

A Volumetric Method for Representing and Comparing Regions of Electrostatic Focusing in Molecular Structure

Brian Y. Chen^{*} and Debdas Paul
Department of Computer Science and Engineering
Lehigh University
19 Memorial Drive
Bethlehem, Pennsylvania
chen@cse.lehigh.edu

ABSTRACT

Algorithms for protein structure comparison employ diverse and effective representations of molecular shape. However, they do not generally represent the shape of the electrostatic potential field, except at the molecular surface. This approach neglects the geometry of the field on the outside of the molecular surface, where electrostatic focusing can play an important role in molecular recognition: Narrow clefts and grooves can partially shield charged atoms from the high dielectric solvent, enhancing potentials inside the cavity and projecting the lines of the electric field outwards from the cavity. This interplay between molecular shape and electrostatic potential is an essential means of recognition in many biomolecular systems. To leverage this phenomenon for more accurate protein structure comparison algorithms, this paper presents the first comparative representation of the region where focusing occurs. We first verified our representation in a case study of superoxide dismutase, where electrostatic focusing was first observed. Our method accurately identified the site where electrostatic focusing was established in the past. We then applied our representation to compare regions of electrostatic focusing with the positions of charged amino acids, to determine where they coincide. Over 866 protein-DNA complexes, our representations correctly detected the enrichment of arginines that contact regions of electrostatic focusing in the minor grooves of DNA. These results indicate that our novel methods precisely represent and accurately compare regions where electrostatic focusing occurs. They also describe a novel and elegant technique for seamlessly integrating molecular shape and electrostatic focusing into the same structure comparison framework.

Categories and Subject Descriptors

J.3 [Life and Medical Sciences]: Biology and Genetics

General Terms

Algorithms, Design, Verification

^{*}Corresponding Author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM-BCB '12, October 7-10, 2012, Orlando, FL, USA
Copyright 2012 ACM 978-1-4503-1670-5/12/10 ...\$15.00

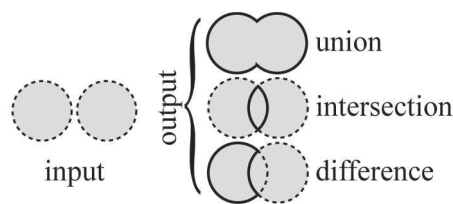


Figure 1: Boolean Set operations on three dimensional solids (left) can detect regions where two solids coincide (intersection), where either solid occupies space (union), or regions inside one solid and not inside the other (difference).

1. INTRODUCTION

Electrostatic interactions are a major contributor to molecular recognition in biological systems [17, 20, 21, 33–35, 41, 44, 49, 58]. Attractive and repulsive forces derived from charged atoms can be influential over considerable molecular distances, giving them a considerable impact on binding affinity [18, 56, 57, 59, 62] and especially on binding specificity [18, 37, 52] because of differential positioning and the energetic cost of displacing solvent [24, 31]. For algorithms that analyze and compare protein structures, precisely representing electrostatic influences is crucial for detecting influences on specificity or similarities in biochemical function (e.g. [26]).

One aspect of electrostatic potential fields not frequently represented in structure comparison algorithms is *electrostatic focusing* [20]. This effect occurs in narrow molecular cavities where it enhances potentials inside the cavity and focuses electric field lines into the high dielectric solvent. First observed in superoxide dismutase [20], this phenomenon is general, has often been observed broadly among proteins [39] and DNA [45]. Our paper presents the first algorithm for integrating the analysis of electrostatic focusing into the comparison of protein structures.

Specifically, we describe how to represent the region where electrostatic focusing occurs as a three dimensional solid. This representation is computationally identical to existing solid representations of binding site shape, which can be compared with Boolean set operations [11] (Figure 1). As a result, the shape of the focusing region can be compared independently, without considering atomic structure, or it

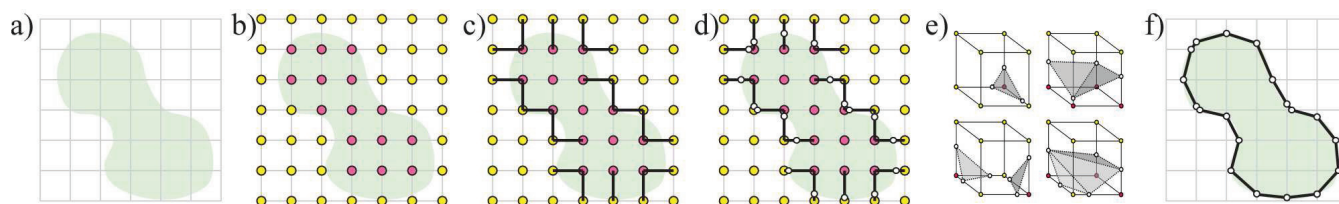


Figure 2: Representing the Focusing Region. (a) The lattice (grey), surrounding the focusing region (teal). (b) Interior points (red), exterior points (yellow). (c) Edges that connect an interior to an exterior point (bold) (d) boundary points (white). (e) Partial lookup table of triangle configurations. (f) Output triangles (dark lines) that approximate the boundary of the focusing region.

can be modified by atomic structure, to provide an integrated analysis. Given that influences on binding specificity can already be identified by solid representations of protein binding sites [7, 8], the integration of electrostatic focusing represents an opportunity to detect a second, nearly orthogonal range of influences on specificity.

