Molecular Surface Abstraction

Gregory Cipriano, *Student Member, IEEE*, and Michael Gleicher

Abstract—
In this paper we introduce a visualization technique that provides an abstracted view of the shape and spatio-physico-chemical properties of complex molecules. Unlike existing molecular viewing methods, our approach suppresses small details to facilitate rapid comprehension, yet marks the location of significant features so they remain visible. Our approach uses a combination of filters and mesh restructuring to generate a simplified representation that conveys the overall shape and spatio-physico-chemical properties (e.g. electrostatic charge). Surface markings are then used in the place of important removed details, as well as to supply additional information. These simplified representations are amenable to display using stylized rendering algorithms to further enhance comprehension. Our initial experience suggests that our approach is particularly useful in browsing collections of large molecules and in readily making comparisons between them.

Index Terms—molecular surfaces, molecular visualization, surfaces, textures, cartographic labeling

1 Introduction

One goal of structural biology is to understand the chemical and physical properties of macro-molecules (especially proteins) and how this enables the chemical reactions behind life's processes. In order to study these large and complex molecules, biochemists rely on visualizations that provide various levels of abstraction. The more abstract visualizations portray a molecule's internal structure. However, protein interactions involve the "functional surface" presented: to a large degree, the internal structure simply exists as scaffolding to place various forces and chemical properties in proper spatial relationships with one another. While visualizations of these functional surfaces exist, they portray all of the detail and complexity of large molecules. The complexity of these visualizations is problematic as they do not afford rapid assessment, and details may obscure larger scale phenomena. To date, the degree of abstraction provided for internal structure has not been shown for external properties.

In this paper, we introduce *abstracted molecular surfaces,* a visualization technique that provides for abstracted views of the boundary of a molecule and the physical and chemical properties at this boundary. Our goal is to provide simplified visual representations of molecules such that scientists can rapidly assess the most significant features of their surfaces, even when drawn at a small size. Such abstracted views are useful for rapid browsing and comparison, but also to study molecules while unencumbered by small details. Once general notions are determined from the abstracted views, a scientist can make a focused examination of a more detailed representation.

Our abstraction mechanism processes the detailed information about the molecule to provide a visually simplified representation. The shape of the molecular surface is simplified, removing small details to better convey the basic shape. Significant shape features, such as clefts and pockets, become more prominent when the visual clutter of smaller features is removed. A comparison with other display methods is shown in Figure 1 and Figure 2. Additionally, our abstracted display is more amenable to stylized rendering that accentuates the shape, retains readability at lower resolutions allowing gallery displays, and allows for the use of surface markings to display other information.

Our technique also uses abstraction on properties other than shape. Scalar fields along the surface, such as electrostatic charge, are simplified for clarity, and other properties are displayed as symbols on the surface. These presentations allow significant features to be seen quickly and clearly.

We are motivated by an increased need for tools that enable quick and comparative visual analysis. Advances in structural biology, such as high-throughput crystallography and NMR spectroscopy, together with better prediction and simulation, have led to a marked increase in the number of proteins for which the three-dimensional atomic structure is known. Repositories for structural information, such as the PDB [2], have in turn grown dramatically. This wealth of information creates the need to look at large collections of molecules, requiring quick judgement.

Our technique, detailed in §3, relies on filtering of both the shape and surface fields to form abstractions. We describe how we adapt standard methods for shape smoothing by explicitly removing regions where the methods are likely to fail. §3.2 describes how surface decals can be used to display removed features, as well as other information. §3.3 describes the methods we use for more effectively portraying scalar fields on the surfaces, and §3.4 describes how stylized rendering techniques are used.

1.1 Background

Because proteins predominantly interact with other molecules non-covalently, their atomic forces effectively create shells that must fit together. The metaphor of a "lock and key" dates back over a century — [33] cites an 1894 paper by Fisher. While flexible water balloons or bean bags may be more appropriate metaphors than rigid pieces of metal, the central property governing how proteins interact is the non-penetration of their boundaries. Because the forces fall off rapidly with distance, treating the boundary as a surface (in the mathematical sense) is appropriate.

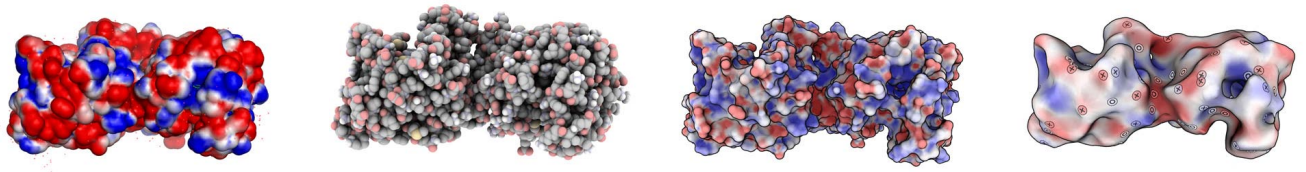
The definition of this geometric boundary surface is varied. The pioneering work of Richards (see [8] for a historical survey) introduced functional notions of these surfaces. The notion of solvent accessibility leads to *molecular surfaces* that provide a smooth description of the boundary. The molecular (solvent excluded) surface is the surface that a spherical probe can contact without intersecting the molecule. Connolly provided practical methods for sampling these surfaces [6, 7], which have subsequently been refined in both efficiency and quality [3, 29, 39]. Even more recent works provide methods, such as in [5], that can produce high quality representations of these surfaces.

The geometric molecular surface is only one of many properties that contribute to protein interactions. There is considerable evidence that electrostatics play a critical role in protein-protein interactions (see [32] for a survey of early work). Other properties that influence interactions include hydrogen bonds (and a molecule's ability to form them), polarity, and hydrophobicity. Because these properties fall off rapidly with distance, it is reasonable to consider them as scalar fields on the molecular surface when studying molecular interactions.

Molecules are continually in motion. In proteins, the motions range

• *Authors are with the Department of Computer Sciences, University of Wisconsin, Madison, E-mail: gregc@cs.wisc.edu gleicher@cs.wisc.edu.*

Manuscript received 31 March 2007; accepted 1 August 2007; posted online 2 November 2007.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.



