

Molecular Modeling Information Transfer with VRML: From small Molecules to large Systems in Bioscience.

G. Moeckel^{1,2}, M. Keil¹, T. Exner¹, and J. Brickmann^{1,3}

¹*Physical Chemistry I, Darmstadt University of Technology, Darmstadt, Germany*

²*Division of Toxicology & Cancer Risk Factors, German Cancer Research Center,
Heidelberg, Germany*

³*Darmstadt Center of Scientific Computing, Darmstadt, Germany*

The suitability of the Virtual Reality Modeling Language (VRML) for the communication of scientists via the internet is demonstrated with recent results from computer assisted cancer research:

I. Substrate channels in cytochrome P450 enzymes.

II. Binding properties of the *wild type* and mutated p53 tumor suppressor protein.

Complex 3D molecular models were used to visualize new insights in the active site access of cytochrome P450 enzymes and in the p53 protein-DNA binding achieved by the use of computational methods. These 3D models of biomolecular systems were transferred into VRML scenarios. Additional implemented features allow users to receive related information interactively. With these examples it is shown that VRML provides an efficient method for scientific information exchange by the use of complex 3D molecular models.

1 Introduction

The investigation of protein functions based on structure and property analysis requires three-dimensional (3D) molecular models, which become more complex with the increase of computational technology on one hand and experimental results on the other hand. Especially the visualization of molecular properties together with molecular structures in a 3D scenario offers new insights towards an understanding of molecular interactions. These complex biomolecular information of 3D models can be transferred via internet or intranet with the Virtual Reality Modeling Language (VRML) [1,2]. In this way, computational scientists can communicate with other researchers, e.g. experimenters using the appropriate 3D molecular models.

The suitability of VRML for 3D biomolecular information transfer should be demonstrated interactively with recent results from two projects in computer assisted cancer research:

I. Substrate channels in cytochrome P450 enzymes.

II. Binding properties of the *wild type* and mutated p53 tumor suppressor protein.

These investigations are conducted in a close cooperation with H. Bartsch and his group at the Division of Toxicology and Cancer Risk Factors of the German Cancer Research Center in Heidelberg, Germany.

2 Molecular models in biochemistry

The increasing complexity of biomolecular systems as well as the availability of additional molecular information requires an extended molecular model compared with models used for small molecules. Representations of biomolecular systems like proteins or protein complexes with thousands of atoms have to be reduced in complexity in order to get insights. One possibility to show related molecular information is the use of the concept of *molecular surfaces*, first introduced by Richards [3] and modified by Connolly [4] to describe the molecular envelope accessible by a solvent molecule. The concept of molecular surfaces is obvious artificial, but the visualization of the 3D shape of a molecule is useful for understanding interactions of molecules. These so called *solvent accessible surfaces* are employed successfully in many biomolecular investigations e.g. docking algorithms [5], free energy approach [6] and visualization of molecular properties [7] e.g. the electrostatic potential as well as lipophilicity.

2.1 Molecular properties

In our investigations, the electrostatic potential in solution was calculated by the use of a finite difference algorithm to solve the Poisson-Boltzmann equation [8]. A smoothing scheme for the dielectric values of the solvent and the solute was introduced to achieve a better approximation of the transition from the molecular interior to the solvent [9].

In order to determine local hydrophilic/lipophilic properties we used a method based on hydrophobic atomic partial values [10]. As an extension for the investigation of proteins our structure analysis incorporates charged amino acid fragments and accounts for their hydrophobic partial values [11]. After a connectivity analysis to characterize the atoms in their individual structural environment, the local hydrophilic and lipophilic values were calculated from the atomic partial values and mapped onto the *solvent accessible surface*. By the use of a special weighting procedure [12], it is guaranteed, that only those atomic fragments that are in close neighborhood to a surface point contribute significantly to the surface value of lipophilicity (see Fig. 1).

3 Biomolecular information transfer with VRML

Recent developments in computer technology as well as the increasing number of biomolecular structure data enable scientists to create large model systems with specialized biomolecular information. With the development of VRML (<http://www.vrml.org/>), it is now possible to transfer these 3D scenarios directly

via the internet and to examine it with standard internet browsers [1,2,13]. In this way, computational scientists can communicate with other researchers e.g. experimenters using the appropriate 3D molecular models.

