webPIPSA: a web server for the comparison of protein interaction properties

Stefan Richter[1],*, Anne Wenzel[1], Matthias Stein[1], Razif R. Gabdoulline[1,2] and Rebecca C. Wade[1]

[1]Molecular and Cellular Modeling Group, EML Research gGmbH, Schloss-Wolfsbrunnenweg 33, 69118 and [2]BIOMS (Center for Modeling and Simulation in the Biosciences), University of Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany

ABSTRACT

Protein molecular interaction fields are key determinants of protein functionality. PIPSA (Protein Interaction Property Similarity Analysis) is a procedure to compare and analyze protein molecular interaction fields, such as the electrostatic potential. PIPSA may assist in protein functional assignment, classification of proteins, the comparison of binding properties and the estimation of enzyme kinetic parameters. webPIPSA is a web server that enables the use of PIPSA to compare and analyze protein electrostatic potentials. While PIPSA can be run with downloadable software (see http://projects.eml.org/mcm/software/pipsa), webPIPSA extends and simplifies a PIPSA run. This allows non-expert users to perform PIPSA for their protein datasets. With input protein coordinates, the superposition of protein structures, as well as the computation and analysis of electrostatic potentials, is automated. The results are provided as electrostatic similarity matrices from an all-pairwise comparison of the proteins which can be subjected to clustering and visualized as epograms (tree-like diagrams showing electrostatic potential differences) or heat maps. webPIPSA is freely available at: http://pipsa.eml.org.

INTRODUCTION

The interactions of biological macromolecules are critical to their physiological function and dependent on their molecular interaction fields. The electrostatic potential is a molecular interaction field of particular importance for determining the specificity and kinetics of molecular binding. PIPSA (Protein Interaction Property Similarity Analysis) (1) may be used to classify a large number of proteins according to the similarities or dissimilarities in

their 3D molecular interaction fields, such as the electrostatic potential (2,3). Some of the applications of PIPSA have been to WW domains (4), electron transfer proteins (5), ubiquitin conjugating enzymes (6), complement control protein modules (7) and dihydrofolate reductases (8). An extension of PIPSA is qPIPSA (quantitative PIPSA) (9). qPIPSA enables the kinetic parameters of a set of enzymes sharing the same function to be related to the molecular interaction field, e.g. the electrostatic potential, at a functional region of the protein. Such a comparison may enable estimation of unknown kinetic parameters for an enzymatic reaction and thereby assist in the modeling and simulation of biochemical pathways (9,10).

webPIPSA allows the user to perform PIPSA to compute and compare the electrostatic potentials of a set of structurally related proteins. Other web servers such as PCE (http://bioserv.rpbs.jussieu.fr/PCE) (11) and PFplus (http://pfp.technion.ac.il) (12) also allow the calculation of protein electrostatic potentials. PFplus was designed to extract and display the largest positive electrostatic patch on a protein surface. These web servers, however, only allow the calculation of electrostatic potentials for single proteins. webPIPSA on the other hand permits calculation of the electrostatic potentials of a large number of proteins and performs an all-versus-all pairwise comparison of the electrostatic potentials around the entire protein and, optionally, over a user-predefined region.

webPIPSA can first superimpose the protein structures by a least-squares fitting procedure. The electrostatic potentials are then computed on a grid by solution of the linearized finite-difference Poisson–Boltzmann equation using the UHBD (13) or APBS (14) software. The similarity or dissimilarity of the electrostatic potential of each pair of proteins in the dataset is quantified for a user-defined region by means of similarity indices and distance measures. The proteins can be clustered according to the relations between their electrostatic potentials. The results are displayed as an epogram (tree-like diagram based on electrostatic differences) and a colored matrix view. The results can be used for the classification of the proteins

---

*To whom correspondence should be addressed. Tel: +49 6221 533 279; Fax: +49 6221 533 298; Email: stefan.richter@eml-r.villa-bosch.de

according to their interaction properties or for the correlation of the molecular interaction fields with functional properties of the proteins.

webPIPSA USAGE

The input required for using webPIPSA is a set of protein coordinate files in PDB format. These can be either (i) user-supplied or (ii) specified by PDB identifier code in the RCSB. The user-supplied structures can be either experimentally determined or generated by comparative modeling techniques using, for example, MODELLER (15) or SWISSMODEL (16) or structures from the respective databases, such as MODBASE (17). An example dataset consisting of triosephosphate isomerases from various species (9) is provided on the web server and described in detail below.

