[Frontiers in Bioscience 14, 2559-2583, January 1, 2009]

Molecular visualization in the rational drug design process

Matthias Keil[1], Richard J. Marhofer[2], Andreas Rohwer[2], Paul M. Selzer[2], Jurgen Brickmann[3] Oliver Korb[4], Thomas E. Exner[4]

[1]Tripos International, 1699 South Hanley Road, St. Louis, MO 63144-2319, USA, current address: Chemical Computing Group Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Quebec, H3A 2R7, Canada, [2]Intervet Innovation GmbH, BioChemInfomatics, Zur Propstei, 55270 Schwabenheim, Germany, [3]Department of Chemistry, Darmstadt University of Technology and MOLCAD GmbH, Petersenstr. 20, 64287 Darmstadt, Germany, [4]Fachbereich Chemie, Universitat Konstanz, 78457 Konstanz, Germany

TABLE OF CONTENTS

1. ABSTRACT

The visualization of molecular scenarios on an atomic level can help to interpret experimental and theoretical findings. This is demonstrated in this review article with the specific field of drug design. State-of-the-art visualization techniques are described and applied to the different stages of the rational design process. Numerous examples from the literature, in which visualization was used as a major tool in the data analysis and interpretation, are provided to show that images are not only useful for drawing the attention of the reader to a specific paper in a scientific journal.

2. OVERVIEW OF THE DRUG DESIGN PROCESS

Molecular visualization is nowadays a vital tool in the life of a chemist or biologist. This is also especially true for the rational drug design process. Well-designed presentations of molecular data help to plan and optimize the design process, judge numerical results and output of modeling programs, and to discuss experimental and theoretical findings. It is important to note that, most of the time, this visualization is not the end product like it is seen in publications. Many of the produced graphics and even animations will only be used temporarily to present different ways of looking at a specific problem, i.e. as

interactive displays to generate and communicate new ideas. To keep the focus on the application, the used techniques will be outlined only briefly in textboxes spread out across this paper. These can be red independently from the rest of the manuscript. We refer to the literature for more thoroughly reviews on the more technical aspects of molecular visualization (1,2,3,4,5,6,7,8,9,10). Also, the large number of visualization tools will not be reviewed or compared. In this publication, the program MOLCAD (11,1), distributed worldwide as part of the molecular modeling package SYBYL (12), was used to create the pictures unless otherwise indicated.

The serendipity approach of random screening of compound collections against biological systems, which dominated the drug design approach for decades, has been replaced by a more rational approach (see Figure 1) some ten years ago. The decoding of the DNA's three-dimensional structure by James Watson and Francis Crick (13) and the large-scale sequencing of many prokaryote and eukaryote genomes including the human genome led to the approach of receptor-based screening. In this approach, the biological system is replaced by a molecular target, which could be any biomolecular macromolecule, e.g. a protein, a DNA- or RNA-molecule. Hence, the initial step in the drug discovery chain is the identification of such a macromolecule, which is involved in a pathway relevant to the disease of interest on the one hand and, on the other hand, which is druggable by small molecule ligands. For the identification of potential targets, classical molecular biology methods are used as well as high-throughput methods like functional genomics and proteomics and computational approaches like bioinformatics. The druggabilty can be assessed by assigning the potential target molecule to its gene family using bioinformatics methods. If one member is known to be influenced by a small molecule ligand, it is assumed that other members of this family are also druggable (14). Once a potential target is identified it must be validated, i.e. it must be shown that affecting the target leads to the desired effect. This is usually done using, amongst others, knock-out and knock-in mutants in animal models (15) or RNA interference experiments. If a small molecule ligand is known for one of the protein family members, the potential target can also be validated chemically which of course is a surplus for the druggability assessment.

The validated target is then passed on to the next step of the drug discovery chain, the lead identification process. In this step small chemical entities are identified which interact with the target in the desired way. An average compound collection available at a pharmaceutical company easily comprises some 100,000 to millions of compounds, consisting of proprietary compounds and commercially available compound collections. Obviously, such a number of compounds can only be tested using high-throughput screening approaches (HTS). The screening of an average compound collection results in a high number of initial hits from which a suitable subset of compounds is selected for further dose-response confirmation by a team of biologists, medicinal chemists, and chemoinformaticians. Finally, the confirmed hits are ranked considering a wide range of information, like intellectual property considerations and predicted physico-chemical and predicted ADME-Tox (absorption, distribution, metabolism, excretion, and toxicity) properties of the individual molecules. From this ranked set, a small number of highly ranked compounds are chosen as lead structures.

