VLDP web server: a powerful geometric tool for analysing protein structures in their environment

Jérémy Esque[1], Sylvain Léonard[2,3,4,5], Alexandre G. de Brevern[2,3,4,5,*] and Christophe Oguey[1,*]

[1]LPTM, CNRS UMR 8089, Université Cergy-Pontoise, F-95302 Cergy-Pontoise, France, [2]INSERM, U665, DSIMB, F-75739 Paris, France, [3]Univ Paris Diderot, Sorbonne Paris Cité, UMR-S665, F-75739 Paris, France, [4]Institut National de la Transfusion Sanguine (INTS), F-75739 Paris, France and [5]Laboratoire d'Excellence GR-Ex, F-75739 Paris, France

ABSTRACT

Protein structures are an ensemble of atoms determined experimentally mostly by X-ray crystallography or Nuclear Magnetic Resonance. Studying 3D protein structures is a key point for better understanding protein function at a molecular level. We propose a set of accurate tools, for analysing protein structures, based on the reliable method of Voronoi–Laguerre tessellations. The Voronoi Laguerre Delaunay Protein web server (VLDPws) computes the Laguerre tessellation on a whole given system first embedded in solvent. Through this fine description, VLDPws gives the following data: (i) Amino acid volumes evaluated with high precision, as confirmed by good correlations with experimental data. (ii) A novel definition of inter-residue contacts within the given protein. (iii) A measure of the residue exposure to solvent that significantly improves the standard notion of accessibility in some cases. At present, no equivalent web server is available. VLDPws provides output in two complementary forms: direct visualization of the Laguerre tessellation, mostly its polygonal molecular surfaces; files of volumes; and areas, contacts and similar data for each residue and each atom. These files are available for download for further analysis. VLDPws can be accessed at http://www.dsimb.inserm.fr/dsimb_tools/vldp.

INTRODUCTION

Protein structures are the support of most major biological functions. Composed of a series of amino acids, proteins owe their specific properties to their side chains (1). Interactions between residues include covalent bonds, such as the disulphide bridges between two cysteines (2), and weaker bonds, such as hydrogen bonds (H-bonds), van der Waals interactions or hydrophobic effects. These interactions are essential for protein folding and for stabilizing protein structures. For instance, repetitive secondary structural elements as 3_{10}-, α-, π-helices, β-sheets and turns (3), which play a key role in protein architecture, are mostly maintained by hydrogen bonds. Thus, tools as DSSP (4,5) and STRIDE (6,7), for determining local folding, are interesting in the structural bioinformatics field.

In this article, we propose a tool for analysing protein structures with strong mathematical grounds, the Laguerre diagram, which is a weighted Voronoi diagram. Voronoi diagrams and their derived variants have been used often for the study of protein structures, protein–protein interactions, packing of the protein core, packing at the interface with water, protein cavities, assessing the quality of protein crystal structures (8,9). A few studies using Voronoi diagrams were also devoted to nucleic acids (10–12). At this day, the number of web servers based on this approach is limited. We can note the following: (i) The NIH web server (http://helixweb.nih.gov/structbio/basic.html) that computes the buried residue volumes in several tessellation ways. (ii) MolMovDB (13), which evaluates the packing of residues using Voronoi volumes (http://www.molmovdb.org/cgi-bin/voronoi.cgi). (iii) Vorolign (14), a

web server dedicated to structural alignment based on Voronoi contacts. (iv) MOLE (15) and its latest version MOLE2.0 (16), focused on the detection of molecular channels, pores and tunnels, based on the search of void paths in protein structures using the classical Voronoi method. It is dedicated to enzyme channels. (v) DiMoVo (17) (http://albios.saclay.inria.fr/dimovo/) proposes a service able to discriminate the crystallographic and biological protein–protein interactions.

In this research field, we propose Voronoi Laguerre Delaunay Protein web server (VLDPws). This tool is based on the Laguerre tessellation, and its aim is to compute a panel of properties of interest: (i) residue volumes, (ii) the residue contacts and (iii) exposure to solvent, i.e. accessibility. An important point is that VLDPws is the only tessellation approach that includes an effective solvent around the protein. Another improvement is the set of finely tuned weights used in the Laguerre tessellation (8,9). It allows us to take into account all the residues, including the accessible ones at the protein surface.