After the structures have been uploaded, the user can choose whether to use the UHBD (13) or the APBS (14) software to calculate the electrostatic potentials. Example input files for UHBD and APBS are given in the online documentation for webPIPSA. At this point, the user is asked to provide an email address to be notified when the PIPSA calculation has been completed. The user receives an email with a link to the results page displaying the heat map and a clustered epogram. The email also provides a link to the results directory on the web server containing intermediate and additional results such as the potential grid files and the complete PIPSA output. A description of how to visualize electrostatic potential grids from the results directory is given in the online documentation.

Figure 1 shows part of the results pages for the example case. Triosephosphate isomerases from 12 different species are clustered according to the all pairwise distances between electrostatic potentials using a color code from red (small distance) to blue (large distance). Proteins with similar electrostatic properties are clustered on the top and left side of the graph. The proteins are assigned to a user-defined number of clusters according to their electrostatic potential similarities. The number of clusters can be specified by selecting from a drop down box at the bottom of the result page. Here, triosephosphate isomerases from 12 species form four subclusters in a comparison of the electrostatic potentials around the whole protein surface. The type of distance measure [Hodgkin similarity index (18), Carbo similarity index or average potential differences (9)] displayed in the heat map can also be selected from a dropdown box. The distance matrix may also be viewed with the proteins in their input order, without clustering.

webPIPSA WORKFLOW

An illustration of the workflow in webPIPSA is given in Figure 2. After the upload of the protein structures is complete, the user can choose to superimpose the structures using one of two different methods. The first method, called 'sup2pdb', starts with one structure (the template) and subsequently does a pairwise sequence alignment (19) between this template structure and the remaining coordinate files. The structures are then superimposed according to the respective alignments. Since the outcome of this algorithm depends on the choice of protein template, in the second option, 'optimized sup2pdb', every coordinate file is considered as a template for superimposing the others. Then the most successful superposition based on the maximum number of matched structures and minimal root mean square deviation (RMSD) of the superimposed structures is selected. For a large number of protein structures, however, this second option can be time consuming.

The next step in the workflow is to add polar hydrogen atoms to the protein structures using WHATIF (20). Hydrogen atom positions are assigned by optimizing the hydrogen-bond network. Standard protonation states at pH 7 are assumed for all residues except for histidine which can be treated as singly or doubly protonated.

For the superimposed set of coordinate files, electrostatic potentials are computed automatically using a set of default parameters assuming an ionic strength of 50 mM and a temperature of 300 K. UHBD or APBS can be used to solve the linearized Poisson–Boltzmann equation (LPBE) treating the protein as a low dielectric with partial atomic charges embedded in a homogeneous high dielectric continuum representing the solvent.

The electrostatic potentials of all the structures are then compared using PIPSA. First, the potentials in the complete protein surface skins are compared. The protein surface is defined by using a probe of radius 2 Å. The skin extends out from the protein surface with a thickness of 3 Å. Additionally, the user can specify spherical regions within which the electrostatic potentials in the skin are compared. A spherical region can, for example, encompass an enzyme's active site. The Cartesian coordinates of the center of the sphere and its radius should be input by the user. This option can therefore only be used when the uploaded input structures have already been superimposed. Several spherical protein regions can be specified and compared separately. Each region is given a name for later identification.