In our results, we demonstrate the accuracy of our method for representing regions of electrostatic focusing in a case study on superoxide dismutase. We then verify our methods at a large scale on 866 protein-DNA complexes: Using Boolean set operations, we compare the position of the focusing region in the minor grooves of DNA with the placement of amino acids from proteins in complex with the DNA. Our results point to wide applications in protein engineering or for determining specific influences on molecular recognition.

2. RELATED WORK

Algorithms for protein structure comparison widely label elements of protein structure with information relating to electrostatic potential. This approach is most convenient for algorithms that represent atoms or amino acids as points in space [9, 36, 38, 47, 48, 50, 51], distance matrices [19] or as geometric graphs [16, 61], because label assignment is made to a discrete entity. Comparison algorithms of this nature build correspondences between atoms or amino acids with similar biological or chemical qualities. Once the correspondence is constructed, geometric similarity is measured using least-squares methods [22, 55]. In these representations, labeling affects comparison by changing the rules by which point-to-point correspondences are generated (e.g. [10]): They insist that geometric comparisons only be made between atoms or amino acids with similar electrostatic properties.

A second general approach to structure comparison is to represent proteins according to their solvent-accessible surface [13, 30], often referred to as the *molecular surface*. The surface itself is often described with triangular meshes [27, 46], three dimensional grids [29], alpha shapes [4, 14], or spherical harmonics [23, 25, 42]. In such cases, the electrostatic potential field can be represented at points on the molecular surface [27]. Surface points are crucial because they represent the field at interfaces that make contact with other molecules.

In all of these representations, electrostatic information is generally recorded in atom identities or charges, or as potentials on the molecular surface. These approaches do not represent the shape of the field in the space outside the pro-

tein, where it influences other molecules. For example, in superoxide dismutase, without accounting for electrostatic focusing, there is no “window” of electrostatic potential by which charged substrates can be attracted to the active site from long distances [28]. Existing methods cannot anyway represent this window, but after the enhancement caused by electrostatic focusing is accounted for, the presence of this window is obvious. The isopotential window can be easily visualized with software like GRASP [38].

This work provides the first evidence that electrostatic focusing can also be leveraged in the comparison of protein structures, once an adequate representation can be established. Our representation delineates a tight boundary around the region where potentials are enhanced, enabling the shape of the focusing region to be compared with Boolean set operations. Furthermore, unlike existing methods, our representation is entirely independent of molecular shape, enabling the focusing region to be represented outside the molecular surface.

3. METHODS

First, we describe our method for representing the region of electrostatic focusing. To evaluate this method on a large scale, we require the identification of amino acids in contact with the minor groove in DNA. Thus we present, second, our method for detecting which amino acids are in the minor groove. Finally, we describe the way we compare the region of electrostatic focusing with the positions of amino acids, and the structure of our experimental data set.

3.1 Generating the Focusing Region

Our approach, inspired by Marching Cubes [32], accepts as input a structure file from the Protein Data Bank (PDB) [3], a threshold K which defines the degree of focusing (used throughout this paper), and a resolution parameter (.5 Å in this work) that controls the geometric detail of the output. As output, our method generates a triangular mesh that bounds the region of electrostatic focusing.

The field of electrostatic potential surrounding the input molecule can be evaluated with uniform and nonuniform dielectric models. The potential field evaluated with uniform dielectric, Φ_U , represents the interior of the input molecule with the same high dielectric as the surrounding solvent. As a result, potentials in Φ_U are not influenced by the shape of the input molecule: no focusing occurs in Φ_U . In contrast, the potential field evaluated with a nonuniform dielectric, Φ_N , represents the interior of the input molecule

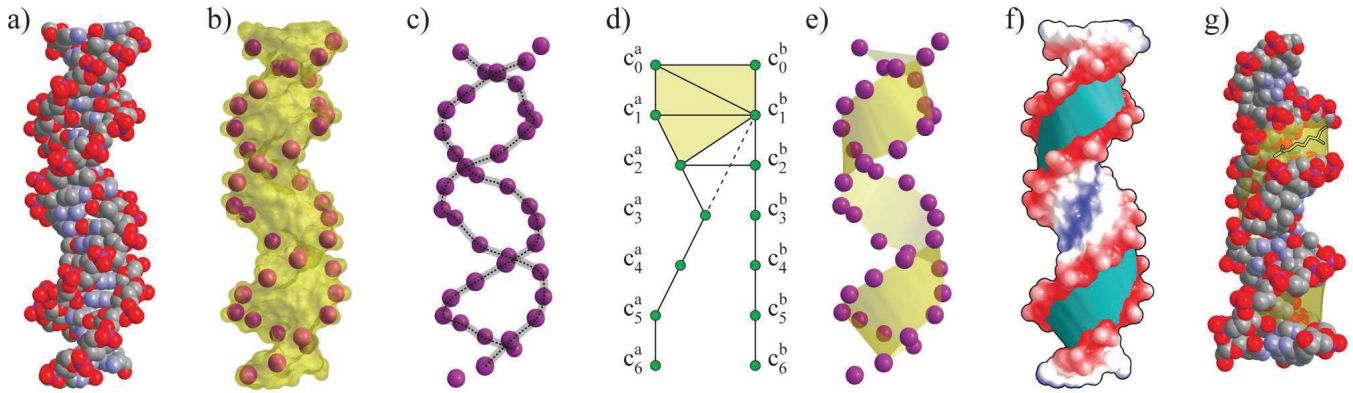


Figure 3: Finding amino acids in the minor groove. (a) Atoms from DNA chains in 1HLZ, from the PDB (spheres). (b) Molecular surface of DNA (transparent yellow), phosphates in DNA backbone (magenta). (c) Catmull-Rom spline (grey). (d) Diagram of triangle ribbon assembly, corner points (green circles), triangle edges (black lines). (e) Finished ribbon (transparent yellow) (f) Finished ribbon (teal). On the molecular surface of DNA (black outline), blue regions are distant, white regions intermediate, red regions are nearby the ribbon. (g) Atomic structure of DNA, rotated (spheres). Ribbon (transparent yellow), arginine intersecting the ribbon and the minor groove (yellow, black outline).

