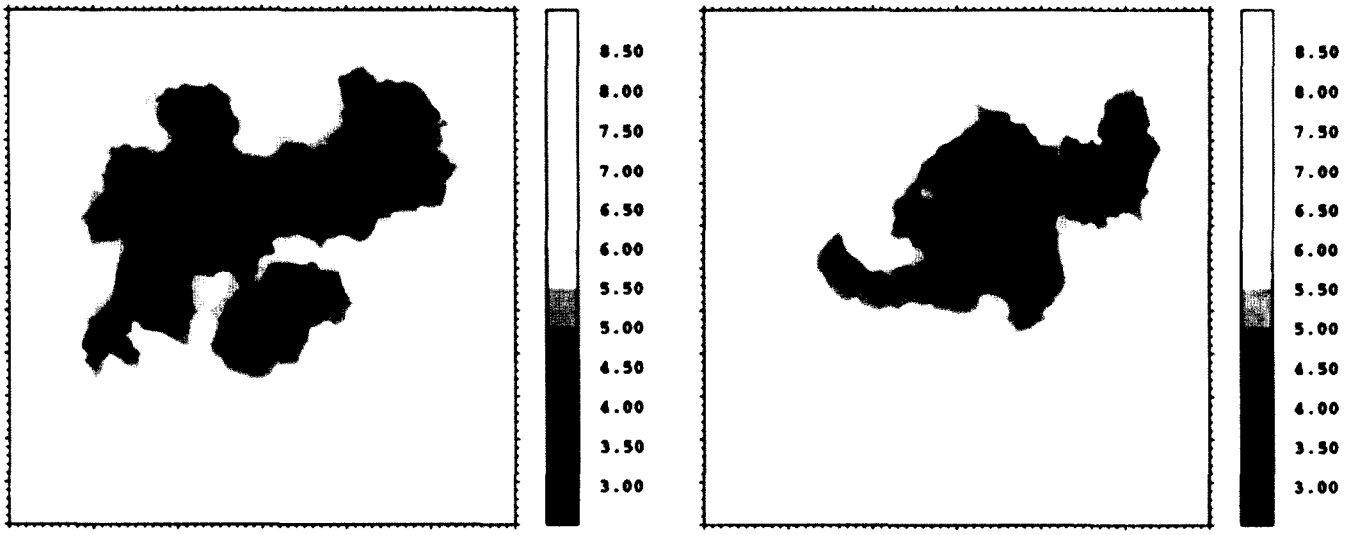

Figure 8. Mapping displaying the interprotein distances (in angstroms for the hGH–hGHR2 (left) *and hGH–hGHR1* (right) *interfaces. The distance is computed at every interface point as a sum of the distances to the closest atoms of the two proteins. The latter distances are derived from the distance functional [Eq. (6)] with atomic radii set to zero.*

the hGHR1 interface (formed mainly by N218 of hGHR1) also has a large buried area. However, it contributes much less to the binding energy, as revealed by mutational analysis.[23] It can be seen from Color Plate 2 that this patch has less hydrophobic complementarity, being composed of hydrophilic residues of hGHR1. Moreover, it is rather mobile, unlike the first patch. It is interesting that it is the mobility rather than hydrophobicity that clearly distinguishes the first patch from the second (see Figure 10 vs. Figure 9), implying that the interactions between less mobile residues define the binding free energy of these two proteins. The residues forming the second patch of the interface to hGHR1 have much smaller interfacial areas on the interface to hGHR2.

The electrostatic potentials of the two proteins show a high degree of complementarity at the interface, as can be seen from Color Plate 2, where regions of positive (negative) potential from one protein match regions of negative (positive) potential from the other. The electrostatic potentials were computed with the program UHBD.[25] The potentials of the hormone at the two binding interfaces are similar as was demonstrated in Ref. 26. This is seen by comparing the right sides of the upper and the lower portions of Color Plate 2, taking into account the different transformations applied during projection: In both cases, the potential of the hormone consists of patches in the sequence: negative–positive–negative–positive (from left to right).