(a) Molecular surface with charge (b) Qutemol (c) Stylized display (d) Abstracted with our method

Fig. 1. Depictions of the surface and electrostatic charge distribution of Adenylate Kinase (1ANK). The standard approach (a), drawn with Pymol [11], shows the molecular surface pseudo-colored (from red to blue) with the charge distribution. Qutemol [34] (b), applies stylized lighting to a space filling representation. (c) applies stylized shading directly to the molecular surface. (d) shows our abstracted surface depicted with stylized rendering. This molecule has binding partners that fit into channels formed in each of its lobes. From this view, one such channel should be visible in the center of the right lobe. This is made more readily visible by abstraction.

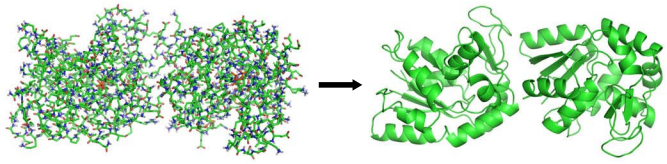


Fig. 2. A ball-and-stick representation (left) of adenylate kinase (Figure 1) contains too much information to be easily understandable. For larger scale views, biologists use abstracted representations such as ribbon diagrams (right). Such abstractions show major internal features of the molecule but do not convey the external surface.

from thermal vibrations to large conformational changes. This incessant movement, coupled with uncertainty in the measurement of the atomic coordinates, means that any static configuration of a molecule is merely a snapshot of its possible state. Therefore, smaller details of the shape and other fields of the molecule are of reduced significance as they are likely changing continuously.

2 Related Work

2.1 Molecular Visualization

Because of the importance of molecular shape, structural biologists have depended on visual tools from the beginning. Visual tools predate computers and continue to be developed to this day (see [8, 35, 16] for historically oriented surveys). Current state-of-the-art systems, such as Chimera [27], PyMol[11], and their competitors, provide large feature sets giving many options for the display of molecules.

Any visualization of a molecule necessarily involves some degree of abstraction. The field has developed a range of visual representations that provide different levels of abstraction; see [16] for a survey. For showing the internal composition of a molecule, many abstractions exist ranging from models that show every atom and bond, such as a ball and stick model, to highly abstracted representations, such as ribbon diagrams (Figure 2). These highly abstracted diagrams are valuable for providing a summary of a large molecule. However, because they do not indicate the "exterior" shape of the molecule, they provide less help in studying how the molecule would interact with others. Our methods provide such abstract representations for external properties.

There are two primary ways for showing the exterior shape of a molecule: space filling diagrams, where each atom is drawn as a solid sphere, and solvent-excluded surfaces, or molecular surfaces. Both views provide the shape of the molecule, however they provide it with a large amount of atomic-scale detail. Such detail is problematic as it can obscure larger scale phenomena and hinder effective portrayal of the shape. Raising the size of the probe sphere leads to molecular surfaces that exclude smaller crevices but may lose important pockets without discarding distracting bumps (Figure 3). Our methods can retain important features while reducing distracting detail.

Surface simplification (see [23] for a survey) creates approximate models with fewer polygons. These methods are useful in improving efficiency while preserving the appearance. Simplification is an essential part of large molecule surface display [13, 28]. In contrast, our approach seeks to alter the appearance to be more abstract, and does not necessarily provide a performance benefit, although abstracted surfaces are amenable to simplification. [13] applies smoothing, similar to our approach, to reduce the blocky appearance of coarse models of large molecules.

Display of other spatio-physico-chemical properties by color coding molecular surfaces became common as soon as surface representations were readily available. An early example was GRASP [24], which showed electrostatic potentials on surfaces. [4] unfolded the surfaces to better show their property distributions. Our methods provide abstracted display of these properties as well as molecular shape.

Work on displaying molecular motion shows the uncertainty in molecular shape. [21] shows uncertainty and vibrational motion by blurring standard representations, and [31] clusters states to provide visual representations of ranges of conformations. [18] uses a combination of point-based rendering and random displacement to convey surface uncertainty in volumetric data. While our work does not explicitly depict motion, it does convey a sense of uncertainty through the lack of detail.

The work of biochemist and artist David Goodsell inspires us by showing the merit of using artistically stylized depictions of molecules. His stunning figures require considerable artistic talent and effort to create. In [15] he describes a system for image processing molecular graphics that simulates a black-and-white line-art look, but which makes no abstraction of the shape.

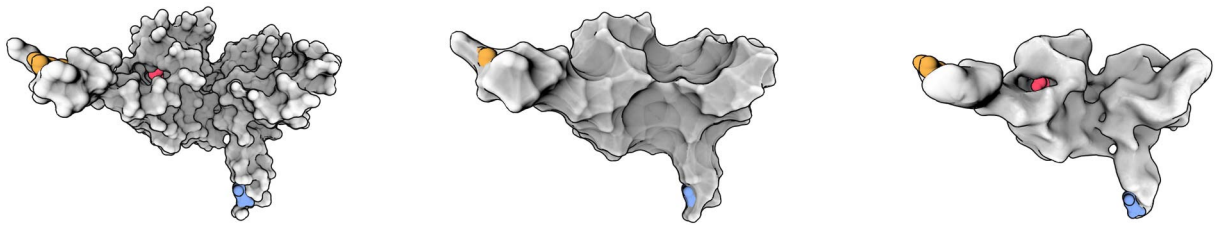
2.2 Conveying Shape

The complexity of molecular shape is hard to convey, as some spatial cues such as size and familiarity are not applicable. To help assist in shape comprehension, biochemists often rely on motion and stereo display to enhance more standard graphics cues, such as lighting, fog, and depth-cueing.

Texture can be an effective cue for aiding in shape perception [17]. To date, it has not been applied in large-molecule visualizations as the molecular surfaces or space filling views are ill-suited for texturing. Our abstracted views make use of texture.

Clever lighting design can greatly enhance the perception of local shape [22]. To provide for better comprehension of global shape, molecular visualizations have recently begun to use both global illumination and stylized shading. QuteMol [34] demonstrates the effectiveness of these techniques and shows how they can be implemented efficiently in hardware. However, QuteMol does not provide for shape abstraction; it provides only all-atom displays or space-filling diagrams. By providing abstracted surfaces, our approach can more easily apply a range of stylized depiction effects.