as having a low dielectric, causing focusing to occur. At any a given point in space p , we can evaluate $f(p)$, the degree of electrostatic focusing at p , with the expression $f(p) = |\Phi_U(p) - \Phi_N(p)|$.

We use DelPhi [43] to compute the two potential fields corresponding to uniform and nonuniform potential fields mentioned above. For the nonuniform dielectric model, we assign an intramolecular dielectric constant of 4 [53], and for the uniform model, we assign 79.9. We use a 0.145 M salt concentration and the AMBER [40] force field for partial charges and atomic radii. Three focusing steps are performed with Delphi on all structures.

Once we have computed both potential fields, we construct a polyhedral surface that approximates the three dimensional boundary of the focusing region. First, we create a three dimensional axis aligned cubic lattice that encloses the entire molecule (Fig. 2a). We describe this lattice in terms of *lattice points*, *lattice segments* and *lattice cubes*. Lattice points are placed incrementally along each axis, forming a grid. The incremental spacing is set by output resolution. Adjacent lattice points on the same axis form a lattice segment. 12 adjacent lattice segments form a lattice cube.

Second, we determine if each lattice point is inside the region of focusing. A point p is inside the region of focusing if $f(p) > K$, otherwise it is outside, as illustrated in Fig. 2b.

Third, we find all lattice segments that have one point inside and another point outside the region of focusing (Fig. 2c). On each segment, we find the point where $f(p) = K$ using linear interpolation. These “boundary points” approximate the boundary of the region of focusing (Fig. 2d).

In the final step, we approximate the surface bounding the region of focusing. For any lattice cube, each of its eight corners are either inside or outside the focusing region, creating up to 256 inside-outside permutations. We use a look up

table (e.g. Fig. 2e), detailed elsewhere [32], to connect each permutation to a triangular approximation of the boundary surface as it passes through the cube. With the boundary points computed above, we use the triangular approximation to create and store triangles corresponding to each lattice cube. The collection of all triangles from all cubes forms a surface that approximates the region of focusing (Fig. 2f).

3.2 Representing the minor groove of DNA

Given a protein-DNA complex as input, this section describes a method for detecting amino acids that enter the minor groove of the DNA. The overall approach is to build a ribbon that connects the phosphate atoms over the minor groove, and then to determine which amino acids intersect that ribbon.

Beginning with a PDB structure containing DNA in complex with protein, we begin by separating the DNA from the protein (Fig. 3a). We then use the find-pair program within X3DNA [63] to detect paired nucleotides within the complex. Unpaired nucleotides or those that lack a phosphate atom are ignored. The pairing enables us to divide the input into pairs of chains. For each chain, we identify the phosphate atom of each nucleotide (Fig. 3b). We use the phosphate atom as a *control point* in a Catmull-Rom spline [5].

Catmull-Rom splines are a class of piece-wise functions that specify the position of a point along a curve using a number of control points. For a Catmull-Rom spline $c(t)$,

$$c(t) = [1 \ t \ t^2 \ t^3] \begin{bmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{bmatrix} \begin{bmatrix} P_{i-2} \\ P_{i-1} \\ P_i \\ P_{i+1} \end{bmatrix}$$

where the spline is parametrized by the variable t and interpolates between the control points P_i . With this equation,

the segments at the end of each spline cannot be calculated, so we add a linearly extrapolated virtual phosphate beyond the end in each direction. The result is a spline with a continuous derivative (i.e. lacking in sharp corners) that fluidly approximates a path through the backbone phosphates (Fig. 3c).

Next, between each successive control point, we place five intermediate points along the spline. If there are k control points, combining the control points with these new intermediate points results in a total of $6k - 5$ *corner points*, designated c_0^a, c_1^a, \dots , for a spline a . We follow this procedure to generate splines for paired phosphate backbones in the input complex, which we refer to as splines a and b .

After generating corner points for both splines, we begin generating triangles between the splines as follows: We begin with the first corner points of each spline, c_0^a and c_0^b , on one end of the DNA. We mark c_0^a and c_0^b *used*. We then measure the distance d_a between points c_1^a and c_0^b and the distance d_b between c_0^a and c_1^b . If d_a is smaller than d_b , we designate c_1^a the *next point*, otherwise this designation is assigned to c_1^b . For example, in Fig. 3d, the distance between c_3^a and c_1^b is larger than the distance between c_2^a and c_2^b , causing us to select c_2^b as the next point. The next point defines a new triangle using the last two used points and the next point. We add the new triangle to a list of triangles that defines the ribbon, and then we also mark the next point used. This triangle generation process then repeats with the last two used points.