(i) In a previous work (8), we showed that computations based on tessellations gave a good agreement with experimental results in the evaluation of amino acid volumes, better than the other equivalent methodologies.
(ii) The protein contacts are a crucial feature to analyse protein structures. The usual definition is based only on geometrical parameters, such as distance-threshold (18), and used for instance to identify small compact units (protein units) (19–21). Information on protein contacts led to pertinent results for protein folding analysis or protein design. An extensive analysis of protein contacts defined by Laguerre tessellations in comparison with the classical approach underlined its interest (22).
(iii) Another important feature is the interaction of the protein with its environment. A classical method is to compute the residue accessibility using NACCESS tool (Hubbard, S.J. & Thornton, J.M. (1993), http://www.bioinf.manchester.ac.uk/naccess/); DSSP also provides such a measure (4). Owing to the addition of a realistic solvent, VLDPws evaluates the residue exposure to solvent, measured as a ratio of exposed surface area over the total surface area and called PIA (Polyhedral Interface Area). PIA was shown to give results compatible with classical approaches, but it also revealed some differences reflecting particular physico-chemical interactions of the residues with the environment, the explicit solvent.

Hence, VLDPws, based on the Laguerre tessellation, is efficient and precise in computing various protein properties. The good accuracy was previously benchmarked (8,22). The computed tessellation can be viewed in a static way owing to PyMOL software, or in an interactive way using Jmol software (23). Moreover, VLDPws is the only online web server including the protein in a realistic solvent for drawing the tessellation.

MATERIALS AND METHODS

The server can be used to study residue packing, residue accessibility and intra-protein contacts. Figure 1 summarizes the main steps of the VLDPws approach. At first, (a) the 3D protein structure is (b) solvated, then (c) the Delaunay tessellation is computed on the complete system including protein and solvent, and (d) the Laguerre tessellation is deduced from the Delaunay diagram. From the Laguerre tessellation, a list of local, global or averaged quantities are computed, including residue volumes, areas, contacts and accessibility (see Figure 2).

Input protein structure and solvation

The 3D protein structures must be in classical Protein DataBank format (24). Preferentially, the structure should be recognized like a monomeric chain (see Figure 1a). A topology file is created using GROMACS 4.5.3 (25). The hydrogens are added using OPLS-AA force field (26). Finally, the SOLVATE program (27) creates a shell of water around the protein. The SOLVATE parameters are a shell thickness of 5 Å, a solvent boundary set up to 2 (ngauss option). This first stage produces a pdb file including the protein and solvent coordinates (see Figure 1b).

Delaunay tessellation

A tessellation is a partition of space, i.e. a collection of polyhedra filling space without overlaps or gaps. Given a set of positions, the Delaunay tessellation is a partition of space into tetrahedra whose vertices are the system points. In the case of VLDPws, the points are the atom centres given in the pdb file (see Figure 1c). The Delaunay tessellation is built on the whole system (protein and solvent) using an incremental point insertion. At each insertion, the weighted circumscribed sphere criterion delimits a region where the tessellation needs to be adapted.