The PIPSA software is used to calculate Hodgkin and Carbo similarity indices of the protein electrostatic potentials as well as average electrostatic potential differences (the difference in electrostatic potentials in $kcal \, mol^{-1} \, e^{-1}$ of two proteins divided by the number of grid points in the comparison region where the two protein skins overlap) described in (9). The similarity indices range from -1 (anti-correlated potential), through 0 (uncorrelated) to $+1$ (identical potentials). These values are converted into distances given by $sqrt(2 - 2^*SI)$ where SI is the respective similarity index. These distances thus range from 0 (identical) to 2 (anti-correlated potentials). The clustering analysis and generation of tree-like diagrams (epograms) on the results page are performed using the statistical program R (21). For visualization purposes, the resulting distance matrix is presented as a heat map and as an epogram. Both the heat map and epogram representations allow the fast identification of inter-protein relations and classifications for a large set of proteins.
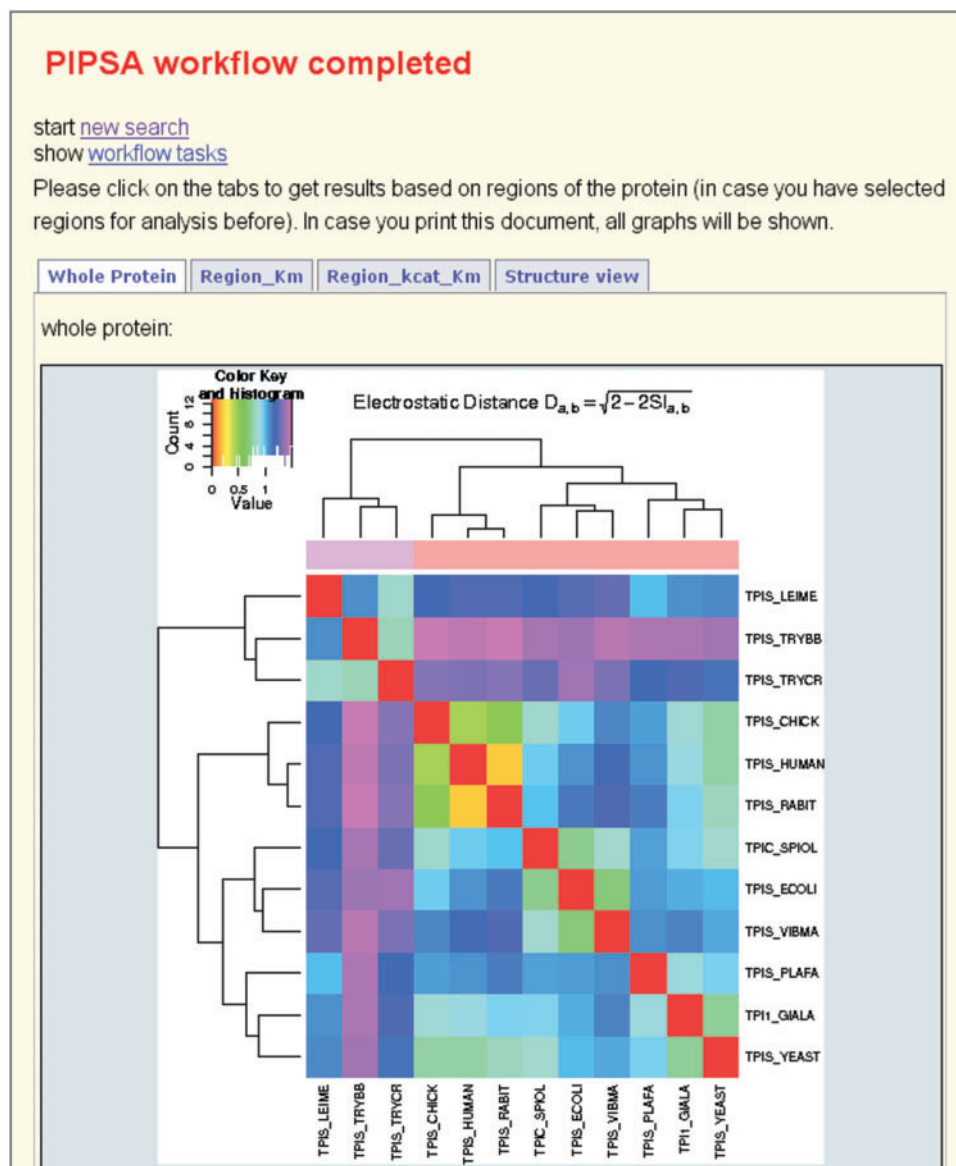
Figure 1. Screenshot of part of the results page for the example provided on the webPIPSA web server for the comparison of the electrostatic potentials of triosephosphate isomerases from 12 different species. The colored matrix (heat map) is shown with color coding corresponding to the distances calculated from the Hodgkin similarity indices for the electrostatic potentials. Enzymes from species with highly similar electrostatic potentials, such as human, rabbit and chicken, are clustered together (second subcluster from top/left).

