SurfStamp: 3D Printer Compatible Molecular Surface Representation

Toshiyuki Oda[1]

[1] Independent software developer

e-mail: odat1248@gmail.com

Abstract

SurfStamp is an application that is used to generate textures for surface models of proteins. The textures contain information about surface residues and the information is drawn directly on the 3D object of the models. This approach is more intuitive than the labeling functions that most three-dimensional (3D) structure viewers use to show residue information. Therefore, the use of this application enables researchers, readers, or audiences to easily determine which residues are contributing the surface they are focusing on.

Availability

The application is provided under the open-source Apache License Version 2.0 (http://www.apache.org/licenses/LICENSE-2.0). The application and source code are available from https://github.com/yamule/SurfStamp-public/releases .

Introduction

Since visualization of protein structures in 3D space is one of the fundamental ways to analyze properties and function of proteins, various techniques have been developed and are now in widespread use; e.g., tube models, ball-and-stick models, surface models, and so forth (Gu and Bourne, 2009). Among them, molecular surface model is one of the more important techniques; because the properties of the surface regions determine their binding affinities to other molecules (e.g. small chemical compounds, DNAs or other proteins) and these affinities are strongly related to the functions of proteins.

For this reason, popular 3D structure visualization software products such as PyMOL (http://www.pymol.org/), VMD (Humphrey, et al., 1996), and JMol (http://www.jmol.org/) each have functions that can be used to construct molecular surface models. Using these software products, researchers can see protein shapes and interesting regions such as cavities, hydrophobic regions, or charged clusters, which will expand their knowledge and provide clues for further experiments.

However, when examining surface models, it is often difficult to see which residues contribute to the regions under examination. While most molecular visualization software is capable of adding labels that represent residues, those labels often need manual rearrangement in order to make them understandable when they are shown with a surface.

In this paper, I introduce SurfStamp, which is a software tool for generating textures which show residue types and numbers on a surface model. Using this application, residue information is drawn directly on the surface object, which makes it easy to understand which residues are contributing to the regions of interest. In addition, since the residue information is included in a 3D object, either as paintings on the object or as part of the object, it can be utilized in 3D printing. Therefore, it is hoped to contribute to many scenes in research and pedagogy (Da Veiga Beltrame, et al., 2017).

Implementation

Work flows

The works which SurfStamp does are quite simple. It can be divided into 5 or 6 steps as follows.

1. Surface model generation

   SurfStamp creates models using the similar approach of molecular surface mode of EDTSurf (Xu and Zhang, 2009), however, the integrity of the algorithm is not validated. Users can import surface models created by other popular software products such as EDTSurf or

PyMOL(Schrodinger, 2015).

2. Grouping faces (polygons)

All faces in those geometries are grouped according to their nearest residues. The distances between the center point of each face and any atoms of residues are used.

3. Topology validation

The topologies of grouped faces are checked. When the default setting is used, each group is separated into two groups heuristically if there are large areas that cannot be seen from one side. If there are loop structures such as tori, which cause failures of UV unwrapping, those structures are removed. Structures such as sharp hills are also removed because they result in extremely skewed textures.

4. UV unwrapping

The faces are unwrapped onto a two-dimensional (2D) plane and the texture coordinates are decided.

5. Texture creation

The unwrapped faces are then painted according to the coloring scheme, after which text showing the residue types and numbers are placed inside the unwrapped textures. Without "-extrude" option, the process stops here and result files are created.

6. Extrusion

With "-extrude" option, SurfStamp doesn't create texture files but extrudes residue information from surface model as solid objects.

Coloring scheme

When the default setting is in place, a coloring scheme similar to ClustalX (Thompson, et al., 2002) is used. The Hydropathy index (Kyte and Doolittle, 1982) and Charge (Zimmerman, et al., 1968) extracted from AAIndex Database (Kawashima, et al., 2008) are provided as presets. The CPK coloring scheme (Corey and Pauling, 1953; Koltun, 1965), B-factor or occupancy columns can be used for atom, too. However, it should be noted that users can apply their own original coloring schemes by using a simple script (the details are written in the readme.txt file provided with the application).

Input file format

For 3D polygon objects, SurfStamp accepts Wavefront .obj format files (https://en.wikipedia.org/wiki/Wavefront_.obj_file), Polygon file (.ply) format files (https://en.wikipedia.org/wiki/PLY_(file_format)), and binary STL format (https://en.wikipedia.org/wiki/STL_(file_format)) as inputs. I optimized the software to import files by PyMOL (v0.99, https://sourceforge.net/projects/pymol/files/Legacy/), which is one of the most popular visualization software tools, and EDTSurf, which is a character user interface (CUI) based molecular surface generation tool that can be used for batch processes. However, it is hoped that the files generated by other software can be processed as well.