Figure 11[27] shows the similarity index for the potentials from hGHR1 and hGH plotted on every interface point, as well as a simple product of the two potentials. To compute the potentials used in Figure 11, the polar hydrogen atoms were simply added to the crystal structure with the HBUILD option in QUANTA.[28] As can be seen from Figure 11, the complementarity of the two potentials (given by negative values of the similarity index) is not uniformly distributed over the interface. There are at least four regions where the potentials from the two molecules are in conflict, having the same sign and having considerable absolute values (see the right side of Figure 11). To evaluate whether these result from assignment of poor hydrogen atom positions, we energy minimized the hydrogen atoms (fixing all heavy atoms) with QUANTA/CHARMm[28] for 150 steps of steepest descent minimization. The lower portion of Figure 11 shows the same properties as the upper portion for the resulting conformation.

As a result of minimization, one highly noncomplementary patch (region B) disappeared. The other noncomplementary patches, however, remain. The conformational changes due to minimization are rather small, as can be seen from the 3D images (shown in Figure 12a–d) of regions corresponding to the marked patches of noncomplentarity in Figure 11. Noncomplementarity in region B is removed by a conformational change in Thr-175 of hGH, the added hydroxyl hydrogen atom of which was in obvious conflict with the side-chain atoms of Arg-43 of hGHR1. The small noncomplementarity in region A is not affected by minimization. Noncomplementarity in region D is reduced, but this is not only due to minimization: we also interchanged the amide nitrogen and oxygen atoms of the side chain of Gln-46 of hGH, which, as a result, has a favorable electrostatic interaction with Glu-120 and Thr-77 of hGHR1. The interactions in region C are very complicated and remained essentially unoptimized after minimization. It is interesting that there is a correlation between electrostatically noncomplementary regions, mobile regions, and the regions providing a small contribution to binding.

In summary, this example shows the different features contributing to binding. The "hot spot" is a particularly rigid (black spot in Figure 10), hydrophobic (large, light gray spot in Figure 9 and white spot in the upper portions of Color Plate 2 and Figure 11) region of high shape complementarity (black spot in the left-hand portion of Figure 8) involving two tryptophans (Color Plate 1) surrounded by polar residues in highly complementary positions (best seen in Figure 11).