Laguerre tessellation and weighting

By a duality relationship, the Laguerre tessellation follows directly from the Delaunay diagram. In the Laguerre tessellation, each polyhedron is convex, and most often surrounds its corresponding atom centre (a vertex of the Delaunay diagram) (In a Voronoi tessellation—special case where all the weights are equal—all the polyhedra contain one and only one data point. In Laguerre tessellations, this one-to-one relation may be lost depending on the weights. However, in VLDPws with finely tuned weights, this atom polyhedron correspondence will only be broken in extreme physical circumstances.). In contrast to Voronoi tessellations, Laguerre tessellations are weighted. The Delaunay and Laguerre tilings and the duality relation all depend on the weights associated to the input points. See (8) and references therein for details. Here, the weights are scaled to the atom types as assessed in Esque *et al.* (8). Hence, the shape of Laguerre polyhedra depends on the weights and mutual positions of neighbouring atoms (see Figure 1d).
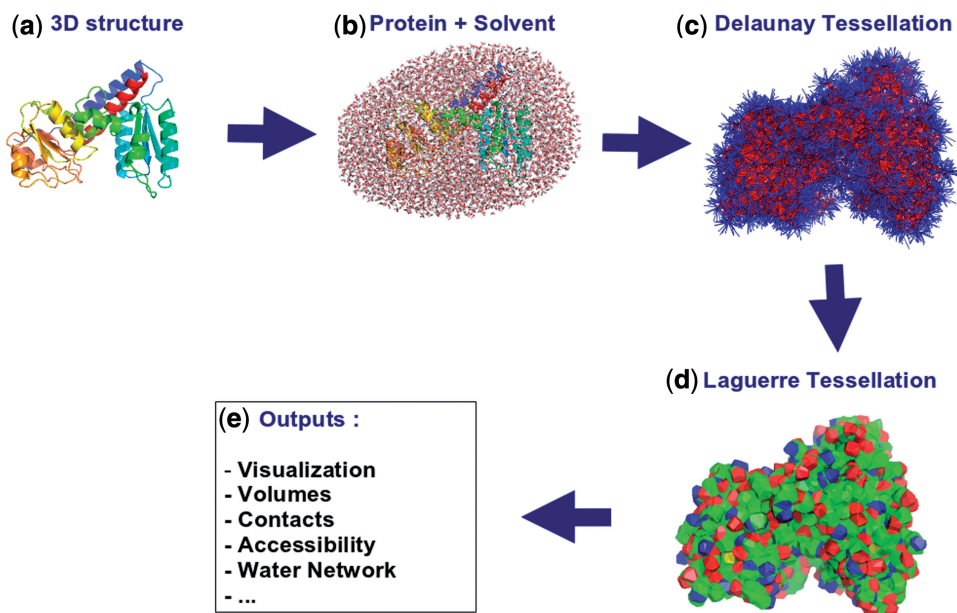
Figure 1. Flow chart of VLDPws' main methods. (**a**) A single 3D protein structure is entered as input, (**b**) solvent is added, (**c**) the Delaunay tessellation is performed on the whole system by VLDP [on the figure, only inter-residue (red) and residue-solvent (blue) contacts are shown], (**d**) the Laguerre tessellation is directly deduced as the dual of the Delaunay diagram. (**e**) From the tessellations, a range of analyses is carried out, i.e. volumes, areas, connectivity, residue contacts, residue accessibility.
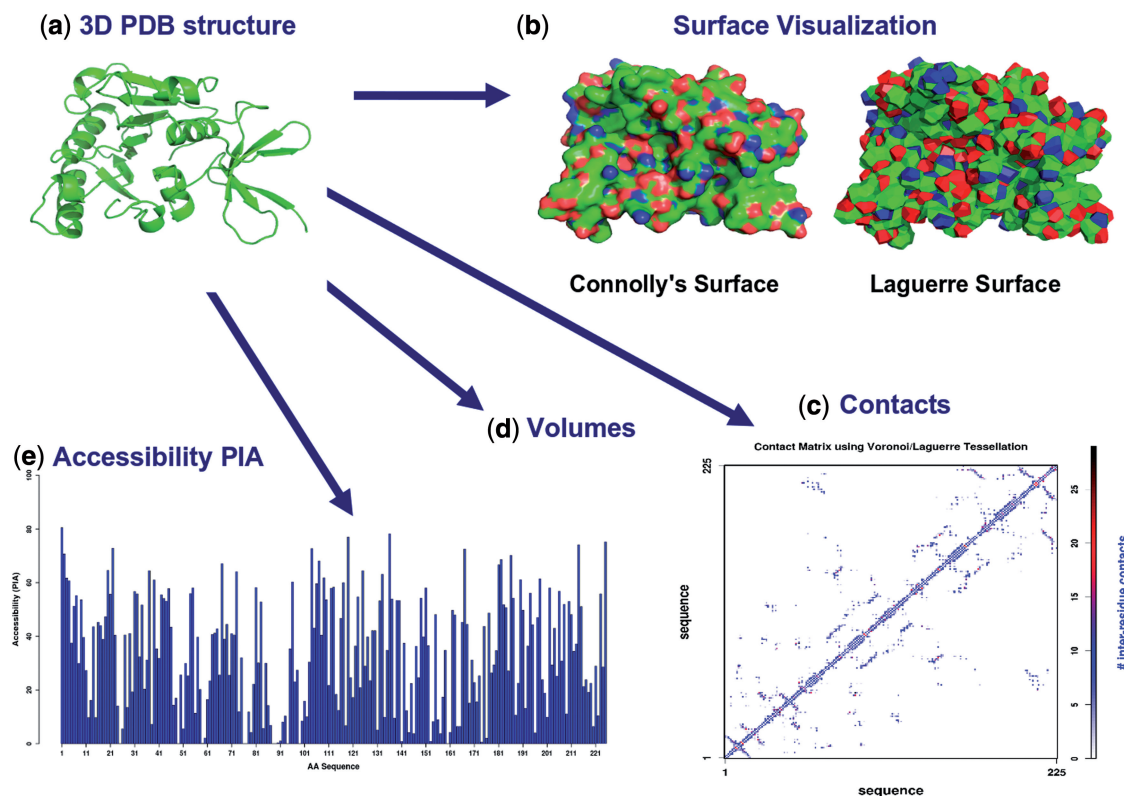