The results are presented on a series of tabs. One tab is given for the analysis using the entire protein skin and additional tabs are used for each spherical region specified for analysis by the user. A further tab has a Java applet with the Astex^{TM} viewer (22) to allow the user to visually check the superposition of the protein structures as well as the positioning of the spherical regions used for focused comparisons of the electrostatic potentials.

webPIPSA EXAMPLE

As an example, the dataset studied in reference (9) is provided on the web server for download. These structures of triosephosphate isomerase (TPI) from 12 different species are already superimposed (five of them are shown in Figure 2). The example shows how their electrostatic potentials in the vicinity of the active site can be analysed and related to enzymatic kinetic parameters. Electrostatic interactions contribute to ligand–enzyme binding and to enzyme catalysis. For example, they have been shown to affect the rate of ligand binding, the affinity of ligand binding and the stability of the transition state. Whether a relation between electrostatic potential differences and kinetic parameters or binding affinities exists in a particular case under study depends on the relative contribution of electrostatic interactions to these quantities and the consistency of the structural and the kinetic or thermodynamic experimental data.

After the upload of the protein coordinates using an applet, one can request comparison of the electrostatic potentials in specific regions. For the triosephosphate isomerase example, regions may be selected as follows
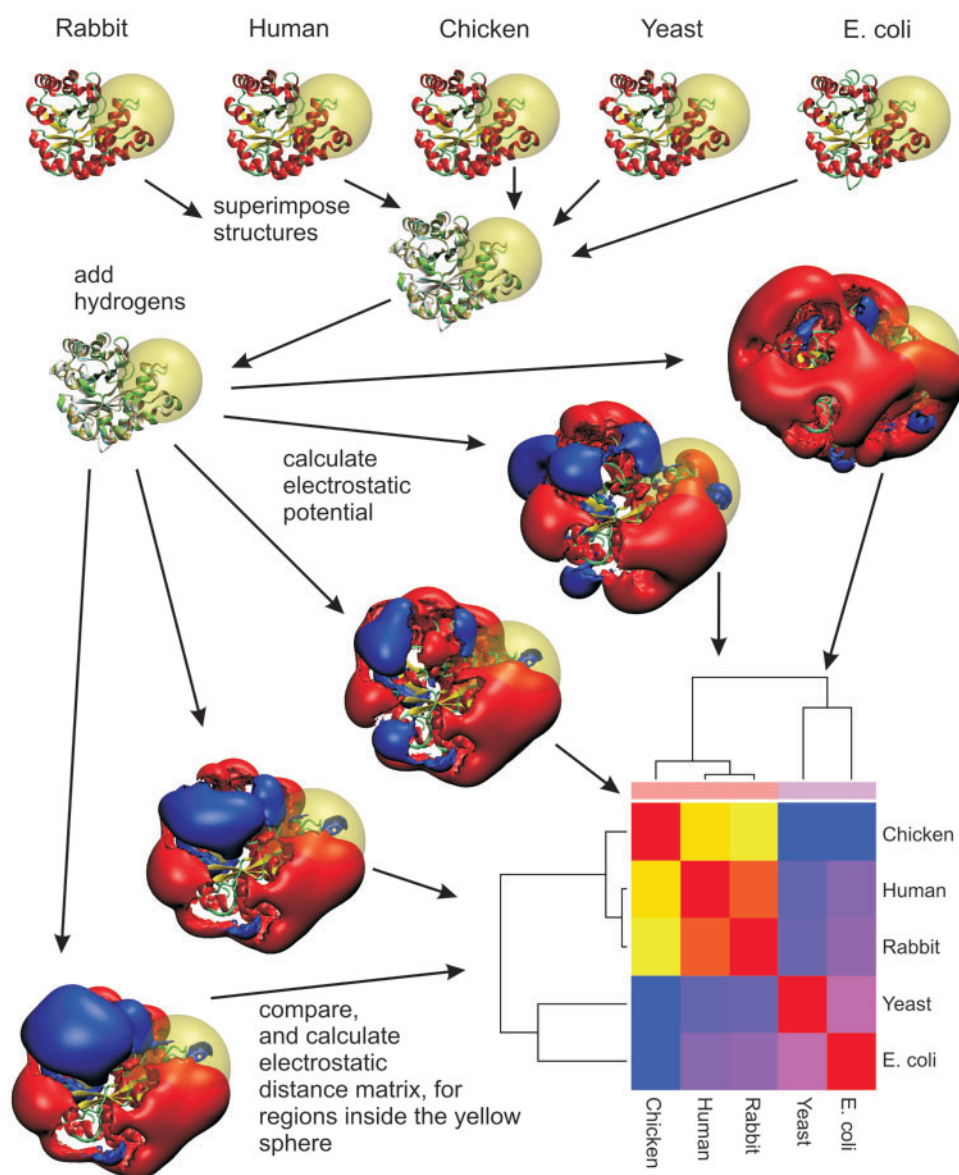
Figure 2. Schematic overview of the workflow employed in webPIPSA. The protein structures are a subset of the triosephosphate isomerases in the example analysis provided on the web server. The region for comparison of the electrostatic potential can be selected after upload of the coordinate files (the substrate binding site in the example given). Polar hydrogen atoms are added to the coordinate files using the WHATIF program. Electrostatic potential grids are calculated using the UHBD or APBS software. PIPSA is used to compare the electrostatic potentials and to calculate a distance matrix. These distances are used to cluster the proteins according to the relations between their electrostatic potentials and the clustering is displayed in a tree-like diagram, an epogram.