The result of the triangle generation process is a ribbon of edge-adjacent triangles (see Fig. 3e) that minimize the length of the edges that extend between the splines. This minimization causes the triangles to always cover the minor groove (e.g. Fig. 3f), except, in theory, if the DNA is in a highly nonstandard conformation. That effect was not observed in our data.

We define amino acids inside the DNA minor groove as having at least one atom that intersects the ribbon defined in Section 3.2. We detect atom-ribbon intersections by measuring the distance from the center of the atom to any triangle of the ribbon. If the distance is within the Van der Waals radius of the atom (based on this reference [6]), we say the triangle and the atom intersect (e.g. the arginine on Fig. 3g).

3.3 Comparing the focusing region

We use VASP [11] to compare the position of the focusing region in the minor grooves of DNA with the position of amino acids from proteins in complex with the DNA. As input, we begin with the region of focusing and the structure of an amino acid. First, we generate the molecular surface of the amino acid using the Trollbase library from GRASP2 [38]. Since molecular surfaces are closed surfaces, we can interpret them as three dimensional solids, enabling them to be compared with Boolean set operations. We then compute the Boolean intersection (e.g. Figure 1) between the region of focusing and the amino acid. The volume of the intersection region measures how much the intersection region and the amino acid coincide in space.

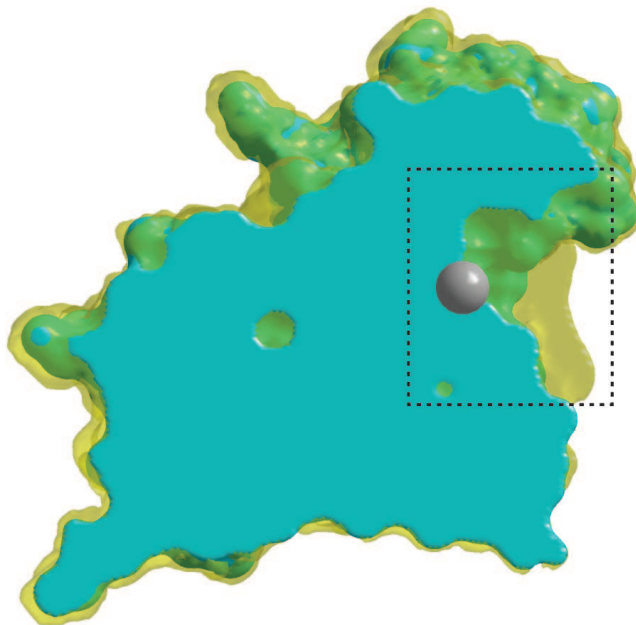


Figure 4: A cross-section of the region of electrostatic focusing in superoxide dismutase. The molecular surface of superoxide dismutase is shown in teal. The region of electrostatic focusing, where the uniform dielectric potential field and the nonuniform dielectric potential field differ by at least $2kt/e$, is shown in transparent yellow. The cavity region where considerable focusing is occurring is indicated by the dotted rectangle. A copper ion cofactor is shown as a grey sphere.

3.4 Data preparation

To evaluate the accuracy of our methods, we use two data sets. The first is a case study on Cu-Zn Superoxide Dismutase (PDB: 2SOD). Charges in Superoxide Dismutase were assigned according to [28]. Most notably His-41 was assigned a charge of +1 and His-61 was assigned -1 (charges were split between the ring nitrogens), as discussed here [1, 15, 54].

The second is composed of 884 DNA-protein complexes analyzed first by Rohs and coworkers [45]. From this list, PDB entries 1GJI, 1EGW, 1R71, 2W7N, 3L4J, 3L4K, 3N78, 3N7B and 3OD8 were eliminated because of missing nucleotide side chains, and 3HJF, 3HK2, 3HM9, 3HO1 and 3HVR were eliminated for lacking paired DNA chains. 2XSD was removed because its generated in error in Delphi.

In the remaining 866 structures, we removed ions, duplicated atoms, and waters. Protonation was corrected using the *reduce* program (ver. 3.14) [60] from MolProbity [12]. Potential fields were computed with DelPhi 5.1 [43].

3.5 Implementation Details

After computations with Delphi are complete, computing the focusing region required an average of 3 seconds on one core of an AMD Opteron 6128 with 2 gigabytes of random access memory (RAM) per core. These processors were part of Corona, a 1040 core cluster at Lehigh University. Visualization for figures 3, 4, and 6 was performed with in-house soft-