In addition to the physical screening as described above, virtual screening has gained an increasing role recently. While initially it had been considered as a replacement of the physical screening, it gained its importance as a complementary method (16). Although, the term high-throughput implies very high speed, the screening of an average compound collection needs considerable time even if 384 well-plates or 1536 well-plates are used. Moreover, a number of constraints raising from the high-throughput conditions, e.g. stable and reproducible assays, must be considered. A reduced number of compounds to screen would be desirable because it would ease these limitations tremendously. For a theoretical prediction of possible ligands, an accurate estimate of the binding free energy would be desirable. But highly sophisticated methods able to provide this accuracy, like free energy perturbation or thermodynamic integration (17), need a large amount of computer time (hours to days on a supercomputer) even for a single compound resulting in considerable costs in time and money even exceeding the ones of physical screening. Even then, a funded knowledge about the placement of the ligand in the active site is needed before starting these calculations. Therefore, to be commercially relevant for the screening of large databases in a pharmaceutical company, the methods to be used must compromise between speed (and thus costs) and accuracy. At this point virtual screening comes into play. Virtual screening is the generic term for any in silico method used in a high-throughput manner for the identification of potential ligands to a receptor under consideration. According to the used methods, virtual screening can be classified as either receptor-based virtual screening, if e.g. molecular docking has been used, or as ligand-based screening, if e.g. pharmacophore modeling has been used. Since in either case virtual screening causes only negligible costs for each additional compound, it is in principle possible to screen an unlimited number of compounds, which in addition, need not necessarily exist already. Only the most promising hits from the virtual screening are subsequently passed on to the physical screening, which then can be carried out in medium-throughput with greatly reduced limitations. For a detailed description of virtual screening methods and their applications in actual projects see (18,14) and references therein.

Once lead structures have been identified, they are passed on to the final step of the drug discovery chain, the lead optimization. In lead optimization the lead structures are chemically modified and their properties are optimized towards a minimum set of required characteristics, like potency, selectivity, suitable ADME-Tox parameters, etc., using medicinal chemistry in a number of optimization cycles. Because a relatively high