(with the Cartesian coordinates of the center and the radius of a sphere in Å given):

(i) the substrate binding site (Region_Km: 1.07, 4.06, 21.11, 15) where the electrostatic potential correlates with substrate K_m values, and

(ii) the active site (Region_kcat_Km: 8.15, 3.54, 34.83, 15) where the catalytic turnover occurs and the electrostatic potential influences enzymatic k_{cat}/K_m values (9).

The electrostatic potentials are computed for the entire proteins using UHBD. For restricting the region of comparison in case (i), a spherical region of radius 15 Å around W168 is chosen (shown as a transparent yellow sphere in Figure 2). This residue is part of the flexible loop region of the TPIs which closes over the active site when the substrate binds. A comparison of the similarity of the electrostatic potentials (measured by the Hodgkin SI) in this region for the five model proteins is shown on the bottom right corner of Figure 2. The five enzymes form two subclusters: TPIs from chicken, human and rabbit form one subcluster, whereas yeast and *Escherichia coli* form a distinct second subcluster. This is in agreement with the different substrate K_m values for the two

subclusters: TPIs from chicken, human and rabbit possess lower K_m values (0.47 mM, 0.49 mM and 0.43 mM, respectively) and yeast and *E. coli* exhibit significantly higher K_m values (1.22 mM and 1.03 mM, respectively). For a thorough discussion, see ref. (9). Thus, the electrostatic potential differences in the TPI loop region can be used as a fingerprint for classifying TPIs from various species. Such a PIPSA analysis can provide insights that are complementary to and not apparent from a protein sequence-based analysis.

TECHNICAL OVERVIEW

webPIPSA is implemented using Java servlets and Java server pages (JSP) running on tomcat 5 (http://tomcat.apache.org). The workflows with significant computational components are implemented as ant scripts (http://ant.apache.org) and are launched from a Java messaging server (JMS, http://activemq.apache.org). This architecture allows the separation of the tomcat web server from the computationally demanding workflows. The user is informed about the current state of the workflow by messages sent from the ant script via the messaging server. All data are stored on the file system and may be removed after 2 weeks without further notice.

CONCLUSIONS AND OUTLOOK

Currently, webPIPSA provides a description and categorization of the electrostatic potential differences between the input protein structures. It does not include all the features of the downloadable PIPSA software which can be used to analyze other types of molecular interaction field and to select conical as well as spherical regions. These features are planned to be added in future developments of webPIPSA. In addition, it is planned to add tools to webPIPSA to enable the user to study the relations between protein molecular interaction fields and enzymatic kinetic parameters, as in qPIPSA (9). On the other hand, webPIPSA is much more user-friendly and, therefore, accessible to non-experts. It also has additional analysis features such as the simultaneous display of colored heat maps and epograms as well as protein structure visualization with the Astex^{TM} viewer.