[26] shows how identified features can be better presented using stylized display. We apply this idea in a new domain and extend it by explicitly removing the features and only portraying them in a stylized fashion.



(a) Original, 1.5 angstrom probe (b) With a 4 angstrom probe (c) Abstraction of original mesh

Fig. 3. A demonstration of how using a larger probe size, while resulting in a slightly smoother mesh, will destroy fine details. Bright yellow, blue and red surfaces denote ligands, included to emphasize the important pockets. In particular, channels containing important ligands are completely removed in (b), along with other structural detail. Our method (c) preserves these features.

3 Abstracted Surfaces

Our abstraction process removes small details from molecular surfaces and their associated properties. These details are unlikely to be biologically significant, but will certainly detract from a viewer's ability to interpret larger patterns.

To create an abstracted representation of a molecule, our approach takes as input a triangle mesh of the molecular surface as well as information about the properties of the molecule. Our implementation uses external tools to create these inputs from PDB files. For the figures in this paper, MSMS[29] was used to create surface meshes and APBS[1] was used to compute electrostatic charge. The triangle mesh must be sampled finely enough to appear smooth at the scale of interest, but need not be uniform. Properties are associated with mesh vertices; scalar fields are sampled at these points before abstraction.

The primary step of abstraction is to remove small details in shape by smoothing the mesh. The choice of the scale of "small" is chosen to be smaller than a residue, but larger than an atom. Our implementation uses Taubin's filter [36, 37] as it is efficient and sufficiently volume-preserving. Local operations are performed on vertices lying in disc surrounding a given point, with importance falling off with the inverse of distance, as given by Fujiwara [12]. We have emprically determined the filter parameters $\lambda = .8$ and $\mu = -.87$, disc radius 4 angstroms, and 10 iterations to correspond to the desired feature size to remove. These parameters are a function of the scale at which we are interested in studying (i.e. larger than atomic scale interactions), not of particular molecules. All examples use the same parameter values.

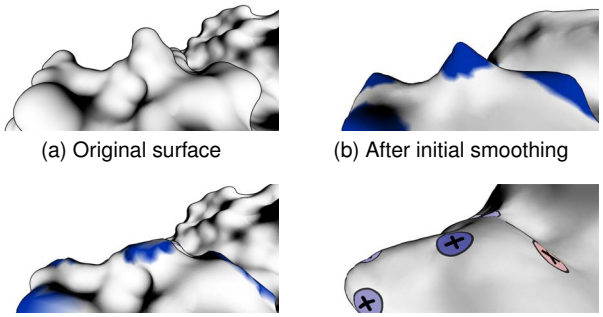
The smoothing process creates a potential problem: "Mid-sized" features that are larger than what is removed reliably by the filter are often distorted by it. These features, such as peaks formed by protruding chemical groups and small divots, may be biologically significant. While a more sophisticated filtering mechanism might better preserve them, even undistorted, these features are undesirable for abstraction as they are still difficult to portray.

To handle mid-sized features, we apply a different strategy shown in Figure 4. Our approach identifies the features and removes them from the mesh, leaving a smoother surface. However, it remembers that a feature was removed and depicts that feature using a surface marking. This approach has the advantages that it avoids artifacts from filtering and provides more control over how features are displayed. Small surface markings are better for abstracted representations than small bumps and divots because they do not detract from the overall shape and are visible from a wide range of viewpoints (see Figure 11).

3.1 Removing Mid-Sized Features

Mid-sized features are identified and removed. At present, our methods identify bumps and bowls that are large enough to be potentially interesting, but small enough to be problematic for smoothing. Similar approaches could be applied for other shape features, such as ridges and valleys.

Our process for removing bumps and bowls consists of several steps, illustrated in Figure 4. Features are identified by finding points of that have high curvature after smoothing, which is indicative of a filtering artifact. An initial round of smoothing is applied specifically



(a) Original surface (b) After initial smoothing

(c) With bumps removed (d) Abstracted result

Fig. 4. The molecular surface (a) is first smoothed, then "mid-sized" features are identified (b). Those features are removed from the original surface (c), then smoothing is applied and decals are used to represent the removed features (d).

for feature detection. Vertices whose curvature are outliers are chosen as features. The system computes the 10th percentile of the absolute value of principle curvatures ($\kappa 1$ and $\kappa 2$) for all vertices over the mesh.

Vertices are chosen to be outliers if their Gaussian curvature ($K = \kappa 1 \kappa 2$) is greater than P times the square of the 10th percentile principle curvature. Empirically, we have chosen $P = 30$ to provide a good balance between mesh smoothness and overly aggressive feature removal. The exact value of this parameter does not matter since precise identification is unimportant; an excess or missed point is likely to be grouped with another point in a later stage.

For each of these seed vertices, our system constructs a group containing other vertices within 2.5 angstroms along the surface. This distance was empirically found to correspond to the approximate size of individual mid-size features. If groups overlap, then it is likely that they are larger aggregates of individual features on the original mesh, so the process repeatedly merges groups until no overlaps are found.

This step results in a collection of patches on the surface representing mid-sized features. These regions are then removed from the original, unsmoothed mesh. To "sand them off," our system removes the majority of the vertices in the region and simultaneously "deflates" those that remain. To accomplish both, it first sorts the vertices according to how far away they are from their nearest seed vertex. It then takes the closest 80% and removes them, one by one, by edge contracting each with its closest neighbor in the graph, provided this contraction doesn't cause topological problems. This ensures that smaller triangles will be removed first, and also that vertices will be removed top-down, in the case of peaks, or bottom-up for divots.

The edge contraction process produces a mesh with its mid-sized features "sanded off" but the removal process often negatively impacts mesh connectivity, leaving many high-order vertices. Our system performs edge flips to improve the mesh by identifying high order vertices and for each one flipping its outgoing edge that is connected the highest order neighbor. It also finds extremely low-order vertices and contracts the outgoing edge connecting to their lowest order neighbor.