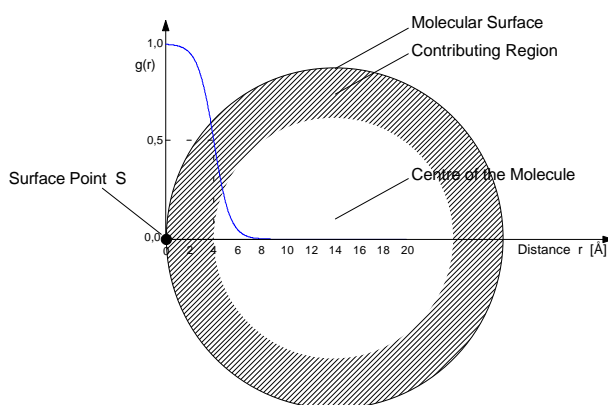


Fig. 1: Contribution to lipophilicity: The lipophilicity value of the surface point S of a globular molecule is determined by atomic partial values of the contributing region. The distance dependency $g(r)$ of the local lipophilicity is shown by the function graph.

VRML is based on a subset of the Open Inventor ASCII File Format [14] of Silicon Graphics, Inc. extended by networking capabilities, such as World Wide Web (WWW) hyperlinks. With these networking capabilities VRML is equivalent to the Hypertext Markup Language (HTML). Like HTML documents specify the layout of pages on the internet, VRML files describe single objects within a scene or even whole scenarios in the 3D space.

VRML provides an object oriented method for the description of these 3D scenarios. They are build of basic elements which are called nodes. There are different categories of nodes affecting different aspects of the scenes. Some of them represent the shape and properties of single objects, others define the point of view or lighting of the scene, etc. All nodes of a scene are arranged in a hierarchical structure called *scene graph*. This *scene graph* defines the ordering of the nodes and therefore the layout of the 3D scene. Analogous to HTML it is possible to build up complex 3D molecular models consisting of linkeded VRML files of single components. However, only static scenarios can be generated with VRML 1.0. This hindrance was overcome with the introduction of VRML 2.0 (<http://vrml.sgi.com/>), supplying additional capabilities, e.g. for the animation of the scenarios.

The generation of VRML scenarios can be done with simple text editors or VRML author tools like Cosmo Worlds (TM) from Silicon Graphics Inc., which

simplifies the creation of VRML scenes considerably. But there are also many tools converting other 3D file formats into VRML scenes. We used the program PDB2VRML (<http://www.pc.chemie.th-darmstadt.de/vrml/pdb2vrml>) developed in our group to convert files from the Brookhaven Protein Data Bank (<http://www.pdb.bnl.gov>) into the VRML file format. PDB2VRML supports several different molecular representation types e.g. wireframe, ball&stick, CPK models, ribbons, etc. A schematic representation of the generation of VRML scenarios from 3D structures is shown in Fig. 2.

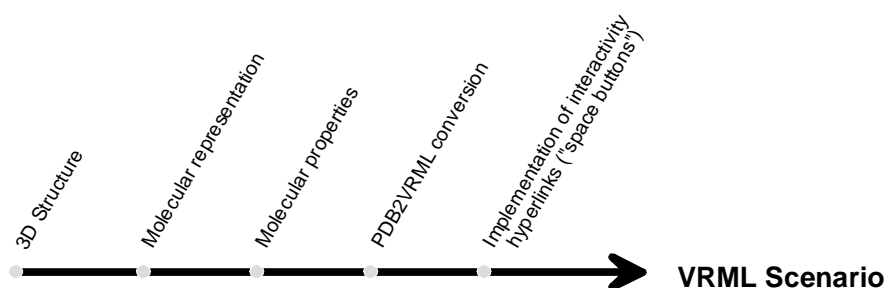


Fig. 2: Generation of VRML scenarios

Furthermore, VRML provides so called WWW anchors (analogues to the anchors in HTML) to build up complex scenarios consisting of linked VRML files. This selectable links within a VRML scene can lead to other VRML models or HTML documents.

In our VRML applications, the molecular models are hyperlinks, which enable the users to choose between different representations of the molecular scenario. Additionally, we have implemented simple geometric objects e.g. cubes, balls, cylinders, cones, etc. as *space buttons* in a VRML scene to provide users with textual information or to change between levels of complexity. In such scenarios the user can interactively choose regions of his own interest.