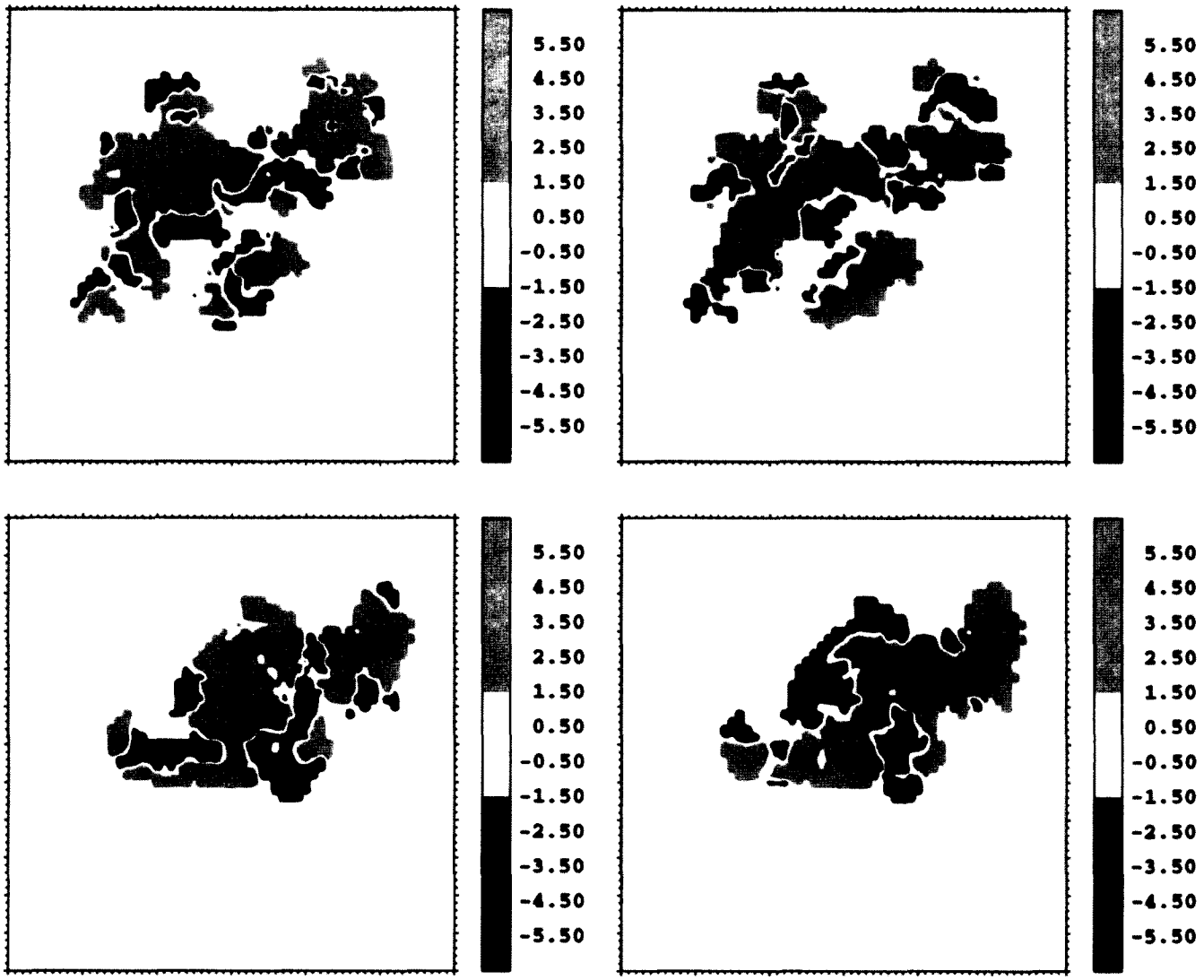

Figure 9. Mapping of the hydrophobicity properties (following Eisenberg and McLachlan[2]) of the closest atoms of hGHR1 (top left) *and hGH* (top right) *onto the hGH–hGHR1 interface. Light gray regions are hydrophobic; dark gray regions are hydrophilic.* Bottom: *The same for the hGH–hGHR2 interface. More features can be seen when displayed in color.*

ACCURACY OF AREA VALUES

There is approximate area conservation during the mapping procedure. When the reference distance for ring generation is D = 1 Å, approximately one point is placed per 1.225 Å^2. However, the distribution of the points is not uniform. Projection of the surface rings to the rings of the polar coordinates on a planar rectangle further distorts the areas.

We repeated the mapping for the two interfaces for hGH binding to hGHR1 and hGHR2, using different starting points. In both cases, the interface remains the same since its definition is independent of mapping. Distortions occur when the surface is projected onto the planar square and these distortions are related to the position of the interface with respect to the center/starting point. There is no strict correlation between the distance to the chosen center and the distortions, although the area around the center will always be less distorted. Mapping with different centers is a good test of area-conserving properties. Figure 13 shows the correlation between the areas assigned to every residue of both proteins at both interfaces during two different mappings. Deviations are within either the absolute limit of x ± 7 points or the relative limit of x ± 0.3x. These deviations depend mainly on the surface curvature, which defines the degree of distortion.

The distortions described above are introduced by the projection procedure. However, the placing of the points on the interface or surface is also subject to error. These distortions can be quantified by relating the accessibility of each residue to the number of surface points generated on the analytical surface that are closer to this residue than any other (see Figure 14). There are two sources of deviation. The first is caused by the deviation of the analytical smooth surface from the hard-sphere surface used to compute resi-