Feedback concerning webPIPSA is encouraged and should be sent to mcmsoft@eml-r.villa-bosch.de.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Blomberg,N., Gabdoulline,R.R., Nilges,M. and Wade,R.C. (1999) Classification of protein sequences by homology modeling and quantitative analysis of electrostatic similarity. *Proteins*, **37**, 379–387.
2. Wade,R.C., Gabdoulline,R.R. and De Rienzo,F. (2001) Protein interaction property similarity analysis. *Int. J. Quant. Chem.*, **83**, 122–127.
3. Wade,R.C. (2005) In Cruciani,G. (ed.), *Molecular Interaction Fields. Applications in Drug Discovery and ADME Prediction*. Wiley-VCH, Weinheim, Vol. 2, pp. 27–42.
4. Schleinkofer,K., Wiedemann,U., Otte,L., Wang,T., Krause,G., Oschkinat,H. and Wade,R.C. (2004) Comparative structural and energetic analysis of WW domain-peptide interactions. *J. Mol. Biol.*, **344**, 865–881.
5. De Rienzo,F., Gabdoulline,R.R., Menziani,M.C., De Benedetti,P.G. and Wade,R.C. (2001) Electrostatic and Brownian dynamics simulation analysis of plastocyanin and cytochrome f. *Biophys. J.*, **81**, 3090–3104.
6. Winn,P.J., Religa,T.L., Battey,J.N., Banerjee,A. and Wade,R.C. (2004) Determinants of functionality in the ubiquitin conjugating enzyme family. *Structure*, **12**, 1563–1574.
7. Soares,D.C., Gerloff,D.L., Syme,N.R., Coulson,A.F.W., Parkinson,J. and Barlow,P.N. (2005) Large-scale modelling as a route to mutliple surface comparisons of the CCP module family. *Protein Eng. Design Selection*, **18**, 379–388.
8. Henrich,S., Richter,S. and Wade,R.C. (2008) On the use of PIPSA to guide target-selective drug design. *Chem. Med. Chem.*, **3**, 413–417.
9. Gabdoulline,R.R., Stein,M. and Wade,R.C. (2007) qPIPSA: relating enzymatic kinetic parameters and interaction fields. *BMC Bioinformatics*, **8**, 373.
10. Stein,M., Gabdoulline,R.R. and Wade,R.C. (2007) In Hicks,M.G. and Kettner,C. (eds), *Proceedings of the 2nd Beilstein Workshop*. Logos Verlag, Berlin, pp. 237–253.
11. Miteva,M.A., Tuffery,P. and Villoutreix,B.O. (2005) PCE: web tools to compute protein continuum electrostatics. *Nucleic Acids Res.*, **33**, W372–W375.
12. Stawiski,E.W., Gregoret,L.M. and Mandel-Gutfreund,Y. (2003) Annotating nucleic acid-binding function based on protein structure. *J. Mol. Biol.*, **326**, 14.
13. Madura,J.D., Briggs,J.M., Wade,R.C., Davis,M.E., Luty,B.A., Ilin,A., Antosiewicz,J., Gilson,M.K., Bagheri,B., Scott,L.R. *et al.* (1995) Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. *Comput. Phys. Commun.*, **1995**, 57–95.
14. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
15. Sali,A. and Blundell,T.L. (1993) Comparitive protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.
16. Schwede,T., Kopp,J., Guex,N. and Peitsch,M.C. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.*, **31**, 3381–3385.
17. Pieper,U., Eswar,N., Davis,F.P., Braberg,H., Madhusudhan,M.S., Rossi,A., Marti-Renom,M., Karchin,R., Webb,B.M., Eramian,D. *et al.* (2006) MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.*, **34**, D291–D295.
18. Hodgkin,E.E. and Richards,W.G. (1987) Molecular similarity based on electrostatic potential and electric field. *Int. J. Quant. Chem. Quant. Biol. Symp.*, **14**, 105–110.
19. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
20. Vriend,G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 52–56.
21. Ihaka,R. and Gentleman,R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.
22. Hartshorn,M.J. (2002) AstexViewer (TM): a visualisation aid for structure-based drug design. *J. Comput.-aided Mol. Des.*, **16**, 871–881.