4 Applications from computer assisted cancer research

The induction of cancer is a multistage process, which can be initiated by exogeneous or endogeneous compounds. We present here recent results of two applications from computer assisted cancer research in tumorigenesis, but these investigations may have implications to other research fields, e.g. drug metabolism or mutation analysis. In the first example we show a comparison of the substrate channels to the active site of cytochrome P450 enzymes, which are responsible for

the metabolic activation of a variety of compounds. In the second example we present results from our analysis of *wild type* and mutated p53 tumor suppressor protein complexed with DNA (see Fig. 3).

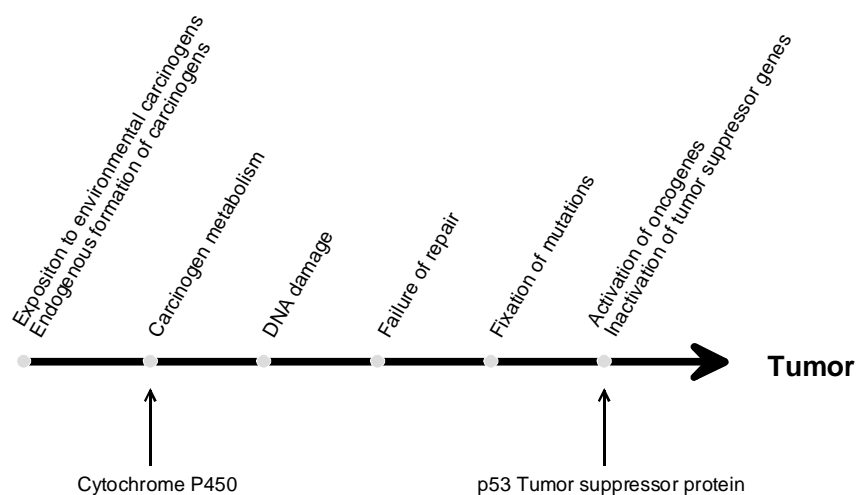


Fig. 3: Scheme of the multistage carcinogenesis process with the function of cytochrome P450 and the tumor suppressor protein p53 (modified from [15]).

The visualization of these results using complex molecular models and the transport via internet with VRML may highlight the capabilities of these techniques for the communication between research groups.

4.1 Substrate channels of cytochrome P450 enzymes.

Cytochrome P450 (P450) enzymes [16] play a central role in most phase I metabolisms by oxidizing a variety of substrates, both endogenous and exogenous origin. Many carcinogens require metabolic activation by P450 enzymes to exert mutagenic or carcinogenic effects. In the case of carcinogenic nitrosamines the P450 mediated oxidation initiates a reaction cascade, which can lead to alkylation of DNA and finally to tumor growth [17]. The substrate oxidation takes place at the active site located in the center of the almost spherical P450 enzymes. To enter (and to leave) the binding site, substrates have to pass through a channel leading from the enzymes surface to the buried active site. Because of its length (about 15 Å) the channel seems to play an essential role in access of a large number of substrates by different P450 enzymes.

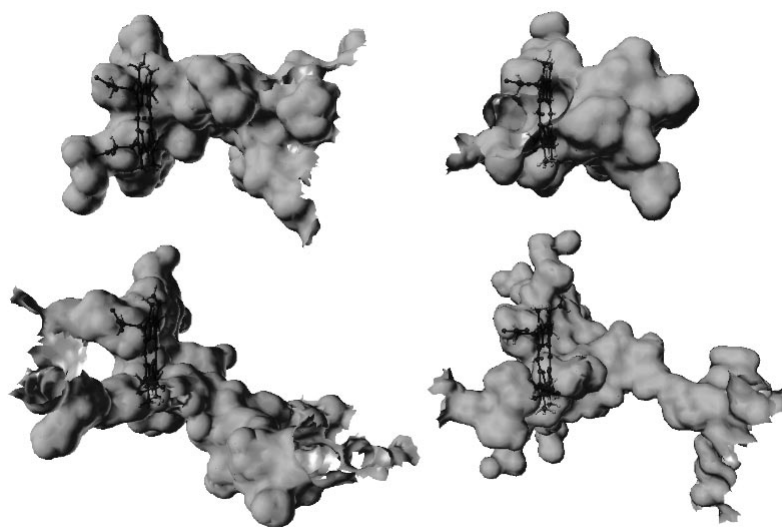


Fig 4: Molecular surfaces of the access channels to the active site of P450 enzymes with known 3D structure: P450_{BM3} (above, left), P450_{eryF} (above, right), P450_{terp} (below, left), and the active-site-access-model of P450_{cam} (below, right). The central heme at the active site is visible through the transparent surface. The aperture of the substrate channel (except for P450_{eryF}) is located at the right side.