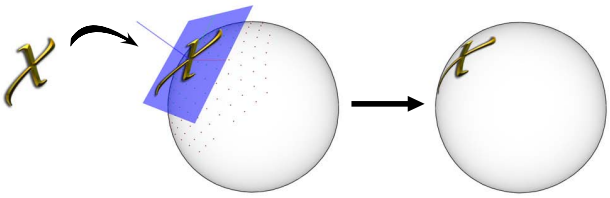


Fig. 5. We texture the surface using local parameterizations generated using an exponential map. First, a plane is constructed tangent to the desired position of the texture. Next, points surrounding that point (here in dark red) are mapped to that plane. Finally, the texture is placed on the surface according to that map.

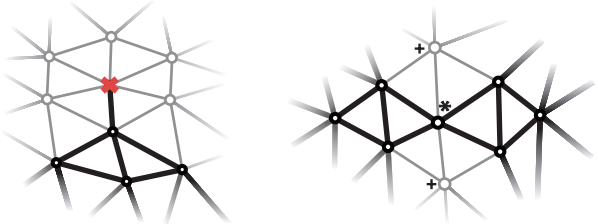


Fig. 6. At left, a vertex in the patch has only one other neighbor that is also in the patch. We remove these. At right, a vertex, denoted by a '*', joins two otherwise locally disconnected sets. We will add its neighbors, denoted by a '+', to the patch.

After these methods remove mid-sized features, the resulting surface is smoothed.

3.2 Decaling

We would like to create surface markings that are independent of the underlying triangulation. Otherwise, a coarse or uneven triangulation might lead to jagged, irregular shaped markings.

Texture mapping provides for surface markings independent of triangulation, but requires a parameterization of the surface to provide texture coordinates. The molecular surfaces are difficult to parameterize globally. We apply the approach of [30] to place textures on regions of the surface. Their approach creates a local parameterization of a region of the surface. This approach works well for our needs because our abstracted surface is relatively smooth, and because the markings we wish to apply are local.

3.2.1 Decal Parameterization

[30] use a discrete exponential map to create a local parameterization of a surface in the neighborhood of a point. Exponential maps take a point on the surface and map the surface surrounding that point to its tangent plane (see Figure 5), in a manner that yields mappings that preserve distances well. That plane serves as a local parameterization of the surface, and can be used to apply a texture with minimal distortion.

The methods of [30] were presented to support interactive decal placement. To apply it within our molecular abstraction process, we must automate the process of choosing the seed point, and of limiting the mapped region. These issues are challenging because poor choices can lead to parameterizations that distort the textures as they get further from the seed point.

To solve these problems our system attempts to locate an ideal starting vertex within the patch. Two competing goals intersect here: this vertex should lie as close as possible to the center of the region it represents, and also the normals on the surface should deviate as little as possible from its normal. This latter property is much more important to the overall quality of the parameterization, so our system first removes from consideration any vertices where it doesn't hold (i.e. $N_{vertex} \cdot N_{plane} < .05$).

If all vertices are removed, then the patch cannot form a good parameterization, and so that patch is not shown. Otherwise, the starting vertex is picked that has minimal distance to its most remote neighbor, which most often is a vertex lying roughly in the center of the patch.

3.2.2 Choosing Decal Placement

We consider two types of markings: fixed sized glyphs centered at a point, and arbitrary shaped regions. The former are used in our system to display symbols, such as circles and checks, to denote various features on the surface. To create a glyph decal, the point position is used as the seed for creating the parameterization.

Regions are represented as a subset of the mesh vertices. To create a decal corresponding to a region on the surface, our approach selects the best vertex (using the criteria in §3.2.1) and builds a parameterization around it. This parameterization determines where each of the vertices in the region lie in the texture plane, providing a 2D mesh that can be drawn on that plane. This patch is drawn to a texture such that the region outside of it is made transparent with alpha blending.

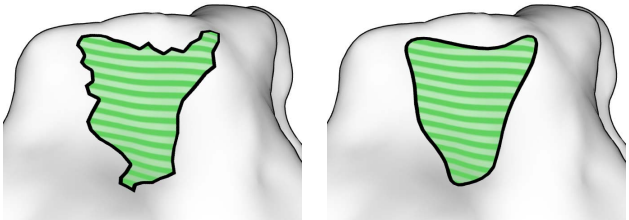


Fig. 7. At left, a patch before boundary smoothing. Nodes on the boundary are placed directly on vertices on the mesh, leaving a jagged exterior. At right, after smoothing.

The shape of a feature may have been distorted by both the filtering operations to create the smooth mesh and the mapping process, which may lead to patches with small holes, disconnected or poorly-connected vertices, and a jagged boundary. Removing these artifacts leads to abstracted markings that not only prevent problems in display, but also fit better into an abstracted representation and dispel any illusion that the fine details of the patch boundary are significant.

Our system abstracts patches in a number of steps. First, it applies standard binary image processing operations adapted to the non-uniform lattice of the 2D mesh. We use morphological operations [14] to remove outlying points and fill in small niches and holes. Dilation and erosion operators are defined based on the neighbors of a vertex. One step of dilation expands the patch out to include all immediate neighbors of the outermost vertices, while one step of erosion contracts the patch to remove all outermost vertices.

Rather than defining larger structuring elements, these immediate connectivity operators are applied repeatedly. We use 4 iterations of the close operator (dilation followed by erosion) to provide a good balance of problem removal and shape preservation.

Morphological operators may leave thin threads and bridges, as shown in Figure 6. These are removed by eliminating vertices with only one connected neighbor, and by expanding the patch around bridge vertices, which are defined as any vertex in the patch that has at least two neighbors not in the patch, and which do not themselves share a neighbor that is not in the patch.

After these cleaning steps, our system then finds all closed loops that lie on the border of the patch. This boundary is then smoothed (in the 2D map) by applying a low-pass filter to the 2D positions in the chains. This boundary is drawn with a stroke around its edge and the enclosed region filled, either with a flat color or with a texture defined over the plane. See Figure 7 for an example.

3.2.3 Using Decals

Our system uses decals to present information about the molecule in several ways. Because decals are semi-transparent, they overlay nicely on one another. However, displaying too much information may lead to clutter, so our system can optionally disable certain types of de-

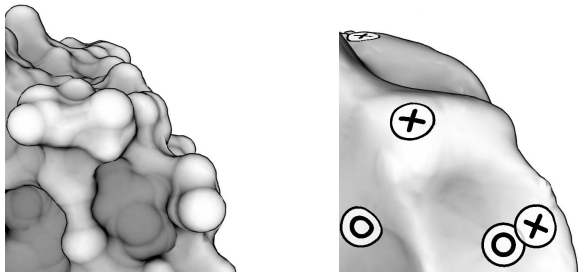


Fig. 8. Left: surface containing peaks and bowls. Right: same surface abstracted; peaks replaced by an X decal and bowls replaced by an O.