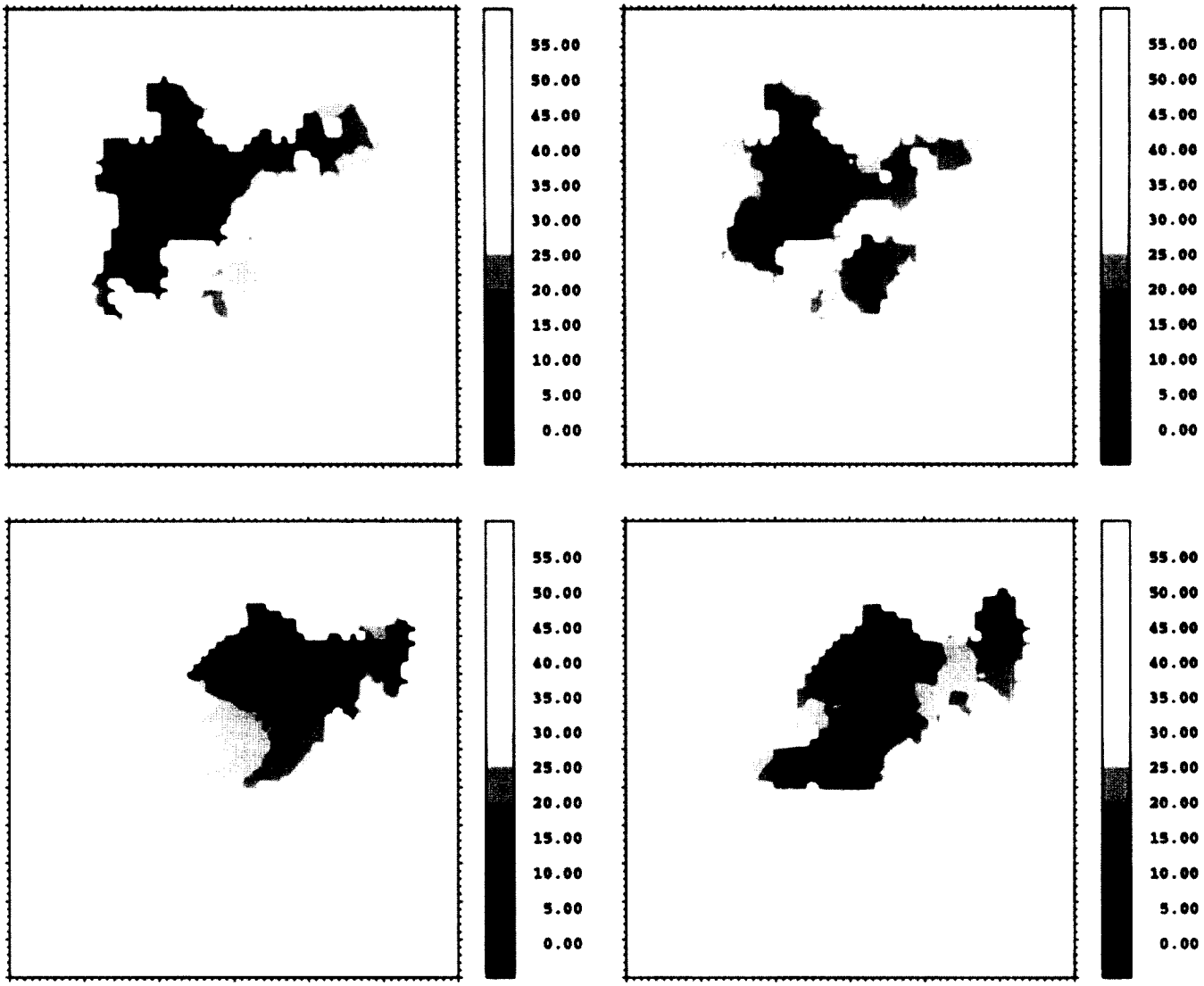

Figure 10. Mapping of the crystallographic temperature factor distribution (in angstroms) of the closest atoms of hGHR1 (top left) *and hGH* (top right) *onto the hGH–hGHR1 interface. Lighter regions are more mobile with temperature factors above 30 Å^2; darker regions are more rigid with temperature factors below 25 Å^2.* Bottom: *The same for the hGH–hGHR2 interface.*

due accessibilities. The second is defined by the algorithm generating the points on the surface, which is designed to generate nonintersecting rings that evolve to a single point when mapping of the surface is complete. As can be seen in Figure 14, although the upper estimate of the error for the surface shown in Figure 2 is 22 Å^2, for the majority of residues the number of surface points per residue correlates much better (with a scaling factor of 1.25) with the residue accessibilities. The scaling factor arises because the surface was constructed so that there is on average 1 point per 1.225 Å^2.

THE PROGRAM

The illustrations for this article were generated by a four-step protocol using the ADS (Analytically Defined molecular Surfaces) program package. In the first step, rings of points are constructed on the interface, using a program that maps the surface of the molecule. In the second step, the required properties are assigned to the points on the surface. In the third step, the surface rings are projected onto a rectangle, producing data files to be used in the fourth step, which is to display the properties and plot the pictures. Output data files are readable by the x-Window-oriented program XFarbe of Preusser,[24] which is used to make a contour representation of the functions on a rectangle. Computer times necessary to perform the manipulations (which have not yet been optimized for speed) are within 2–5 min on an SGI Indy RC 4600 workstation for proteins having ca. 1 500 atoms.