In our project, ‘Substrate Channels in Cytochrome P450 Enzymes’, we have investigated the access to the active site of different P450’s. Although more than 500 P450 genes have been found in prokaryotes and eukaryotes (including mammalian and plants) [16] only four 3D structures of bacterial, soluble P450 enzymes are known and are used in our 3D-structure-function analysis: P450 101 (P450_{cam}) [18], P450 102 (P450_{BM3}) [19], P450 107A (P450_{eryF}) [20], and P450 108 (P450_{terp}) [21]. The crystal structure of P450_{cam} from bacterium *Pseudomonas putida* has been studied in great detail by a variety of methods, including crystallography with different substrates and in a substrate free form [22]. P450_{cam} serves as a model system containing all necessary characteristics of P450 enzyme systems [23]. However, an access channel to the active site could not be found by X-ray analysis. In contrast to this P450_{BM3} was crystallized with a large cleft opened to allow access of substrates (long chained fatty acids) to the heme active site [24].

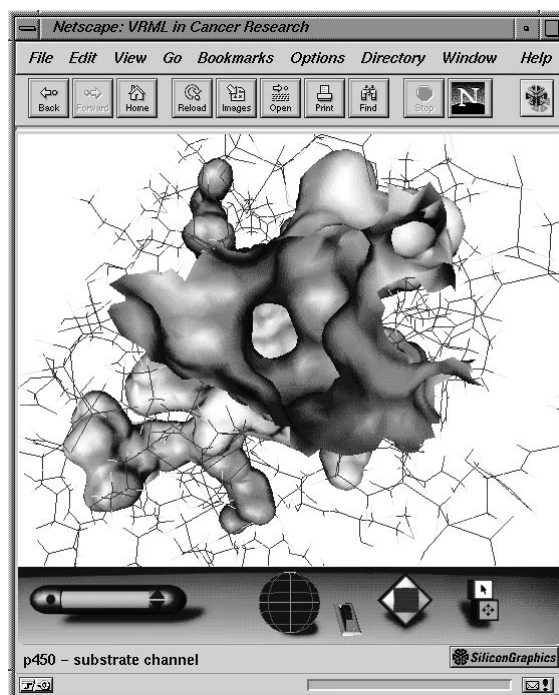


Fig. 5: The aperture of the active-site-access-model of P450_{cam} is shown as a VRML scene.

Using molecular modeling techniques we have found evidence for two possible channels leading to the active site of P450_{cam}. We introduce here a substrate channel for the entrance of ligands and a water channel, allowing water molecules to escape from the active site. Both are combined in an active-site-access-model of P450_{cam}. This 3D model is compared with recently estimated structures of P450_{eryF}, P450_{terp}, and P450_{BM3} using molecular surfaces and local biophysical properties. Fig. 4 shows the substrate channels of the four structures (side view). The 3D models were extracted from the molecular surface of the complete P450's by an automated procedure. The molecular surfaces of the access channels are used as screens for the representation of local molecular properties, e.g. electrostatic potential or lipophilicity with respect to the structural environment. The calculated local hydrophilic/lipophilic properties of the access channels are quite different. While the substrate channel is a lipophilic path especially in the middle region, the local properties of the water channel are much more hydrophilic.

A sequence comparison of residues forming the access channels of P450_{cam} and of P450_{BM3} was performed using the structure based alignment from Haseman et al.

[23]. Most of the P450_{cam} channel residues were matched with channel residues of P450_{BM3}. These findings gave hints for the common substrate access channel of the P450 enzymes and confirms the supposition of Haseman et al. that P450_{cam} represents the basic P450 structure [23]. In Fig. 5 the VRML scene of the active-site-access-model of P450_{cam} represented as part of the solvent accessible surface together with the appropriate amino acids is shown.