For atom coordinates, standard file formats for protein structures, the Protein Data Bank (PDB) file format, and the macromolecular Crystallographic Information File (PDBx/mmCIF) format are supported. For ligand molecules, SDF file format is also supported.

Output file format

The default mode of SurfStamp produces 3D graphics files in .obj and Material library (.mtl) formats (https://en.wikipedia.org/wiki/Wavefront_.obj_file#Material_template_library), while texture image files are provided in Portable Network Graphics (.png) formats (https://www.w3.org/TR/PNG/). Since these formats are widely used in the 3D or 2D graphics field, they are supported by most 3D or 2D graphics software applications.

In addition, texture images are also provided in the Scalable Vector Graphics (.svg) format (https://www.w3.org/TR/SVG/), which can be processed with vector draw software such as Inkscape (https://inkscape.org/). Using this format (and its related software), users can change the texts, fonts, colors, and resolution rates more easily than can be done with raster-type formats such as .png.

For the extrusion mode, it exports objects in binary STL format.

Additional features

SurfStamp can produce ball-and-stick molecule models. This feature was added to allow researchers to determine whether a molecular surface has been successfully created. (Since some visualization software produce 3D objects according to its camera position, the coordinates of produced objects sometimes differ from coordinates of atoms.) Note, however, in addition to this debugging purpose, the feature is also useful for showing the positions of ligands, as can be seen in Fig. 1.

Running, copying, and distributing

I coded SurfStamp entirely in Java language and did not use any external libraries. As a result, it can

be run in any operating system (OS) as long as Java VM is installed. In addition, since I provide this software under the open-source Apache License version 2.0 (http://www.apache.org/licenses/LICENSE-2.0), anyone can use, modify, and distribute this software freely as long as they follow the license term.

Output example

Figure 1 shows a SurfStamp output example applied against the Galectin carbohydrate recognition domain (CRD), 3ZSJ (Saraboji, et al., 2012) molecule. As you can see, the surface consisting of residues, THR181, GLU184, HIS158, and so forth, that are interacting with the beta-lactose shown in the ball-and-stick model.

Conclusion

Since SurfStamp provides intuitive representations of molecular surfaces, the contents made with this application will be useful in usual research, education, creating figures in academic papers, presentations, and so forth. In addition, since a generous license policy has been adopted, and since the application lacks significant platform dependencies, it can be bundled with other software and can be included in web services. As a result, SurfStamp can be expected to increase the power of molecular visualization techniques, and thereby facilitate research and education.

Acknowledgements

I really thank staffs of NIH 3D print Exchange (https://3dprint.nih.gov/), especially James Tyrwhitt-Drake (In 2016. As I've lost his/her contact information, I didn't know about the prefix.), for helpful discussions and suggestions. "-extrude" option for mono-color 3D printing was implemented following their suggestions. I sincerely sorry for having provided programs with a lot of bugs in the early days. I thank Dr. Yoshinori Fukasawa. "-color_scheme 4" and "-color_scheme 5" which utilize B-factor and occupancy column, respectively, were implemented following his suggestions. I thank Mr. Masaki Kato. "-color_dx" option which utilizes result of apbs(Baker, et al., 2001) was implemented following his suggestions.

Conflict of Interest

The author is an employee of Lifematics Inc. This work was privately done by the author using his own resources.

Figure Legends

Figure 1. (A) Example of SurfStamp output. SurfStamp was applied to the entry 3ZSJ. The ClustalX (Thompson, et al., 2002) coloring scheme was used. Beta-lactose is shown using a ball-and-stick model with green carbon and red oxygen atoms. A rendered image was captured from the result constructed using the simple .obj viewer made by me, which uses Three.js (http://threejs.org/) for 3D graphics library. (B) Focused image of the 3ZSJ beta-lactose binding site. The structure was created with options; -tile -font_size 24 -nosep, which was potentially much more appropriate for 2D images in manuscript.

Figure 2. An example 3D printed model created with "-extrude" option using 3D object made by EDTSurf (Xu and Zhang, 2009). The 3D printing was performed using DMM make (https://make.dmm.com/) service. 1 Å was expanded to 2.5 mm. The data is available from https://3dprint.nih.gov/discover/3dpx-004144 .