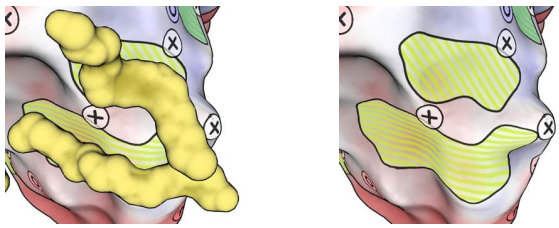


Fig. 9. At left, surface features are obscured by binding ligands. At right, projecting each ligand's location onto the surface allows simultaneous viewing of both ligand location and underlying surface properties.

cals, if desired. Decal positions can be determined from a number of tools, or can be provided manually for annotation. New methods for identifying features to mark can be easily added to our system.

For specific positional features, such as the location of hydrogen bond acceptors, our system chooses a single position on the surface and places a symbolic decal like the X in Figure 5. A surface point near an internal feature, such as an atom center, is chosen somewhat arbitrarily as small differences in positions are not important in the abstracted representation.

Our approach uses decals to indicate the mid-sized features removed in §3.1. While the set of vertices in the feature that remain after the removal process could be used to denote a region, our experience is that after contraction and smoothing, this patch bears little relationship in shape that of the removed feature. Therefore, our system instead uses a circular symbol of fixed size (1.5 angstroms in radius), as a circle doesn't imply anything (for better or worse) about the original shape. Glyphs within the circles are used to differentiate peaks from bowls. Examples can be seen in Figure 8.

Our system also uses decals to indicate larger regions corresponding to other information that is known about the molecule. Biologists use a myriad of tools to attempt to locate biologically significant areas on a molecule's surface. When binding partners are known, regions of the surface near ligands can be marked. This representation makes visible the portion of the surface involved in the interactions (Figure 9). The output of region detectors, such as pocket finders, can also be displayed this way. Our system presently includes an an implementation of Ligsite [19] to identify potential pockets. The output of these detectors is noisy, so before constructing decals, small or low-confidence regions are removed to avoid clutter, and excessively large regions are also culled because they are usually errors from the pocket finder and are problematic for the exponential map creator.

The system displays the output of different region detectors using different patterns for each, allowing multiple features to be shown simultaneously and compared (Figure 10).

3.3 Abstracting Surface Fields

Many important properties beyond the atomic forces that form the molecular surfaces, such as electrostatic charge and hydrophobicity are typically represented as scalar fields on the molecular surface and displayed using pseudo-coloring. Such properties suffer from the same profusion of small details as the shape itself, with similar issues in comprehension and display at small size.

Therefore, we abstract the scalar fields on the surface of the molecule. The scalar properties are attached to vertices before any simplification. This is particularly important because the true geometry determines the value (i.e. the position used to sample a volumetric scalar field such as electrostatic charge). Simplification should account for the real geometry in finding features (such as regions) in the scalar field as property values attached to vertices can be moved by the shape simplification. While this will distort the shape of the field, significant features of the field, such as regions of large magnitude, will remain with approximately the same shape and value.

The field abstraction process aims to remove small, less relevant details but to also preserve the coherent regions where the field has a definite value. Therefore, we apply a boundary-preserving low-pass filter to the surface scalar field. Specifically, we adapt the bilateral filter[38] to the irregular lattice of the triangle mesh.

As in the image case, the bilateral filter computes a new value at a vertex by taking a weighted average of the other vertices in its neighborhood, where these weights are determined by using both the spatial and value differences. We achieve larger kernel sizes by iterated application of smaller ones.

The general formula for our bilateral filter at iteration i for vertex v with value $val_i(v)$, where $d(v,w)$ represents Euclidean distance from vertex v to w, and $N(v)$ represents the set of v's neighbors, is:

$$val_{i+1}(v) = \frac{1}{k_i(v)} \cdot \sum_{w \in N(v)} val_i(w) \cdot c_i(w,v) \tag{1}$$

$$c_i(w,v) = e^{-\frac{d(v,w)^2}{2}} \cdot e^{\frac{(val_i(v)-val_i(w))^2}{2 \cdot \sigma_i^2}}, \quad k_i(v) = \sum_{w \in N(v)} c_i(w,v) \tag{2}$$

Thus, our system applies a Gaussian filter for both distance and value-similarity weights. For the latter, though, we have found that by using a larger kernel (σ_i) for the first few iterations and then progressively reducing it at later iterations, we can prevent areas of uniform value from completely diffusing into areas of different value.

Our method to adapt kernel size is to use a kernel proportional to the standard deviation over the values at each vertex. Since the variance of the values themselves will be converging as a result of smoothing, this results in a progressively smaller sigma, which in turn gives higher weight to the value-similarity kernel.

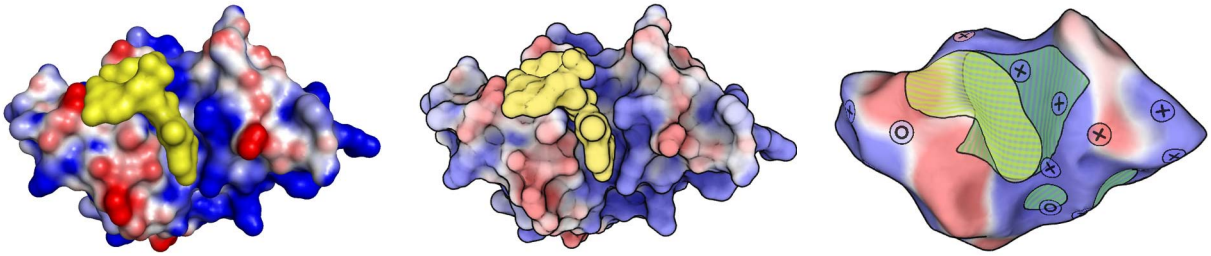
Our system iterates until asymptotic convergence, which is reached when the average over all vertices v of $val_{i+1}(v) - val_i(v)$ is less than ε. In our experiments, $\varepsilon = .005$.