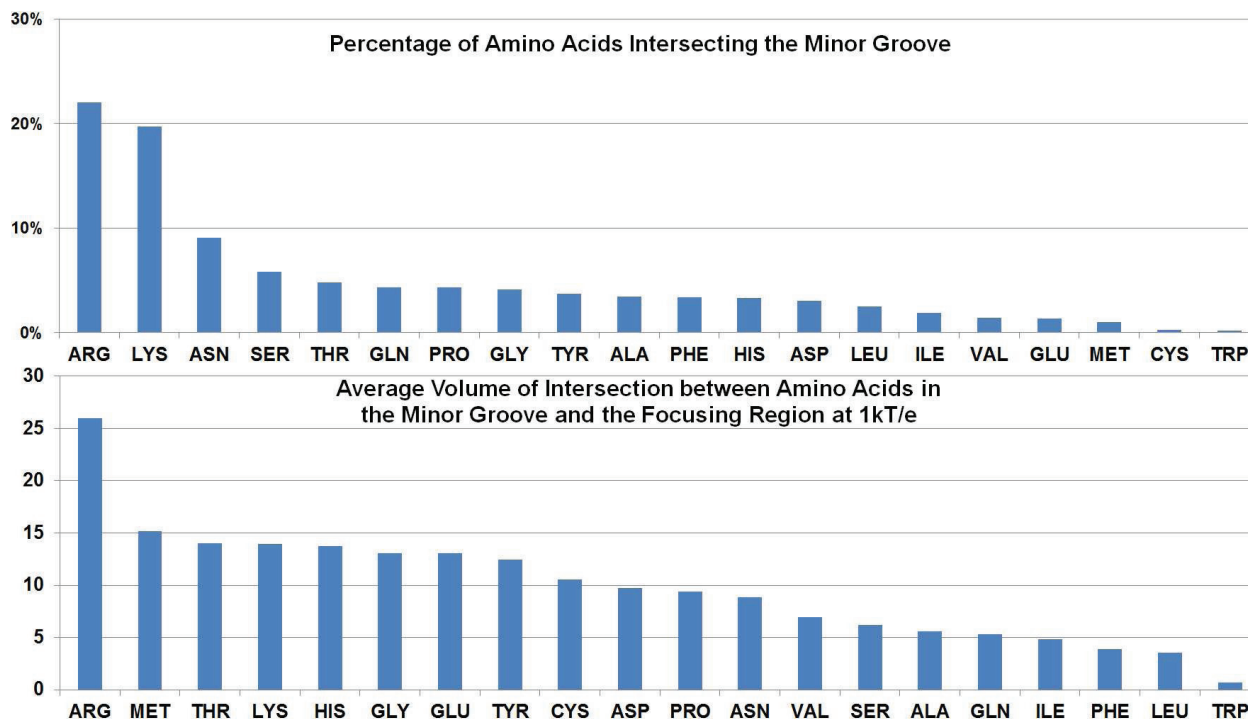


Figure 5:

ware using OpenGL on video cards with the Nvidia Geforce GT 330 chipset. Cross-sectional views in Figures 4 and 6 were computed with Boolean set operations.

4. EXPERIMENTAL RESULTS

4.1 Case study: Superoxide dismutase

To demonstrate the accuracy and applicability of the methods described here, we first demonstrate that our method detects regions near protein structures that correspond to established instances of electrostatic focusing. For this demonstration, we use Superoxide dismutase (*SOD*). Superoxide dismutase catalyzes the dismutation of superoxide, a toxic compound, into oxygen and hydrogen peroxide. This action prevents the highly active superoxide from reacting with sensitive chemicals in biological systems, or from forming other damaging compounds [2].

SOD functions by attracting superoxide molecules into its active site with an electric field. The strong positive charge attracts the negatively charged superoxide, enabling the enzyme to catalyze reactions. However, if we evaluate the potential field using the uniform dielectric model, the entire protein appears, from a long distance, to be enclosed by negative potential [28]. If this was true, *SOD* would repel its own substrates. Klapper and coworkers [28] resolved this dilemma when they observed that electrostatic focusing in a nonuniform dielectric model enhances potentials inside the cavity and makes the positive potential of the cavity detectable at long distances.

Beginning with the structure of *SOD* (PDB: 2SOD), we evaluated the region where the electrostatic potential is enhanced by more than 2 kT/e (e.g. $K > 2$). In a thin layer at

most 1 Å in depth, the focusing region covered parts of the molecular surface on the exterior of the protein, but it entirely filled the binding cavity, which is nearly 8 Å in depth in some places. This effect is illustrated in Figure 4 with a cross section of *SOD* that cuts directly through the cavity region, showing the concentration of electrostatic focusing in the cavity. We also observed this concentration at several other potential differences ($K = \{1.5 \text{ kT/e}, 1.0 \text{ kT/e}, 0.5 \text{ kT/e}\}$). These observations localize electrostatic focusing in exactly the same region as that identified by Klapper and coworkers, illustrating the accuracy of our method.

4.2 Focusing in Protein-DNA Complexes

Proteins take advantage of many aspects of DNA sequence and shape in order to selectively recognize regions on the double helix. In 2009, Rohs and coworkers observed that narrowness in the minor groove, electrostatic focusing inside the groove and arginine binding were strongly correlated [45]. Arginines in the minor groove under these circumstances were particularly enriched over other amino acids. These observations, on a large number of protein-DNA complexes, revealed a mode of protein-DNA interaction that depends on electrostatic focusing.

In this work, we reproduced part of these observations to test the accuracy of our method. First, we counted the number and type of amino acids that intersected the minor groove in every DNA-protein complex, using the method described in section 3.2. These totals are plotted in the top of Figure 5, and they reproduce the same strong enrichment of arginines contacting the minor groove, as observed by Rohs et al. [45]. Lysines, the other amino acid with a distinctly positive net charge and a natural electrostatic complement to the negatively charged DNA, were also enriched in the

minor groove.

Next, for every DNA structure, we generated the focusing region with $K = 1\text{kT}/e$. For each amino acid in contact with the minor groove, as identified above, we compared the region occupied by the focusing region with the region occupied by the amino acid, using VASP as described in section 3.3. Categorizing the data by amino acid type, we computed the average volume of intersection between each type of amino acid and the focusing region. These data are plotted in the bottom of Figure 5. Arginines occupied an average of 25.9 \AA^3 inside the focusing region, 72% greater than the second largest occupants, methionines (15.1 \AA^3), and double or nearly double that of all other amino acids. These results thus reproduce earlier observations that arginines are enriched in regions of electrostatic focusing [45], verifying the accuracy of our methods.