Figures and Tables

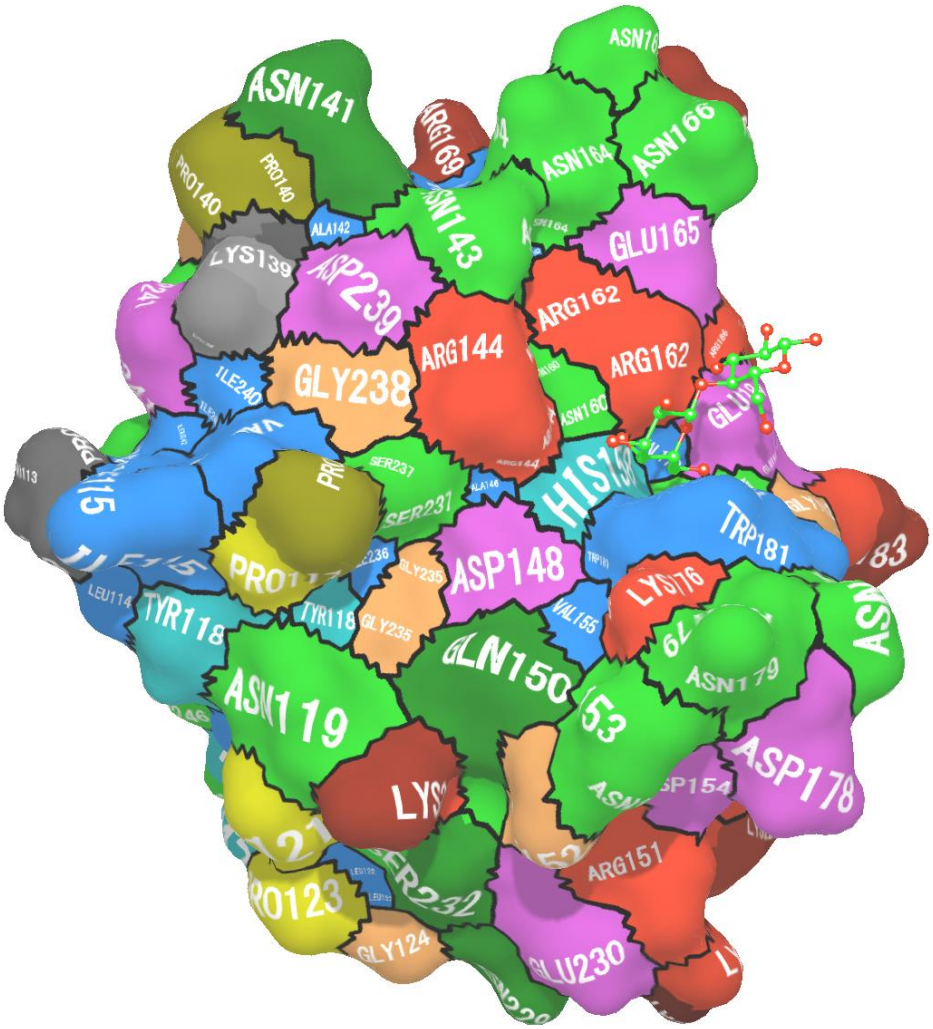

Figure 1A

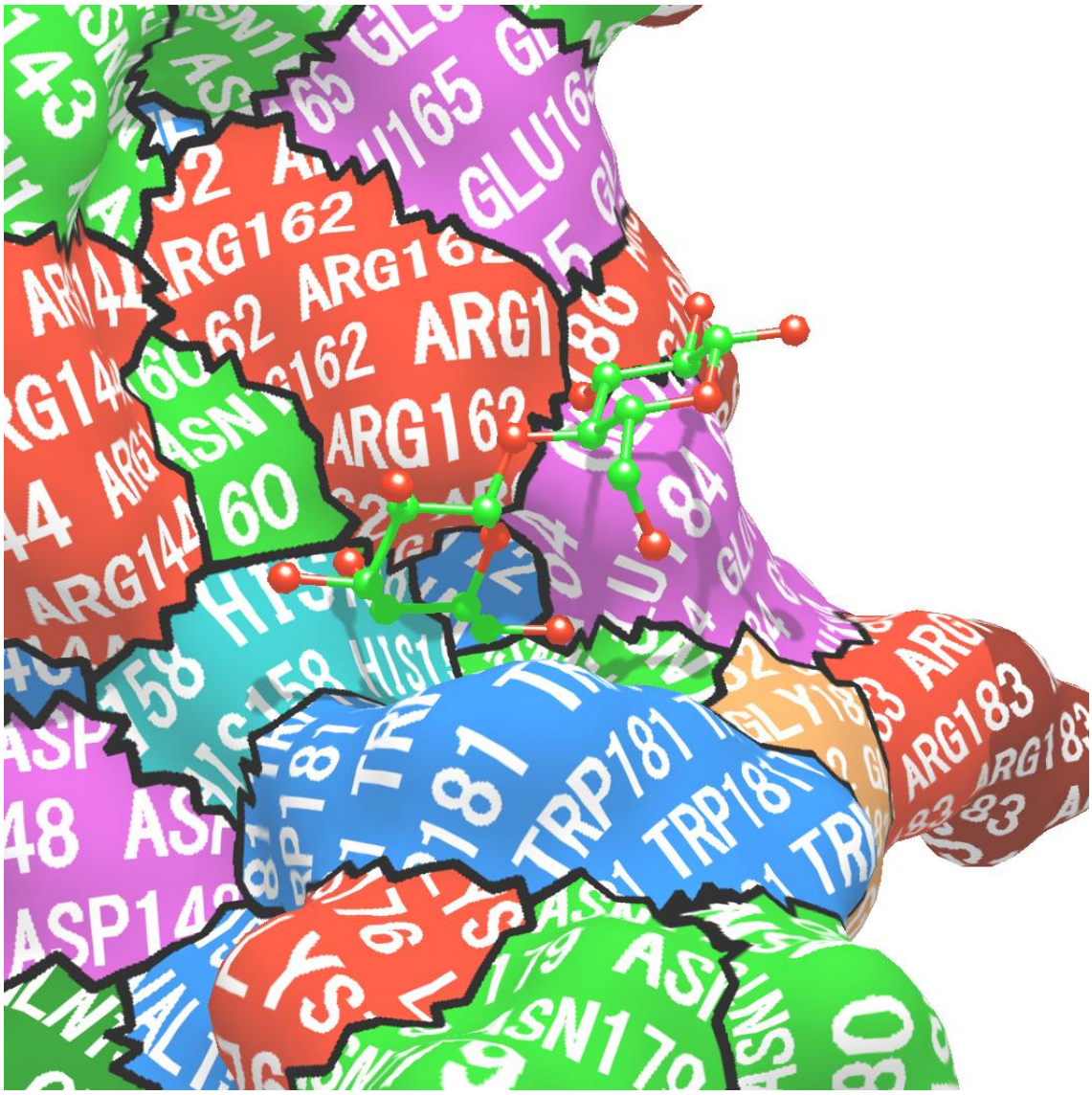
ARG162
ARG163
ARG183
TRP181
GLU165
ASN179
ASP148
HIS158
ASN160

Figure 1B

Figure 2

References

Baker, N.A., *et al.* Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* 2001;98(18):10037-10041.

Corey, R.B. and Pauling, L. Molecular Models of Amino Acids, Peptides, and Proteins. *Review of Scientific Instruments* 1953;24(8):7.

Da Veiga Beltrame, E., *et al.* 3D Printing of Biomolecular Models for Research and Pedagogy. *J Vis Exp* 2017(121).

Gu, J. and Bourne, P.E. Structural bioinformatics. Hoboken, N.J.: Wiley-Blackwell; 2009.

Humphrey, W., Dalke, A. and Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* 1996;14(1):33-38, 27-38.

Kawashima, S., *et al.* AAindex: amino acid index database, progress report 2008. *Nucleic Acids Res* 2008;36(Database issue):D202-205.

Koltun, W.L. Space filling atomic units and connectors for molecular models. U. S. Patent; 1965.

Kyte, J. and Doolittle, R.F. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 1982;157(1):105-132.

Saraboji, K., *et al.* The carbohydrate-binding site in galectin-3 is preorganized to recognize a sugarlike framework of oxygens: ultra-high-resolution structures and water dynamics. *Biochemistry* 2012;51(1):296-306.

Thompson, J.D., Gibson, T.J. and Higgins, D.G. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* 2002;Chapter 2:Unit 2 3.

Xu, D. and Zhang, Y. Generating triangulated macromolecular surfaces by Euclidean Distance Transform. *PLoS One* 2009;4(12):e8140.

Zimmerman, J.M., Eliezer, N. and Simha, R. The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol* 1968;21(2):170-201.