The source codes (in Fortran 77) of the ADS program package are currently under development and it is intended that they will be made available on request.

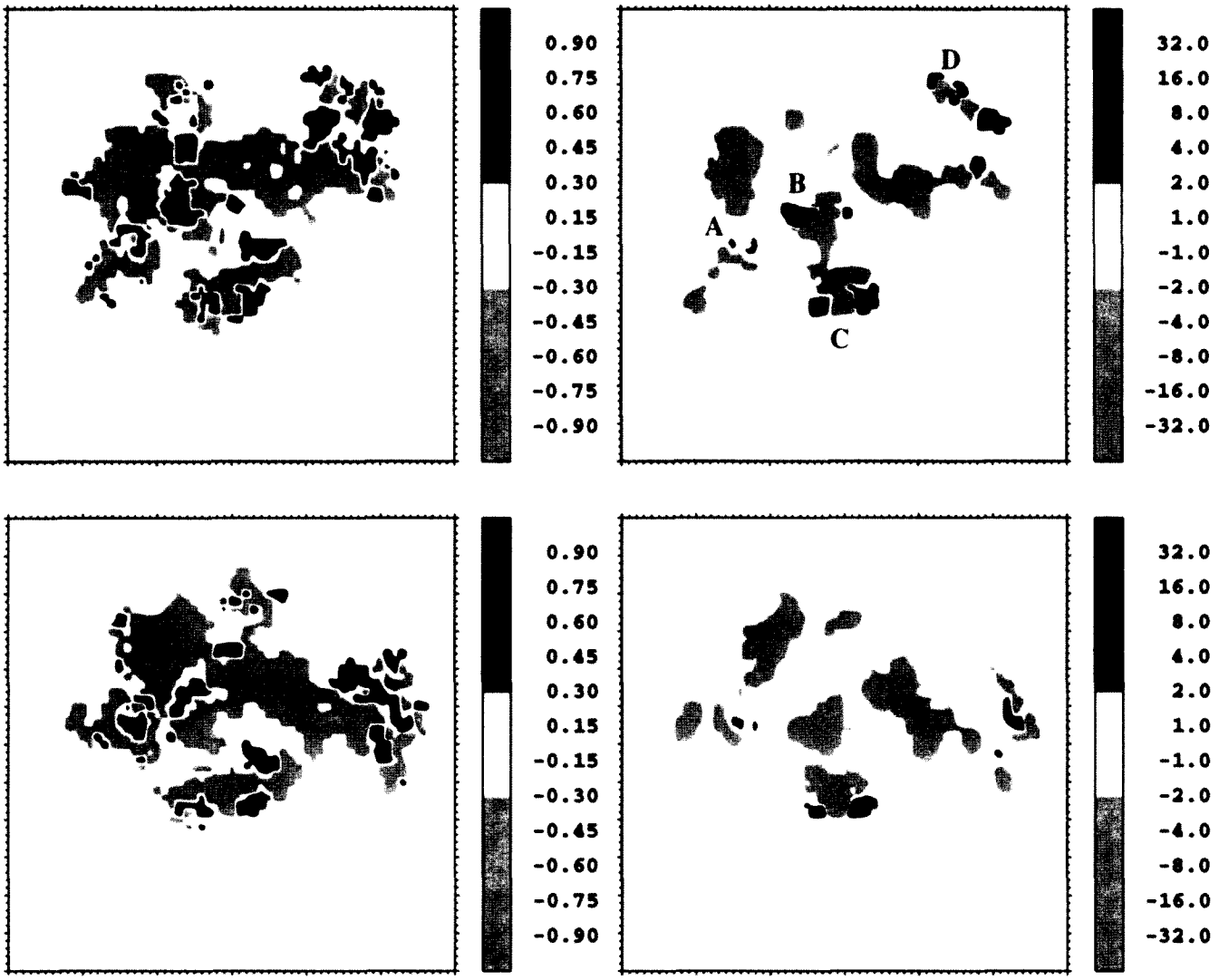

Figure 11. Similarity index (known as the Hodgkin index[27]) as a function of the interface point (top left) *and the product* (top right) *of the two molecular potentials for the hGH–hGHR1 interface. Namely, given the potentials from molecule 1,* $\Phi_1(\mathbf{r})$*, and molecule 2,* $\Phi_2(\mathbf{r})$*, at the point* $\mathbf{r}$ *of the interface, the function shown on the left side is* $2\Phi_1(\mathbf{r})\Phi_2(\mathbf{r})/[\Phi_2^2(\mathbf{r}) + \Phi_2^2(\mathbf{r})]$ *and on the right:* $\Phi_1(\mathbf{r})\Phi_2(\mathbf{r})$**.** Bottom: *The same for the energy-minimized conformation of the hGH–hGHR1 complex. More features can be seen when displayed in color.*