Figure 2. Protein analysis example. (**a**) The analysis of a fibrillarin homologue protein structure [PDB code 1FBN (29)]. (**b**) After constructing the Laguerre tessellation, the polygonal surface is displayed (as image or interactively with Jmol) in comparison with the standard method (Connolly's Surface), (**c**) the contact matrix is displayed, (**d**) the residue volumes are given in a partial table (not illustrated) or downloadable in a text file, and (**e**) the PIA accessibility is shown as a bar plot. All analysis products are downloadable individually or in a zip file.

Computational aspect of VLDP outputs

The source code of the program VLDP is written in Fortran 90. Several functionalities and output formats are available (see Figure 1e): (i) Visualization of the tessellation is performed by PyMol (The PyMOL Molecular Graphics System, Version 1.5, Schrödinger, LLC, www.pymol.org) and interactively with Jmol (23). The output for PyMol is in Compiled Graphics Objects format, whereas the command 'draw' is used for Jmol. (ii) The residue volumes are given as sums of corresponding atomic volumes. An atomic volume is the volume of its corresponding Laguerre polyhedron. The Laguerre polyhedra are subdivided into elementary tetrahedra, the sum of which gives the polyhedral volume (8). (iii) By definition, the contacts are specified by the Delaunay edges (22). (iv) The PIA is computed as the ratio of the residue Laguerre surface of contacts with the solvent divided by the total surface (of contacts) of the residue. PIA is a novel measure of accessibility (8). Some additional results are provided in flat files for expert users interested in solvation and porosity, namely, the topological genus of the protein surface, the water network decomposition into connected components, stratification into layers starting from a pre-defined origin. For the stratification analyses, two cases are computed: starting from the protein surface or from the boundary of the water box. The genus characterizes the protein/solvent interface. The connected components and stratification concern the organization of the water network.

DISCUSSION

The VLDP program is based on the Laguerre tessellation used for geometrical analyses of protein structures. From this tessellation, both metric and topological measures can be deduced opening the way for finer analyses. The VLDP program performs a panel of analyses, such as (i) the calculation of residue volumes, (ii) the determination of residue contacts, (iii) the residue accessibility and (iv) the organization of the water network.

Experimentally, the protein volume is obtained by measurements of partial specific volumes, e.g. by densitometry, or dilatometry, with the help of thermodynamic and hydrodynamic equations. Few developments had been made to compute it through *in silico* approaches. Only Tsai *et al.* (28) gave predictions directly deduced from the sequence information, but this method is not available on any web server. A benchmark on volumes confronting results evaluated with VLDP to data from the literature [values are taken from Tsai *et al.* (28)] underlined the quality of our approach (8). It shows that the Laguerre tessellation provides a good descriptor of the protein surface and packing.