When a mid-sized feature is removed, scalar field information contained in that feature is lost. This information is potentially important: for example, highly charged protrusions may be biologically significant. To preserve this important field information, the removal process associates a field value with the decal representing a removed feature, determined by averaging the values of that feature's vertices.

3.4 Display of Abstracted Surfaces

The resulting abstracted surfaces still provides a shape display problem. The visual presentations must not impede the use of the surface markings indicating small shape features. Also, we prefer a visual style consistent with the abstraction, rather than the "realistic" shiny plastic more commonly used to display molecular surfaces. Therefore, our system's primary display is stylized. Qutemol [34] showed the utility of stylized shading for molecular depiction. We apply several of their concepts to molecular surfaces. Our system also applies the stylized rendering to non-abstracted surface models (seen in many figures throughout the paper).

To enhance shape portrayal, we apply per-pixel silhouette shading [20], which sets $brightness = 1 - p^{\frac{cos^{-1}(n_z)}{\pi/2}}$, where n_z is the z component of the surface normal and p is a tunable constant that we have set to .3 in our figures. This shading sets pixels on faces that orient directly toward the viewer at maximal brightness, with decreasing brightness as the face normal orients away.



(a) Pymol (b) Stylized (b) Abstracted

Fig. 10. An example of Bullfrog Ribonuclease (1M07) before and after our abstraction process. The green striped areas represent parts of the surface that were identified as putative ligand binding sites. The yellow, ligand shadows, or areas of the surface nearest to known ligand locations.

Ambient occlusion (AO) lighting [20] is applied as it accentuates global shape. As pointed out by [34], the regions made darker by ambient occlusion because of lower lighting accessibility are related to the regions with lower chemical accessibility. Interior points of clefts and pockets are made darker. Our implementation of AO uses the graphics hardware to sample light directions.

As a final step, our system strokes along contours of the mesh, which can be defined as those edges that border both a front-facing and a back-facing face. This not only enhances shape perception, but gives a stylized look that gives a constant reminder of the degree of abstraction in the representation. The smoothness of abstracted surfaces makes more sophisticated contouring methods unnecessary. Experiments with Suggestive Contours [10, 9], show that they add few contours beyond the simple ones on our surfaces, and that these additional contours were typically small enough to be difficult to notice.

As in traditional molecular surface display, scalar fields are indicated on the surface by pseudo-coloring. For the examples shown in this paper, electrostatic charge is displayed using the red to blue scale that is commonly used. Colors for other decals are chosen such that they can be seen when overlayed on these colors.

4 Results

We have implemented our molecular surface abstraction techniques in our bespoke visualization testbed that runs under Windows on PCs. Our system relies on standard tools for computing the surfaces and other properties. For all examples in this paper, we use MSMS [29] to generate the molecular surface meshes, APBS [1] to compute electrostatic charge distributions, and an implementation of Ligsite [19] to identify putative binding pockets.

We emphasize that abstraction is not the same as changing the probe size. Larger probe sizes will fill in crevices that may be important pockets, while leaving bumpy details that detract from comprehension as shown in Figure 3. Indeed, abstraction can be applied to surfaces generated with any probe size. All examples in this paper were generated using a probe size of 1.5 Angstroms, (except Figure 3b).

Figures throughout the paper show the results of our methods applied to various molecules. Figure 10 shows how important aspects of the molecule are made clear by abstraction. Figure 12 provides a gallery of examples.

Major shape features, such as pockets and clefts, become very clear on abstracted surfaces because there are fewer small details to distract the viewer, and the smoothness allows silhouette shading, contouring and ambient occlusion lighting to emphasize the shape.

The asymptotic complexity of our method scales linearly with the number of vertices of the input mesh, which scales (at worst) linearly with the number of atoms in the molecule. This was confirmed empirically in our performance evaluation.

To assess performance of our prototype, we selected 60 proteins of various sizes from the Astex test set [25]. When determining the timings for the abstraction process, we consider smoothing, decal construction, and surface field relaxation; but not the preprocesses to determine the initial mesh, electrostatic charges, or binding pockets. We also exclude the time required to load data from disk and compute ambient occlusion lighting. Timings were performed on a PC with an Athlon 4400 CPU, 2GB of RAM, and NVidia 7900GT graphics.

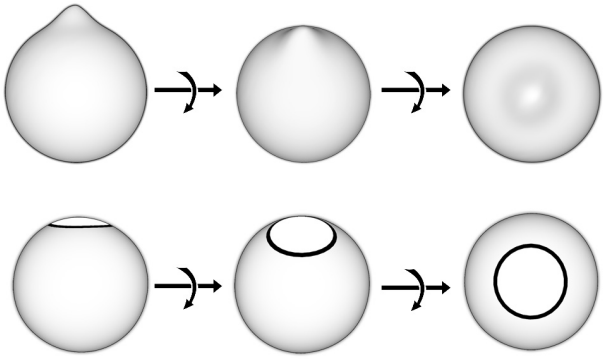


Fig. 11. A sphere with a bump facing upward, rotated forward $45°$, and at right $90°$. Top: traditional geometric display. Bottom: our abstracted view makes the location more apparent.

On our test set, the time needed to perform abstraction ranged between 7 and 113 seconds. These correspond to the smallest molecule in the set (1CBS, 137 residues, 1092 atoms) and the largest (1CX2, 2200, 21764). The expected linear performance scaling was observed.

Once abstracted, the models are displayed in real time using the graphics hardware. Our system maintained at least 30fps on all molecules in the test set. Our ambient occlusion precomputation implementation is not realtime (about 10 seconds for the largest molecule in the test set), but is computed in parallel and its results continuously displayed on the model's surface. The abstracted model, including its ambient occlusion lighting, could be precomputed and stored.

5 Discussion

The symbolic display of smaller features has a number of advantages. It makes shape features more readily apparent from a wider range of viewing directions (Figure 11). Symbols are visible in static displays, while shape features often are only obvious in regular displays when the object is moving. At small sizes, sampling issues may make small geometric features difficult to display. With surface textures, texture sampling hardware can perform sampling using mip-mapping, and decals can be omitted at very small sizes.