The interaction of P450's with possible substrates is studied with carcinogenic and non carcinogenic nitrosamines using the active-site-access-model of P450_{cam}. As a result of the analysis of local molecular properties common characteristics of carcinogenic nitrosamines activated by P450 enzymes were identified. The dynamics of substrate access was calculated with high temperature annealing simulations. The estimated reaction path along the substrate channel as well as the local molecular properties should be presented as a moving VRML scene in order to visualize these complex interactions at the beginning of tumor induction.

4.2 Binding properties of wild type and mutated p53 tumor suppressor protein.

The p53 tumor suppressor protein controls the cell cycle checkpoint responsible for maintaining the integrity of the genome. When DNA is damaged, the p53 level increases and the cell cycle is stopped to allow DNA repair followed by normal cell growth or induction of apoptosis [25]. The inactivation of the p53 tumor suppressor function by single missense point mutations is found in almost half of human cancers. Until Jan. '97, more than 6100 tumor specific mutations have been stored in the p53 mutation database which has been developed by M. Hollstein and coworkers [26]. It is now maintained at the International Agency for Research on Cancer (IARC) [27] and is accessible through the European Bioinformatics Institute (EBI, <ftp://ftp.ebi.ac.uk/pub/databases/p53/>).

It has been found that the majority of these mutations occur in the central p53 core region containing the DNA binding domain and the zinc binding domain. The 3D structure of this core domain of the *wild type* p53 protein complexed with DNA has been crystallized by N. Pavletich and coworkers [28] and can be obtained from the Brookhaven Protein Data Bank (<http://www.pdb.bnl.gov>). This allows investigation of the influence of mutations on DNA binding and protein functions.

We have analyzed the p53-DNA complex as well as some p53 mutants using the molecular modeling package SYBYL/MOLCAD [7]. The amino acid substitutions at the mutation hotspots were performed with the molecular dynamics simulation program CHARMM [29]. The investigations of molecular properties are based on molecular surfaces. These *solvent accessible surfaces* show the 3D size and topography of the molecules. Additionally the surfaces are used as maps for a color coded representation of statistics i.e. mutation frequency and local molecular properties such as hydrophilicity/lipophilicity or electrostatic potential.

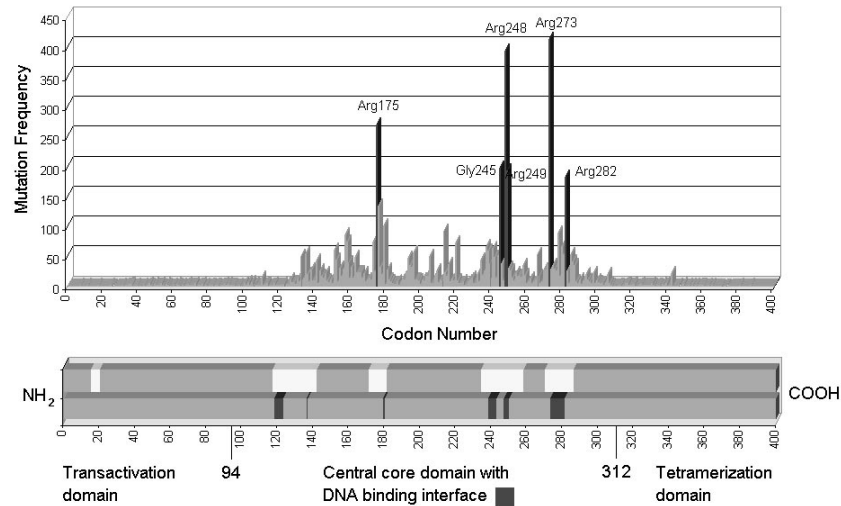


Fig. 6: Frequency of mutations found in human tumors obtained from the p53 mutation database (above). The mutation hotspots are labeled. On the graph (below) highly conserved amino acids (white) and amino acids located within the DNA binding region (black) are shown.

The frequency of mutations stored in the p53 mutation database is shown in Fig. 6 together with a graph indicating the corresponding conserved, the highly conserved amino acids [26] as well as the DNA binding interface of p53. We have projected the mutation frequency onto the molecular surface of the p53 protein (see Fig. 7). The black patches representing a high mutation rate are primarily situated at the binding interface of p53 and DNA. The majority of mutations found in human tumors are located in this region indicating its sensitivity for the function of p53 as a tumor suppressor.