The notion of contact comes out naturally in the form of a polygonal surface. Classical approaches are based on arbitrary distance thresholds (18) while tessellations discard this arbitrariness. The tessellation method provides a more realistic representation of the local packing in the structure; the contacts deduced from tessellation essentially consist of a complete list of nearest

neighbours around any residue. The method is flexible and adapts itself to non-homogeneously dense systems. An extensive benchmark showed that the Laguerre tessellation with well-tuned weights provides a better account of the geometry of the contacts (20). Apart from VLDP, no web server is available to perform such a task.

The presence of the solvent, numerically added to the macromolecular data, provides a new definition of the residue accessibility as a relative exposure to solvent, computed from the area of Laguerre faces. A systematic study has shown a rough linear correlation with the standard Accessible Surface Area (8), but there is a significant difference in precision and sensitivity mostly for partially buried residues.

Because all these properties are important in the structural bioinformatics field and others, we have implemented these analyses in the VLDPws.

FUNCTIONS AND USAGE

Web server VLDPws

VLDPws provides a user-friendly web interface to the analysis of protein structures that combines metrics and topology. The homepage contains a short summary describing the interest of program and the properties computed. The only input that must be provided is the protein structure (preferentially monomeric or merged into one chain), in PDB format. Two possibilities are offered: either direct download from the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) (24) or supply of a file, the filename of which must be given in a second window frame. In both cases, the pdb code and the used chain should be imperatively specified. At the bottom of the page, an example test with an url link is proposed. Additional tabs are as follows: (i) 'Contact', pointing to the authors' homepage; (ii) 'About', giving details on the methodology; (iii) 'Example', explaining or showing, on a concrete case, the input and output of the server (see later in the text). Figure 1 describes the successive steps that lead from a protein structure to the output results of the Laguerre tessellation.

Input

A single PDB structure must be provided (Figure 1a). A check is performed to ensure that only natural amino acids are used and that the specified chain is here.

Background step 'VLDP running'

After checking the input PDB format, a shell of water is first computed with SOLVATE software (27) (see 'Materials and Methods' section and Figure 1b). Then, using the atom coordinates of the extended system (protein + solvent), the Delaunay tetrahedra are built (Figure 1c) by the sequential insertion algorithm (see 'Materials and Methods' section). The Laguerre tessellation is constructed simultaneously as the geometric dual of the Delaunay diagram (Figure 1d). Then, all the properties of interest are computed by VLDP program (Figure 1e), e.g. volume and face area of Laguerre

polyhedra, inter-atomic and inter-residue contacts, graphical outputs and so forth.

Output

Once the job is finished, the home page is updated to show the results. The results are given in various ways. The first information, on the top of the web page, is a global summary on the system under investigation: count of residues, count of protein atoms and the selected protein chain label. The data are confidentially stored during 2 months for later use. Confidentiality is guaranteed by an ID supplied to the user. Three links point to the data available for download: the first place contains a copy of the submitted pdb file; the second one has the solvated protein (pdb format) and the third link is a zip file containing all results.

The protein (29) is displayed online through three nice graphical representations generated with PyMol software: cartoon, Connolly's surface and Laguerre surface (see Figure 2a and b). Each image can be zoomed by clicking on it. Another graph displays the inter-residue contact matrix with colour scale representing the contact strength (number of atomic contacts for one given residue, see Figure 2c); the raw values can also be downloaded in a text file.

The 'volume and area' section shows residue volumes (and area) in a partial table. The complete data can be downloaded as a text file. It also gives the volume per protein atom and the volume per water molecule at the solute-solvent interface.

The second plot on the Result web page is the PIA accessibility along the sequence. Again, a partial table is displayed, whereas the complete table is available for download.