Displaying symbols as decals on the surface provides a mechanism for indicating a variety of properties about the underlying molecule. Such decaling would be difficult on non-abstracted surfaces: their non-smoothness would make parameterization difficult, and the small features would obscure the symbols with clutter and occlusion. Because the symbolic display is imprecise, they fit in better with the abstracted surfaces as abstracted surfaces imply a lack of positional precision.

Abstracted surfaces are also more amenable to mesh decimation for performance enhancement. This can be particularly important when

displaying multiple molecules simultaneously in an interactive system.

Together, these features of abstracted surfaces suggest that they will be useful in creating gallery views that allow a number of molecules to be shown simultaneously for browsing and comparison. Without a reduction in detail, a gallery of surfaces would be overwhelming. In a gallery display, the size and resolution of each molecule's depiction is limited. There is also less opportunity for interactively rotating each molecule to find views that show shape features.

Many of the the limitations of our initial prototype should be addressed. Better shape feature finding would allow us to translate more of the shape into surface symbols to provide further abstraction. Incorporating dynamics would help us better target unstable features for removal. Better decal design should include textures that help convey shape more precisely. And methods are required to apply decals correctly on decimated meshes and patches with non-planar topology. To be truly useful, abstracted surfaces will need to be integrated into existing tools and workflows.

Our abstraction has been applied to study features at a specific scale. Exploring other scales would require retuning the methods, and possibly designing a new set of feature detectors. Studying molecules at scales much different than the atomic level interactions we consider, such as macro-molecular assemblies, would provide more challenges, including performance.

The most important step for our work is to assess how effective these representations are for scientists. To date, our testing of abstracted molecular surfaces with our biochemist collaborators has been limited, so our observations are anecdotal. In all cases, their initial reactions were extremely positive. They immediately appreciated the simplified views. On molecules familiar to them the views matched what they "expected" them to look like. In several cases they would make comments like "I never noticed that before, I wonder ..." which is particularly encouraging as it implies a new way of looking at things might lead to new hypotheses. In viewing unfamiliar molecules, they were able to spot important functional areas quickly.

Acknowledgements

We thank George Phillips Jr. and his lab for their guidance in this project. Aaron Bryden helped with software pipeline issues and PyMol figures, and Cody Robson contributed some of our mesh processing implementation. Nick Reiter provided molecules and feedback. Rachel Heck assisted with video production. Cipriano was supported by NIH training grant NLM-5T15LM007359.

References

[1] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. National Academy of Sciences*, 98(18):10037–10041, Aug 2001.

[2] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[3] J.-D. Boissonnat, O. Devillers, J. Duquesne, and M. Yvinec. Computing connolly surfaces. *J. Molecular Graphics*, 12(1):61–62, March 1994.

[4] M. Chapman. Mapping the surface properties of macromolecules. *Protein Science*, 2(3):459–469, 1993.

[5] H.-L. Cheng and X. Shi. Quality mesh generation for molecular skin surfaces using restricted union of balls. In *Proceedings IEEE Visualization Conference*, 2005.

[6] M. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612):709–13, August 19, 1983 1983.

[7] M. L. Connolly. The molecular surface package. *J. Molecular Graphics*, 11(2):139–141, June 1993.

[8] M. L. Connolly. Molecular surfaces: A review. Network Science online article http://www.netsci.org/Science/Compchem/feature14.html, 1996.

[9] D. DeCarlo, A. Finkelstein, and S. Rusinkiewicz. Interactive rendering of suggestive contours with temporal coherence. In *Proceedings NPAR*, pages 15–24,145, 2004.

[10] D. DeCarlo, A. Finkelstein, S. Rusinkiewicz, and A. Santella. Suggestive contours for conveying shape. In *Proceedings SIGGRAPH*, pages 848–855, 2003.

[11] W. DeLano. The PyMOL molecular graphics system, 2002.

[12] K. Fujiwara. Eigenvalues of laplacians on a closed riemannian manifold and its nets. *AMS*, pages 2585–2594, 1995.

[13] T. Goddard, C. Huang, and T. Ferrin. Software extensions to ucsf chimera for interactive visualization of large molecular assemblies. *Structure*, 13:473–482, March 2005.

[14] R. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2002.

[15] D. Goodsell and A. Olsen. Molecular illustration in black and white. *J. Molecular Graphics*, 10(4):235–40, 1992.

[16] D. S. Goodsell. Visual methods from atoms to cells. *Structure*, 13:347–354, March 2005.

[17] G. Gorla, V. Interrante, and G. Sapiro. Texture synthesis for 3d shape representation. *IEEE Transactions on Visualization and Computer Graphics*, 9(4):512–524, 2003.

[18] G. Grigoryan and P. Rheingans. Probabilistic Surfaces: Point Based Primitives to Show Surface Uncertainty. In *Proceedings IEEE Visualization*, 2002.

[19] M. Hendlich, F. Rippmann, and G. Barnickel. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Molecular Graphics and Modelling*, 15(6):359–363, 1997.

[20] H. Landis. Production ready global illumination. In Siggraph Course Notes, 2002.

[21] C. Lee and A. Varshney. Representing thermal vibrations and uncertainty in molecular surfaces. In *SPIE Conference on Visualization and Data Analysis*, 2002.

[22] C. H. Lee, X. Hao, and A. Varshney. Light Collages: Lighting Design for Effective Visualization. In *IEEE Visualization*, pages 281–288, 2004.

[23] D. P. Luebke. A developer's survey of polygonal simplification algorithms. *IEEE CG&A*, 21(3):24–35, 2001.

[24] A. Nicholls, R. Bharadwaj, and B. Honig. Grasp: Graphical representation and analysis of surface properties. In *37th Meeting of the Biophysical Society*, volume 64, page A166, 1993.

[25] J. W. Nissink, C. Murray, M. Hartshorn, M. Verdonk, J. Cole, and R. Taylor. A new test set for validating predictions of protein-ligand interaction. *Proteins: Structure, Function and Genetics*, 49(4):457–471, 2002.

[26] M. Pauly, R. Keiser, and M. Gross. Multi-scale Feature Extraction on Point-Sampled Surfaces. *Comp. Graphics Forum*, 22(3):281–289, 2003.