To study the influence of mutations on the *wild type* structure of p53 we have generated altered p53 structures according to the main amino acid changes occurring at the mutation hotspots. These comparative investigations of point mutations at hotspots found in the p53 database are related to changes in local biochemical properties leading to a loss of specific binding to DNA. As an example the most mutated amino acid, Arg273 (p53 mutation database version Jan. '97), binds to the phosphate backbone of DNA. The corresponding 'contact mutant', Arg273His, loses direct binding contact to the DNA phosphate group. The Arg249Ser mutant - an example for a 'structural mutant' - prohibits the formation of binding of Arg248 into the minor groove, due to a disruption of an intramolecular bond to Glu171 and local structural changes.

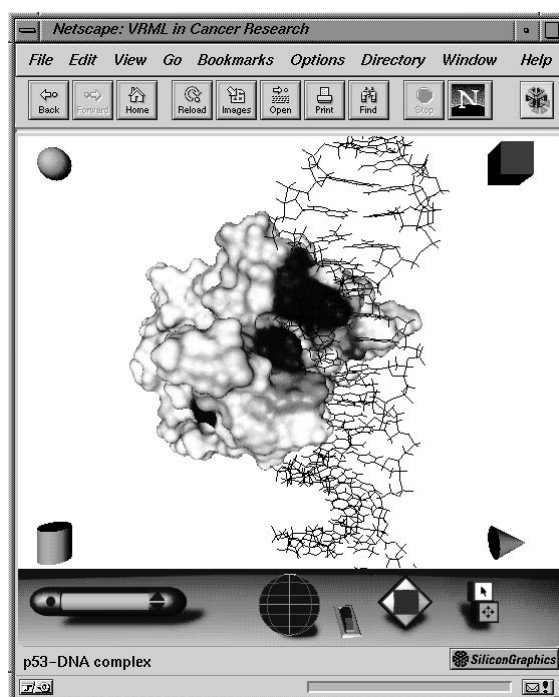


Fig. 7: VRML scene of the p53-DNA complex. The molecular surface of p53 is color coded according to the mutation frequency (white=low and black=high frequency) obtained from the p53 mutation database.

For communication with other scientists we use VRML to transfer the calculated and visualized results of our computer assisted investigations. To study the influence of a specific mutation one can select codons, e.g. a mutation hotspot in a clickable map of the mutation spectrum of the p53 gene (see Fig. 6) within the HTML/VRML pages. Using a link to the 3D structure of the *wild type* p53-DNA complex, the position of this corresponding amino acid as well as their local molecular properties visualized on the molecular surfaces are shown. Selecting a specific mutation occurring at this codon, the user can study the changes in geometry and also in local binding properties within the calculated 3D structure of the corresponding mutated p53-DNA complex. In this way a scientist can easily compare the effects of single point mutations on the structure and molecular properties of the intra- and intermolecular binding network of the p53-DNA complex.

5 Conclusion

It is shown with these examples from molecular modeling in cancer research, that the concept of molecular surfaces as screens for the representation of different properties corresponding to the local 3D structure is very useful for the investigation of complex biomolecular systems, i.e. specific intermolecular interactions. Molecular models of the access channels of the active site of P450 enzymes together with local molecular properties are used to study the interaction with carcinogenic and non carcinogenic nitrosamine substrates. The comparative analysis of *wild type* and mutated p53 complexed with DNA highlights the influence of single amino acid substitutions caused by point mutations on the p53 structure and its function as a tumor suppressor.

It is also demonstrated, that VRML is an excellent tool for the transfer of complex 3D information of molecular scenarios on the internet. Related information, e.g. other HTML documents or a new VRML scene can be selected interactively by the use of linked 3D objects (e.g. space buttons) implemented in the VRML scenarios. These 3D molecular scenarios can be explored with standard internet browsers. Thus VRML provides an effective information tool for the world wide communication among scientists using complex biomolecular models.

Acknowledgement

Many helpful discussions with Dr. M. Hollstein, Dr. B. Spiegelhalder and Prof. H. Bartsch from the Division: Toxicology & Cancer Risk Factors at the German Cancer Research Center are gratefully acknowledged. This work was supported by the "Deutsche Forschungsgemeinschaft" (project No. Br306/24-1).