Finally, an interactive visualization can be performed through the Jmol Applet (25). The Laguerre surface appears by clicking on the 'Show Tessellation' button. Several levels of transparency are proposed: 0, 25, 50 and 75%. Some other options are available like choosing a representation (Trace, Backbone or Cartoon) or assigning colours (to atoms with Corey-Pauling-Koltun convention (CPK), Amino group or Secondary Structure). The user can also display the solvent and the surface (molecular or Solvent-Accessible Surface (SAS)). The surfaces are coloured according to the secondary structure on the white-red-blue scale.

An example

Figure 2 illustrates the results of a protein structure analysis of a fibrillarin homologue [PDB code 1FBN (29)], using Laguerre tessellation. As a confidence descriptor, the Laguerre surface is visually compared with Connolly's surface (Figure 2b). To study the protein folds, the connectivity between residues is shown as a contact matrix indicating the strength of each inter-residue contact (Figure 2c). The contacts along the diagonal indicate, in general, a helix pattern, whereas off-diagonal contacts, parallel or perpendicular to the main diagonal, indicate parallel or anti-parallel strands (most often in β-sheets) (30).

The residue volumes given by VLDPws can be compared with our previous study (8) (Figure 2d). Finally the surface versus core distribution of the residues can be probed by the accessibility analysis through the PIA measure (Figure 2e). As we can see along the sequence on the barplot, the fibrillarin homologue has regions of low PIA values, corresponding to buried regions, and high PIA regions which are more exposed.

Implementation

The program VLDP is written in Fortran 90. The graphical plots are done using R software, version 2.15 (http://cran.r-project.org/). The PyMOl scripts are written in Python. The front-end interface is based on html and php. Perl/cgi programs control the input while VLDP and other programs carry out the processing behind.