[27] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF Chimera - a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605 – 1612, 2004.

[28] M. F. Sanner. A component-based software environment for visualizing large macromolecular assemblies. *Structure*, 3(3):447–462, Mar 2005.

[29] M. F. Sanner, A. J. Olson, and J.-C. Spehner. Fast and robust computation of molecular surfaces. In *SCG '95: Proceedings of the eleventh annual symposium on Computational geometry*, pages 406–407, 1995.

[30] R. Schmidt, C. Grimm, and B. Wyvill. Interactive decal compositing with discrete exponential maps. *ACM Transactions on Graphics*, 25(3):603–613, 2006.

[31] J. Schmidt-Ehrenberg, D. Baum, and H. C. Hege. Visualizing dynamic molecular conformations. In *Proceedings IEEE Visualization*, pages 235–242, 2002.

[32] F. B. Sheinerman, R. Norel, and B. Honig. Electrostatic aspects of proteinprotein interactions (review). *Current Opinion in Structural Biology*, 10(2):153–159, 2000.

[33] G. R. Smith and M. J. E. Sternberg. Prediction of protein-protein interactions by docking methods (review). *Current Opinion in Structural Biology*, 12(1):28–35, Feb 2002.

[34] M. Tarini, P. Cignoni, and C. Montani. Ambient occlusion and edge cueing for enhancing real time molecular visualization. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):1237–1244, 2006.

[35] J. Tate. Molecular visualization. In P. Bourne and H. Weissig, editors, *Structural Bioinformatics*, chapter 23. Wiley-Liss, 2003.

[36] G. Taubin. A signal processing approach to fair surface design. In *Proceedings of SIGGRAPH 95*, pages 351–358, Aug. 1995.

[37] G. Taubin. Geometric signal processing on polygonal meshes. In *EUROGRAPHICS - State of the Art Reports*, 2000.

[38] C. Tomasi and R. Manduchi. Bilateral filtering for gray and color images. In *ICCV*, pages 839–846, 1998.

[39] A. Varshney and F. P. Brooks, Jr. Fast analytical computation of Richard's smooth molecular surface. In *Proceedings IEEE Visualization*, pages 300–307, 1993.

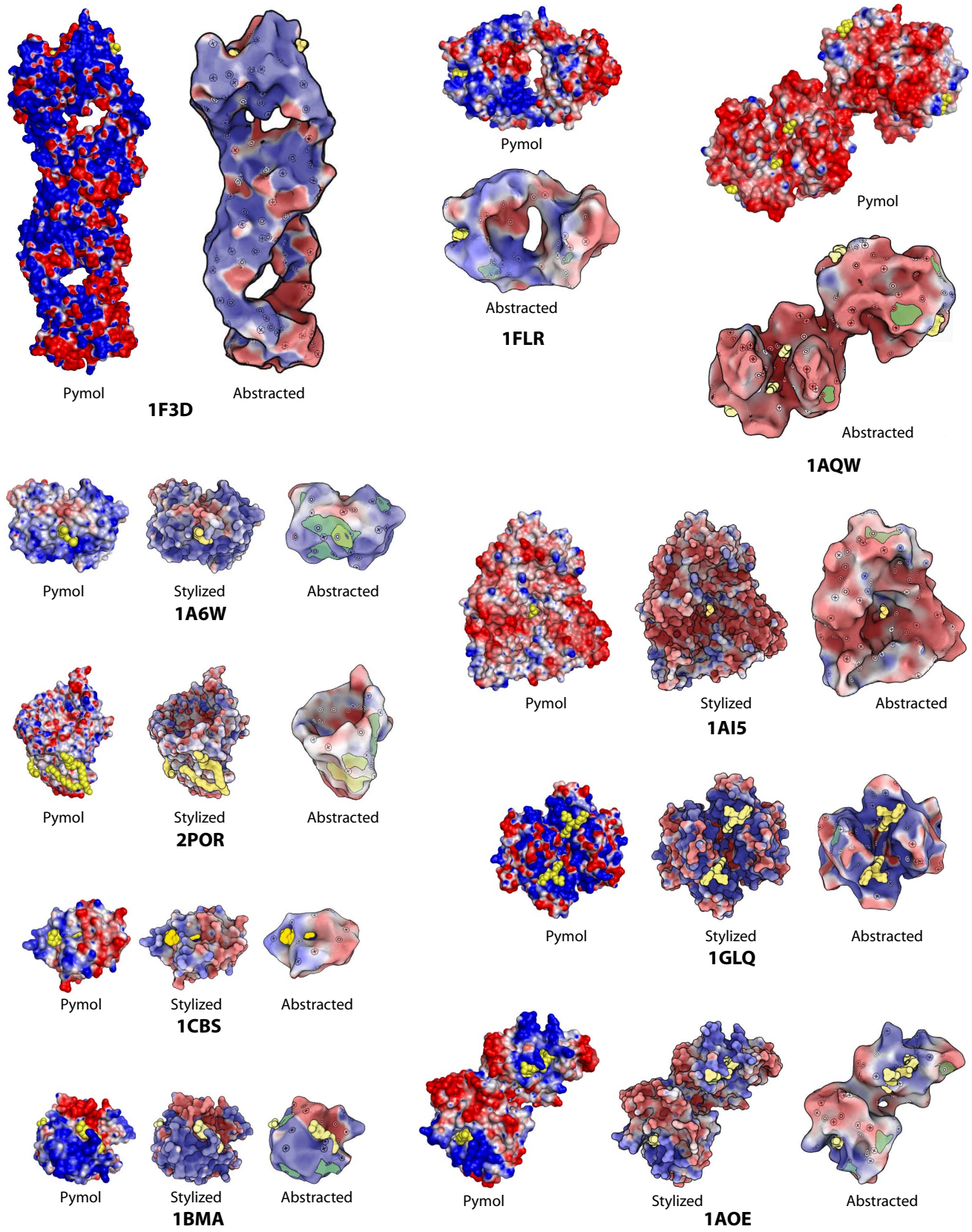


Fig. 12. A gallery of example proteins of various sizes, shown before and after abstraction. Traditional images, rendered with [11], show molecular surfaces for the proteins, and spheres for the ligands. Stylized images use our rendering techniques on non-abstracted surfaces.