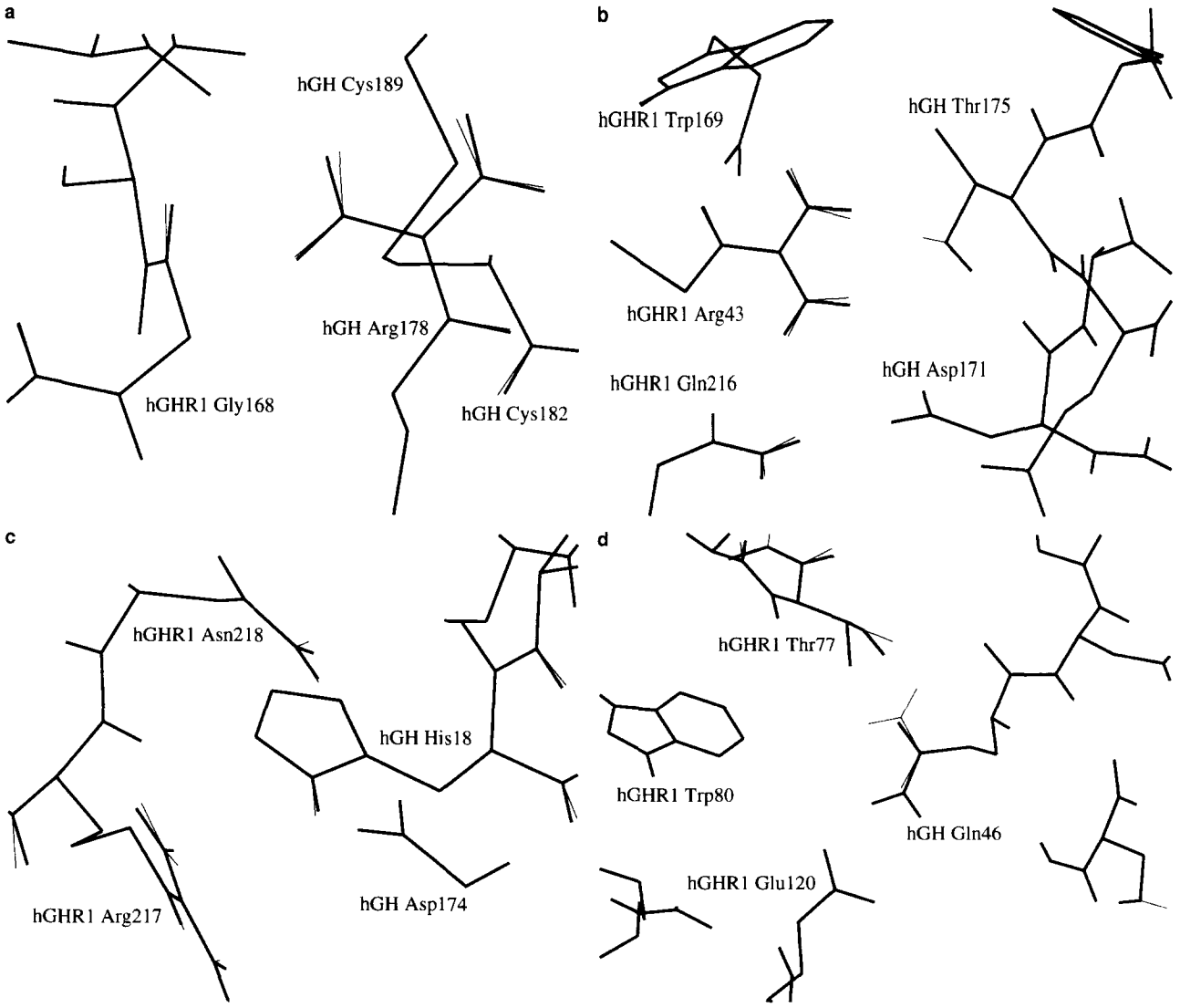

Figure 12. (a–d) Three-dimensional view of the residues in regions A, B, C, and D (respectively) marked on Figure 11 Conformations before (thin lines) and after (thick lines) minimization are shown.