CONCLUSION

Very few web servers are dedicated to the analysis of protein structures using topological graph representation and offering several analysis methods. We propose an original tool that combines metric and topological information. We chose to study mostly the properties of 3D protein structures. Three main characteristics are extracted and given in details: residue volumes for the packing, residue contacts for the protein folds and residue exposure to solvent for accessibility. For expert users interested in hydration, the server also provides information on the water network organized in successive layers. As a criterion of confidence, the results can be compared with our previous studies (8,22), specially for the tabulated volumes. VLDPws brings also an opportunity to compare novel definitions of contacts and accessibility with standard methods. The availability of our method through VLDPws will serve structural biology and inspire new experiments on local or global properties of proteins.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Taylor,W.R. (1986) The classification of amino acid conservation. *J. Theor. Biol.*, **119**, 205–218.
2. Bhattacharyya,R., Pal,D. and Chakrabarti,P. (2004) Disulfide bonds, their stereospecific environment and conservation in protein structures. *Protein Eng. Des. Sel.*, **17**, 795–808.
3. Weiss,M.S., Brandl,M., Suhnel,J., Pal,D. and Hilgenfeld,R. (2001) More hydrogen bonds for the (structural) biologist. *Trends Biochem. Sci.*, **26**, 521–523.
4. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
5. Joosten,R.P., te Beek,T.A., Krieger,E., Hekkelman,M.L., Hooft,R.W., Schneider,R., Sander,C. and Vriend,G. A series of PDB related databases for everyday needs. *Nucleic Acids Res.*, **39**, D411–D419.
6. Frishman,D. and Argos,P. (1995) Knowledge-based protein secondary structure assignment. *Proteins*, **23**, 566–579.
7. Heinig,M. and Frishman,D. (2004) STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.*, **32**, W500–W502.
8. Esque,J., Oguey,C. and de Brevern,A.G. (2010) A novel evaluation of residue and protein volumes by means of Laguerre tessellation. *J. Chem. Inf. Model.*, **50**, 947–960.
9. Poupon,A. (2004) Voronoi and Voronoi-related tessellations in studies of protein structure and interaction. *Curr. Opin. Struct. Biol.*, **14**, 233–241.
10. Ismer,J., Rose,A.S., Tiemann,J.K., Goede,A., Rother,K. and Hildebrand,P.W. (2012) Voronoia4RNA–a database of atomic packing densities of RNA structures and their complexes. *Nucleic Acids Res.*, **41**, D280–D284.
11. Mahdavi,S., Mohades,A., Salehzadeh Yazdi,A., Jahandideh,S. and Masoudi-Nejad,A. (2012) Computational analysis of RNA-protein interaction interfaces via the Voronoi diagram. *J. Theor. Biol.*, **293**, 55–64.
12. Nadassy,K., Tomas-Oliveira,I., Alberts,I., Janin,J. and Wodak,S.J. (2001) Standard atomic volumes in double-stranded DNA and packing in protein–DNA interfaces. *Nucleic Acids Res.*, **29**, 3362–3376.
13. Echols,N., Milburn,D. and Gerstein,M. (2003) MolMovDB: analysis and visualization of conformational change and structural flexibility. *Nucleic Acids Res.*, **31**, 478–482.
14. Birzele,F., Gewehr,J.E., Csaba,G. and Zimmer,R. (2007) Vorolign–fast structural alignment using Voronoi contacts. *Bioinformatics*, **23**, e205–e211.
15. Petrek,M., Kosinova,P., Koca,J. and Otyepka,M. (2007) MOLE: a voronoi diagram-based explorer of molecular channels, pores, and tunnels. *Structure*, **15**, 1357–1363.
16. Berka,K., Hanak,O., Sehnal,D., Banas,P., Navratilova,V., Jaiswal,D., Ionescu,C.M., Svobodova Varekova,R., Koca,J. and Otyepka,M. (2012) MOLEonline 2.0: interactive web-based analysis of biomacromolecular channels. *Nucleic Acids Res.*, **40**, W222–W227.
17. Bernauer,J., Bahadur,R.P., Rodier,F., Janin,J. and Poupon,A. (2008) DiMoVo: a Voronoi tessellation-based method for discriminating crystallographic and biological protein-protein interactions. *Bioinformatics*, **24**, 652–658.
18. Faure,G., Bornot,A. and de Brevern,A.G. (2008) Protein contacts, inter-residue interactions and side-chain modelling. *Biochimie*, **90**, 626–639.
19. Gelly,J.C. and de Brevern,A.G. (2011) Protein Peeling 3D: new tools for analyzing protein structures. *Bioinformatics*, **27**, 132–133.
20. Gelly,J.C., de Brevern,A.G. and Hazout,S. (2006) 'Protein Peeling': an approach for splitting a 3D protein structure into compact fragments. *Bioinformatics*, **22**, 129–133.
21. Gelly,J.C., Etchebest,C., Hazout,S. and de Brevern,A.G. (2006) Protein Peeling 2: a web server to convert protein structures into series of protein units. *Nucleic Acids Res.*, **34**, W75–W78.
22. Esque,J., Oguey,C. and de Brevern,A.G. (2011) Comparative analysis of threshold and tessellation methods for determining protein contacts. *J. Chem. Inf. Model.*, **51**, 493–507.
23. Gutow,J.H. (2010) Easy jmol web pages using the jmol export to web function: A tool for creating interactive web-based instructional resources and student projects with live 3-D images of molecules without writing computer code. *J. Chem. Educ.*, **87**, 652–653.
24. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
25. Hess,B., Kutzner,C., van der Spoel,D. and Lindahl,E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.
26. Kaminski,G.A., Friesner,R.A., Tirado-Rives,J. and Jorgensen,W.L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, **105**, 6474–6487.
27. Grubmuller,H., Heymann,B. and Tavan,P. (1996) Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, **271**, 997–999.
28. Tsai,J., Taylor,R., Chothia,C. and Gerstein,M. (1999) The packing density in proteins: standard radii and volumes. *J. Mol. Biol.*, **290**, 253–266.
29. Wang,H., Boisvert,D., Kim,K.K., Kim,R. and Kim,S.H. (2000) Crystal structure of a fibrillarin homologue from methanococcus jannaschii, a hyperthermophile, at 1.6 A resolution. *EMBO J.*, **19**, 317–323.
30. Sadoc,J.F., Jullien,R. and Rivier,N. (2003) The Laguerre polyhedral decomposition: application to protein folds. *Eur. Phys. J. B*, **33**, 355–363.