The variations in electrostatic focusing that occur in the minor groove can be significant. Figure 6 illustrates close-up views of two different regions in the minor groove (from pdb 1HF0). In Figure 6b, an arginine residue occupies the entire depth of the focusing region, where it can completely immerse its guanidinium group. Ten base pairs away, where the minor groove is wider, the focusing region occupies only a thin film 1 \AA in depth.

5. CONCLUSIONS

We have presented a novel computational representation of regions where electrostatic focusing occurs. Our representation enables the focusing region to be compared as a three dimensional solid with Boolean set operations. Our methods are also the first enable the region of electrostatic focusing to be explicitly visualized.

We demonstrated the accuracy of our method in a case study of superoxide dismutase, the protein on which electrostatic focusing was first observed [28]. We showed that electrostatic focusing occurs to some extent throughout the surface of the protein, but that focusing inside the binding site was particularly significant. In this case, our representation enabled us to visualize the region where focusing occurs, but because the region was represented as a three dimensional solid, its shape could be readily compared with existing methods (i.e. VASP [11]). This comparison capability creates new opportunities for a detailed analysis of the electrostatic character of multiple binding sites, pointing to new applications in protein engineering and for determining influences on binding specificity. We also demonstrated, at a large scale, that our method capably represents the character of electrostatic focusing in the minor groove of DNA. Specifically, we verified that our representation correctly detected the enrichment of arginines in regions of electrostatic focusing, as observed earlier by Rohs et al. [45].

The results presented here are the first to show that electrostatic focusing can be represented and compared as a geometric entity. This representation captures a crucial aspect of the electrostatic field outside the molecular surface, where it influences other molecules, and it is the first representation to enable the geometry of the focusing region to be seamlessly integrated into geometric comparisons of molecular shape. Having tested our methods in detail and on

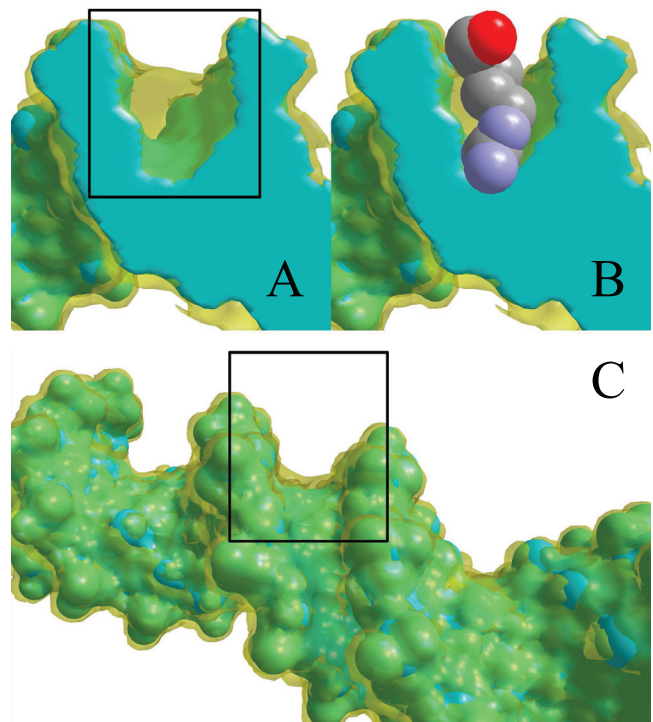


Figure 6: Variations in electrostatic focusing in the minor groove. The molecular surface of DNA (pdb: 1HF0) is shown in teal and the surface of the focusing region at $K = 1\text{kT}/e$ is shown in transparent yellow. The top two figures illustrate a cross section of the minor groove of DNA (black box) where electrostatic focusing is occurring, both without (A) and with (B) bound arginine. The bottom figure illustrates a different region of the minor groove (black box) where less focusing is occurring.

a large scale, and also on both proteins and DNA, our results represent an initial proof of concept that electrostatic focusing can be broadly and robustly applied in structural comparison algorithms.

Acknowledgment. The author sincerely thanks Remo Rohs for supplying the DNA data set.

6. REFERENCES

- [1] S. A. Allison, G. Ganti, and J. A. McCammon. Simulation of the diffusion-controlled reaction between superoxide and superoxide dismutase. i. simple models. *Biopolymers*, 24(7):1323–1336, 1985.
- [2] R. G. Alscher, N. Erturk, and L. S. Heath. Role of superoxide dismutases (sods) in controlling oxidative stress in plants. *Journal of Experimental Botany*, 53(372):1331–1341, 2002.
- [3] 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 Res*, 28(1):235–42, Jan. 2000.
- [4] T. A. Binkowski and A. Joachimiak. Protein functional surfaces: global shape matching and local spatial alignments of ligand binding sites. *BMC Struct*