References

1. Brickmann, J.; Vollhardt, H. *Trends in Biotechnology* **1996**, *14*, 167.
2. Vollhardt, H.; Henn, C.; Moeckel, G.; Teschner, M.; Brickmann, J. *J. Mol. Graphics* **1995**, *13*, 368.
3. Richards, F.M. *Ann. Rev. Biophys. Bioeng.* **1977**, *6*, 151.
4. Connolly, M.L. *Science* **1983**, *221*, 709.
5. Exner, T.; Brickmann, J. *J. Mol. Model.* **1997**, *3*, 321.
6. Pixner, P.; Heiden, W.; Merx, H.; Moeckel, G.; Moeller, A.; Brickmann, J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1309.
7. Brickmann, J.; Goetze, T.; Heiden, W.; Moeckel, G.; Reiling, S.; Vollhardt, H.; Zachmann, C.-D. *in Data Visualisation in Molecular*

Science: Tools for Insight and Innovation, Addison-Wesley Publishing Company Inc. Reading, Mass. **1995**.

8. Honig, B.; Nicholls, A. *Science* **1995**, 268, 1144.
9. Brucoleri, R.; Novotny, J.; Davis, M. *J. Comp. Chem.* **1997**, 18, 268.
10. Viswanadhan, V.N.; Ghose, A.K.; Revankar, G.R.; Robins, R.K. *J. Chem. Inf. Comput. Sci.* **1989**, 29, 163.
11. Keil, M. *Computergestützte Untersuchung der Bindungsregion des p53-DNA-Komplexes*, Darmstadt University of Technology, Diploma Thesis, Darmstadt, **1996**.
12. Heiden, W.; Moeckel, G.; Brickmann, J. *J. Comput. -Aided Mol. Design* **1993**, 7, 503.
13. Casher, O.; Rzepa, H.S. *J. Mol. Graphics* **1995**, 13, 268.
14. Wernecke, J. *The Inventor Mentor: Programming Object-Oriented 3D Graphics with Open Inventor, Release 2*, Addison-Wesley, Reading, Massachusetts, **1994**.
15. Ohshima, H.; Bartsch, H. *Mutation Research* **1994**, 305, 253.
16. Nelson, D.R.; Koymans, L.; Kamataki, T.; Stegeman, J.; Feyereisen, R.; Waxman, D.J.; Waterman, M.R.; Gotoh, O.; Coon, M.J.; Estabrook, R.; Gunsalus, I.C.; Nebert, D.W. *Pharmacogenetics* **1996**, 6, 1.
17. Bartsch, H.; Hietanen, E. *Env. H. Presp.* **1996**, 104(Suppl 3), 569.
18. Poulos, T.L.; Finzel, B.C.; Howard, A.J. *J. Mol. Bio.* **1987**, 195, 687.
19. Ravichandran, K.; Boddupalli, S.; Hasemann, C.; Peterson, J.; Deisenhofer, J. *Science* **1993**, 261, 731.
20. Cupp-Vickery, J.R.; Poulos, T.L. *Nat. Struct. Biol.* **1995**, 2, 144.
21. Hasemann, C.; Ravichandran, K.; Peterson, J.; Deisenhofer, J. *J. Mol. Bio.* **1994**, 236, 1169.
22. Poulos, T.L. *Current Opinion in Structural Biology* **1995**, 5, 767.
23. Hasemann, C.; Kurumbail, R.; Boddupalli, S.; Peterson, J.; Deisenhofer, J. *Structure* **1995**, 2, 41.
24. Poulos, T.L.; Huiying, L. *Structure* **1994**, 2, 461.
25. Arrowsmith, C.; Morin, P. *Oncogene* **1996**, 12, 1379.
26. Hollstein, M.; Sidransky, D.; Vogelstein, B.; Harris, C. *Science* **1991**, 253, 49.
27. Hollstein, M.; Shomer, B.; Greenblatt, M.; Soussi, T.; Hovig, E.; Montesano, R.; Harris, C. *Nucl. Acid Res.* **1996**, 24, 141.
28. Cho, Y.; Gorina, S.; Jeffrey, P.D.; Pavletich, N.P. *Science* **1994**, 265, 346.
29. Brooks, B.R.; Brucoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem.* **1983**, 4, 187.