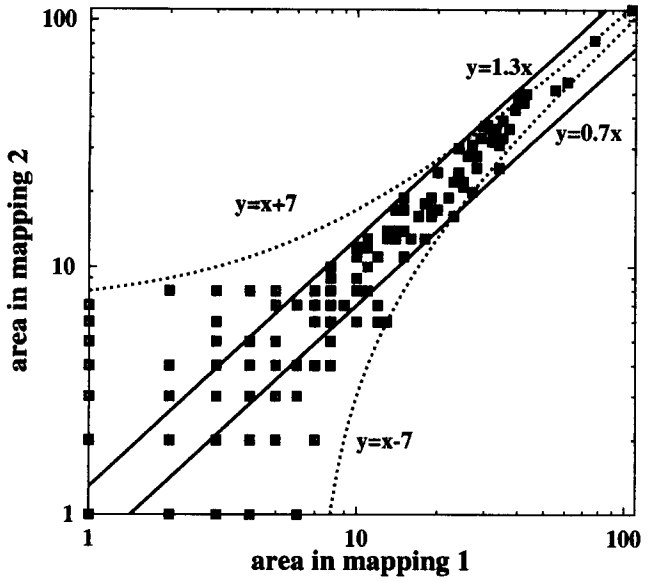

Figure 13. Correlation between areas assigned to residues at the hGH–hGHR1 and hGH–hGHR2 interfaces when mapping is started at points separated by 2 Å. An absolute tolerance limit of x ± 7 *is shown by dashed lines and a relative tolerance limit of* x ± 0.3x. *is shown by solid lines. The scale is logarithmic.*

CONCLUSIONS

We have introduced an analytical definition of smooth molecular surfaces as isocontours of the sum of exponential functions centered on atoms. Depending on the smoothing parameter, the analytically defined surfaces can approximate van der Waals or solvent-accessible surfaces of a molecule with any desired accuracy. We constructed a distance functional to make the implicit definition of the surface easy to use numerically. Moreover, the interface between two molecules may be defined using this definition of distance to the molecules.

We exploited the main property of these surfaces, their smoothness, in order (1) to couple potential functions to a surface and locate extremal points of the potential on the surface by gradient methods, (2) to define pseudo-Euclidean (on part of the surface) or pseudo-spherical (on the whole surface) coordinates on the surface, and (3) to define pseudo-Euclidean coordinates on the interface between two molecules. These representations are useful in visualizing and studying molecular surfaces and interfaces.

The main advantage of these representations is that their complexity or simplicity can be tuned by the user. This attribute is especially important for analyzing the many properties of protein surfaces and interfaces.

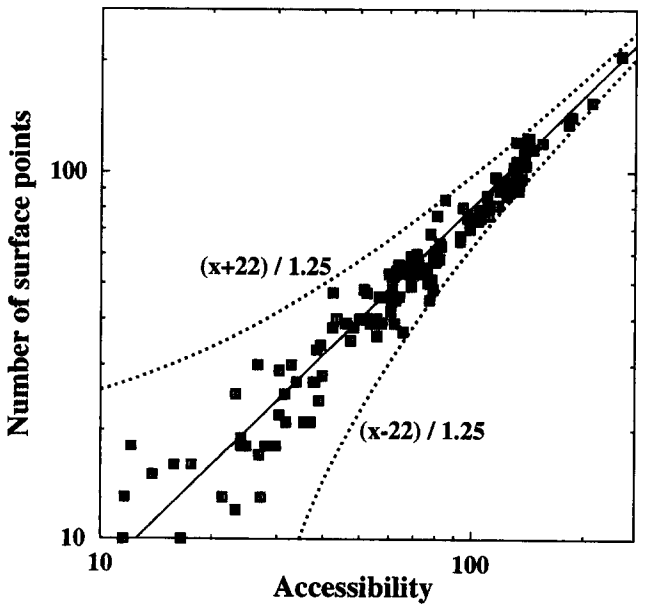

Figure 14. Correlation between the number of surface points per residue on the surface shown in Figure 2 for hGH and residue accessibility.