- Biol*, 8:45, Jan. 2008.
- [5] E. Catmull and R. Rom. *A class of local interpolating splines*. Academic Press, 1974.
 - [6] R. Chauvin. Explicit periodic trend of van der waals radii. *J Phys Chem*, 96:9194–9197, 1992.
 - [7] B. Y. Chen and S. Bandyopadhyay. A Statistical Model of Overlapping Volume in Ligand Binding Cavities. In *Proceedings of the Computational Structural Bioinformatics Workshop (CSBW 2011)*, pages 424–31, 2011.
 - [8] B. Y. Chen and S. Bandyopadhyay. VASP-S: A Volumetric Analysis and Statistical Model for Predicting Steric Influences on Protein-Ligand Binding Specificity. In *Proceedings of 2011 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 22–9, 2011.
 - [9] B. Y. Chen, V. Y. Fofanov, D. H. Bryant, B. D. Dodson, D. M. Kristensen, A. M. Lisewski, M. Kimmel, O. Lichtarge, and L. E. Kavragi. The MASH pipeline for protein function prediction and an algorithm for the geometric refinement of 3D motifs. *Journal of Computational Biology*, 14(6):791–816, 2007.
 - [10] B. Y. Chen, V. Y. Fofanov, D. M. Kristensen, M. Kimmel, O. Lichtarge, and L. E. Kavragi. Algorithms for structural comparison and statistical analysis of 3D protein motifs. In *Pacific Symposium on Biocomputing.*, volume 345, pages 334–45, Jan. 2005.
 - [11] B. Y. Chen and B. Honig. VASP: A Volumetric Analysis of Surface Properties Yields Insights into Protein-Ligand Binding Specificity. *PLoS Comput Biol*, 6(8):11, 2010.
 - [12] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson. Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica D: Biological Crystallography*, D66:12–21, 2010.
 - [13] M. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612):709–713, Aug. 1983.
 - [14] J. Dundas, L. Adamian, and J. Liang. Structural signatures of enzyme binding pockets from order-independent surface alignment: a study of metalloendopeptidase and nad binding proteins. *Journal of Molecular Biology*, 406(5):713–729, 2011.
 - [15] E. D. Getzoff, J. A. Tainer, P. K. Weiner, P. A. Kollman, J. S. Richardson, and D. C. Richardson. Electrostatic recognition between superoxide and copper, zinc superoxide dismutase. *Nature*, 306, 1983.
 - [16] J. F. Gibrat, T. Madej, and S. H. Bryant. Surprising similarities in structure comparison. *Curr Opin Struct Biol*, 6(3):377–85, June 1996.
 - [17] S. C. Harvey. Treatment of Electrostatic Effects in Macromolecular Modeling. *Proteins Struct Func and Genetics*, 5:78–92, 1989.
 - [18] Z. S. Hendsch and B. Tidor. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein science : a publication of the Protein Society*, 3(2):211–26, Feb. 1994.
 - [19] L. Holm and C. Sander. Mapping the protein universe. *Science*, 273(5275):595–603, Aug. 1996.
 - [20] B. Honig and A. Nicholls. Classical electrostatics in biology and chemistry. *Science*, 268(5214):1144–1149, 1995.
 - [21] B. H. Honig, W. L. Hubbell, and R. F. Flewelling. Electrostatic Interactions in Membranes and Proteins. *Annual Review of Biophysics and Biophysical Chemistry*, 15:163–193, 1986.
 - [22] W. A. Kabsch. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallographica A*, 34:827–828, 1978.
 - [23] A. Kahraman, R. J. Morris, R. a. Laskowski, and J. M. Thornton. Shape variation in protein binding pockets and their ligands. *J Mol Biol*, 368(1):283–301, Apr. 2007.
 - [24] E. Kangas and B. Tidor. Electrostatic Complementarity at Ligand Binding Sites: Application to Chorismate Mutase. *The Journal of Physical Chemistry B*, 105(4):880–888, Feb. 2001.
 - [25] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors. In *European Symposium on Geometry Processing 2003*, 2003.
 - [26] K. Kinoshita and H. Nakamura. Protein informatics towards function identification. *Current Opinion in Structural Biology*, 13(3):396–400, June 2003.
 - [27] K. Kinoshita and H. Nakamura. Identification of the ligand binding sites on the molecular surface of proteins. *Protein Sci*, 14:711–718, 2005.
 - [28] I. Klapper, R. Kagstrom, R. Fine, K. Sharp, and B. Honig. of Electric in the Active of Ionic Site of Cu-Zn Effects. *Proteins: Struct. Funct. Genet.*, 1:47–59, 1986.
 - [29] R. A. Laskowski. SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graph*, 13(5):323–30, 307–8, Oct. 1995.
 - [30] B. Lee and F. M. Richards. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*, 55(3):379–400, Feb. 1971.
 - [31] L. P. Lee and B. Tidor. Optimization of electrostatic binding free energy. *Journal of Medical Physics*, 106(21):8681–8690, 1997.
 - [32] W. E. Lorensen and H. E. Cline. Marching Cubes: A High Resolution 3D Surface Construction Algorithm. In *Proceedings of the 14th annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH '87)*, volume 21, pages 163–170, 1987.
 - [33] J. B. Matthew. Electrostatic effects in proteins. *Annual review of biophysics and biophysical chemistry*, 14:387–417, Jan. 1985.
 - [34] S. McLaughlin. The electrostatic properties of membranes. *Annual review of biophysics and biophysical chemistry*, 18:113–36, Jan. 1989.
 - [35] H. Nakamura. Roles of electrostatic interaction in proteins. *Quarterly Reviews of Biophysics*, 29:1–90, 1996.
 - [36] C. A. Orengo and W. R. Taylor. SSAP: Sequential Structure Alignment Program for Protein Structure Comparison. *Method Enzymol*, 266:617–635, 1996.
 - [37] E. K. O’Shea, R. Rutkowski, and P. S. Kim. Mechanism of specificity in the Fos-Jun oncoprotein