REFERENCES

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

2 Eisenberg, D. and McLachlan, A.D. Solvation energy in protein folding and binding. *Nature (London)* 1986, **319,** 199–203

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

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

5 Nicholls, A., Bharadwaj, R., and Honig, B. GRASP: Graphical representation and analysis of molecular surfaces. *Biophys. J.* 1993, **64,** A166

6 Lijnzaad, P., Berendsen, H.J.C., and Argos, P. Hydrophobic patches on the surfaces of protein structures. *Proteins Struct. Funct. Genet.* 1996, **25,** 389–397

7 Fischer, D., Norel, R., Wolfson, H., and Nussinov, R. Surface motifs by a computer vision technique: Searches, detection, and implications for protein–ligand recognition. *Proteins Struct. Funct. Genet.* 1993, **16,** 278–292

8 Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., and Scharf, M. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* 1995, **16,** 273–284

9 Duncan, B. and Olson, A.J. Shape analysis of molecular surfaces. *Biopolymers* 1993, **33,** 231–238

10 Grant, J.A. and Pickup, B.T. A Gaussian description of the molecular shape. *J. Phys. Chem.* 1995, **99,** 3503–3510

11 Laskowski, R.A. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J. Mol. Graphics* 1995, **13,** 323–330

12 Agishtein, M.E. Fuzzy molecular surfaces. *J. Biomol. Struct. Dynam.* 1992, **9,** 759–768

13 deVos, A.M., Ultsch, M., and Kossiakoff, A.A. Human growth hormone and extracellular domain of its receptor: Crystal structure of the complex. *Science* 1992, **255,** 306–313; pdb entry 303hhr, notations hGH, hGHR301,

and hGHR302 are used for chains A, B, and C, respectively
14 Bacon, D.J. and Moult, J. Docking by least-squares fitting of molecular surface patterns. *J. Mol. Biol.* 1992, **225,** 849–858
15 Korn, G.A. and Korn, T.M. *Mathematical Handbook for Scientists and Engineers.* McGraw-Hill, New York, 1961
16 Murray, J.S., Brinck, T., and Politizer, P. Relationships of molecular surface electrostatic potentials to some macroscopic properties. *Chem. Phys.* 1996, **204,** 289–299
17 Richard, A.M. Quantitative comparison of molecular electrostatic potentials for structure–activity studies. *J. Comput. Chem.* 1991, **12,** 959–969
18 Masek, B.B. Molecular surface comparisons. In: *Molecular Similarity in Drug Design* (P.M. Dean, ed.). Blackie Academic and Professional, Glasgow, Scotland, 1995, pp. 163–186
19 Gasteiger, J., Li, X., Rudolph, C., Sadowski, J., and Zupan, J. Representation of molecular electrostatic potentials by topological feature maps. *J. Am. Chem. Soc.* 1994, **116,** 4608–4620
20 Gasteiger, J., Li, X., and Uschold, A. The beauty of molecular surfaces as revealed by self organizing neural networks. *J. Mol. Graphics* 1994, **12,** 90–97
21 Chau, P.L. and Dean, P.M. Molecular recognition: 3D surface structure comparison by gnomonic projection. *J. Mol. Graphics* 1987, **5,** 97–100
22 Duncan, B. and Olson, A.J. Approximation and characterization of molecular surfaces. *Biopolymers* 1993, **33,** 219–229
23 Clackson, T. and Wells, J.A. A hot spot of binding energy in a hormone–receptor interface. *Science* 1995, **267,** 383–386
24 Preusser, A. Algorithm 671–FARB-E-2D: Fill area with bicubics on reactangles—a contour plot program. *ACM Trans. Math. Soft.* 1989, **15,** 79–89
25 Madura, J.D., et al. Electrostatics and diffusion of molecules in solutions: Simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* 1995, **91,** 57–95
26 Demchuk, E., Mueller, T., Oschkinat, H., Sebald, W., and Wade, R.C. Receptor binding properties of four-helix-bundle growth factors deduced from electrostatic analysis. *Protein Sci.* 1994, **3,** 920–935
27 Hodgkin, E.E. and Richards, W.G. Molecular similarity based on electrostatic potential and electric field. *Int. J. Quantum Chem. Quantum Biol. Symp.* 1987, **14,** 105–110
28 *QUANTA Molecular Modeling Software Package. Molecular Simulations Inc.* Waltham, Massachusetts, 1992