- heterodimer. *Cell*, 68(4):699–708, 1992.
- [38] D. Petrey and B. Honig. GRASP2: visualization, surface properties, and electrostatics of macromolecular structures and sequences. *Method Enzymol*, 374(1991):492–509, Jan. 2003.
- [39] F. Polticelli, B. Honig, P. Ascenzi, and M. Bolognesi. Structural determinants of trypsin affinity and specificity for cationic inhibitors. *Protein Science*, 8(12):2621–2629, 1999.
- [40] J. W. Ponder and D. A. Case. Force fields for protein simulations. In V. Daggett, editor, *Protein Simulations*, volume 66 of *Advances in Protein Chemistry*, pages 27 – 85. Academic Press, 2003.
- [41] M. T. Record, C. F. Anderson, and T. M. Lohman. Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Quarterly Reviews of Biophysics*, 11(2):103–178, 1978.
- [42] D. W. Ritchie and G. J. L. Kemp. Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *J Comput Chem*, 20(4):383, Mar. 1999.
- [43] W. Rocchia, E. Alexov, and B. Honig. Extending the applicability of the nonlinear poisson-boltzmann equation: Multiple dielectric constants and multivalent ions. *Journal of Physical Chemistry B*, 105(28):6507–6514, 2001.
- [44] N. K. Rogers. The modelling of electrostatic interactions in the function of globular proteins. *Progress in Biophysics and Molecular Biology*, 48(1):37–66, 1986.
- [45] R. Rohs, S. M. West, A. Sosinsky, P. Liu, R. S. Mann, and B. Honig. The role of DNA shape in protein-DNA recognition. *Nature*, 461:1248–1253, 2009.
- [46] M. Rosen, S. L. Lin, H. Wolfson, and R. Nussinov. Molecular shape comparisons in searches for active sites and functional similarity. *Protein Eng*, 11(4):263–77, Apr. 1998.
- [47] R. B. Russell. Detection of protein three-dimensional side-chain patterns: new examples of convergent evolution. *J Mol Biol*, 279(5):1211–27, June 1998.
- [48] S. Schmitt, D. Kuhn, and G. Klebe. A New Method to Detect Related Function Among Proteins Independent of Sequence and Fold Homology. *J Mol Biol*, 323(2):387–406, Oct. 2002.
- [49] K. A. Sharp and B. Honig. Electrostatic Interactions in Macromolecules: Theory and Applications. *Annual Review of Biophysics and Biophysical Chemistry*, 19:301–332, 1990.
- [50] M. Shatsky, A. Shulman-peleg, R. Nussinov, and H. J. Recognition of Binding Patterns Common to a Set of Protein Structures. *Lect Notes Comput Sc*, 3500:440–455, 2005.
- [51] I. N. Shindyalov and P. E. Bourne. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng*, 11(9):739–47, Sept. 1998.
- [52] C. V. Sindelar, Z. S. Hendsch, and B. Tidor. Effects of salt bridges on protein structure and design. *Protein Science*, 7(9):1898–1914, 1998.
- [53] K. Soman, A. S. Yang, B. Honig, and R. Fletterick. Electrical potentials in trypsin isozymes. *Biochemistry*, 28(26):9918–9926, 1989.
- [54] J. A. Tainer, E. D. Getzoff, K. M. Beem, J. S. Richardson, and D. C. Richardson. Determination and analysis of the 2 a-structure of copper, zinc superoxide dismutase. *J. Mol. Biol.*, 160(2):181–217, 1982.
- [55] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991.
- [56] C. D. Waldburger, J. F. Schildbach, and R. T. Sauer. Are buried salt bridges important for protein stability and conformational specificity? *Nature Structural Biology*, 2(2):122–128, 1995.
- [57] L. Wang, T. O’Connell, A. Tropsha, and J. Hermans. Energetic decomposition of the alpha-helix-coil equilibrium of a dynamic model system. *Biopolymers*, 39(4):479–489, 1996.
- [58] A. Warshel and S. T. Russell. Calculations of electrostatic interactions in biological systems and in solutions. *Quarterly Reviews of Biophysics*, 17(3):283–422, 1984.
- [59] W. C. Wimley, K. Gawrisch, T. P. Creamer, and S. H. White. Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proceedings of the National Academy of Sciences of the United States of America*, 93(7):2985–90, Apr. 1996.
- [60] J. M. Word, S. C. Lovell, J. S. Richardson, and D. C. Richardson. Using hydrogen atom contacts in the choice of sidechain amide orientation. *Journal of Molecular Biology*, 285(4):1735–1747, 1999.
- [61] L. Xie and P. E. Bourne. Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments. *Proc Natl Acad Sci U S A*, 105(14):5441–6, Apr. 2008.
- [62] A. S. Yang and B. Honig. Free energy determinants of secondary structure formation: I. alpha-Helices. *Journal of Molecular Biology*, 252(3):351–365, 1995.
- [63] G. Zheng, X.-J. Lu, and O. WK. Web 3dna - a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Research*, 37:W240–W246, 2009.