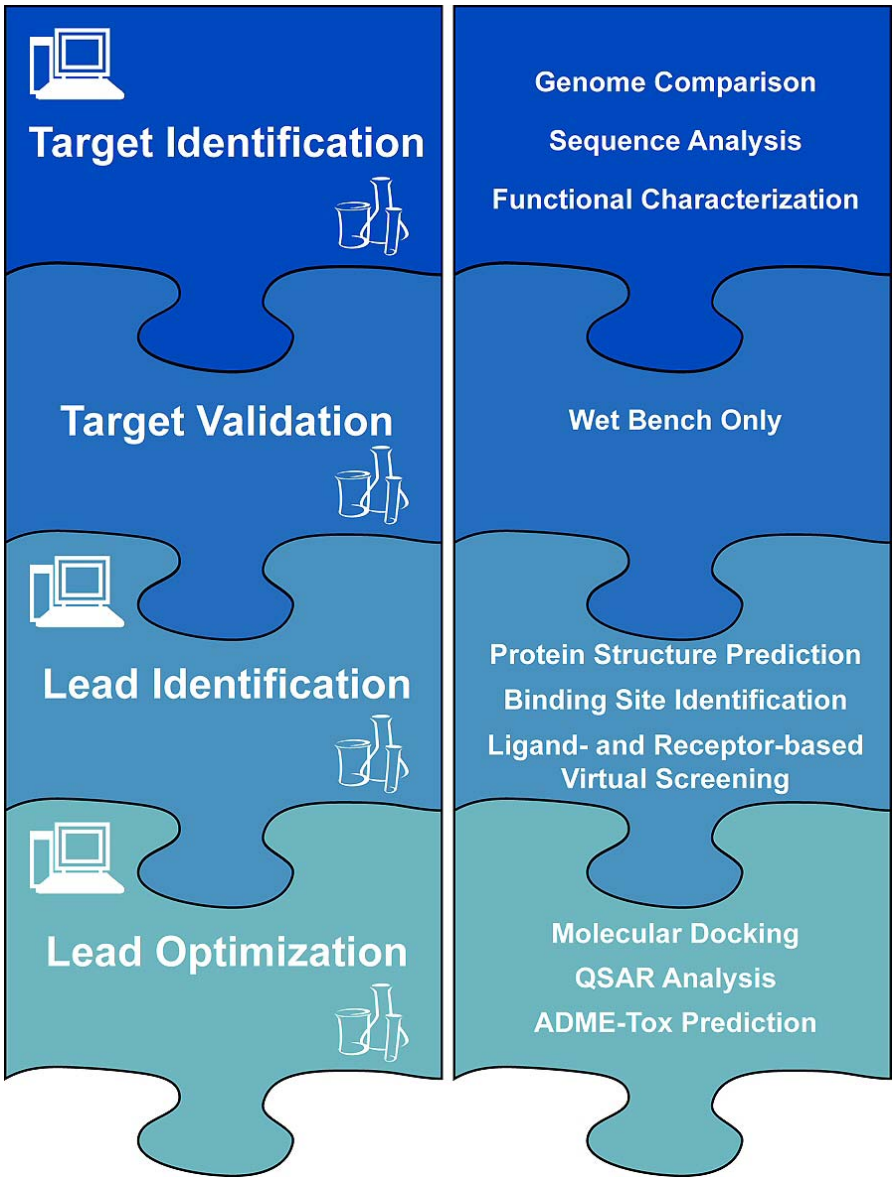

Figure 1. Schematic representation of the drug discovery workflow. In the left panel the four major steps are depicted. The icons in the upper left and lower right corner symbolize support by computational methods (computer) and wet bench work (flasks and tubes), respectively. The right panel lists examples of computational methods, which play a major role in the individual steps of drug design. Some of these examples are described in greater detail throughout the coming sections.

number of properties must be optimized in parallel, lead optimization is a very complex and time consuming task. It is therefore usually supported by applying a number of chemoinformatics methods, like pharmacophore modeling, molecular docking, QSAR analysis etc. The result of the lead optimization is a drug candidate, showing the desired set of properties, which is finally transferred to a drug development program.

2.1. Textbox 1. Standard representations

Molecular structures are commonly drawn in three dimensions or as two-dimensional projections in the form of lines representing the bonds between atomic centers. These 2D and 3D drawings are referred to as Lewis structures (Figure 2), named after Gilbert N. Lewis (109), and as line models, Dreiding models (110), or wire-frame representations (Figure 3), respectively. Atoms are only represented as junctions of these lines or in Lewis structures with their atomic label. Colors are used to code atom types, different molecules in a complex or sometimes to highlight physicochemical properties.

The advances in computer graphics allow the interactive generation of more user-friendly representations like balls-and-sticks, capped sticks, and Corey-Pauling-Koltun (CPK) (111) models giving some indications of the depth and three-dimensional conformation of the molecule

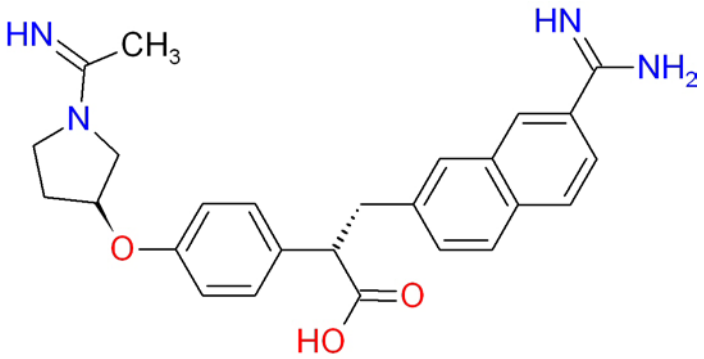

Figure 2. Lewis structure of the inhibitor DX-9065a of coagulation factor Xa (73).

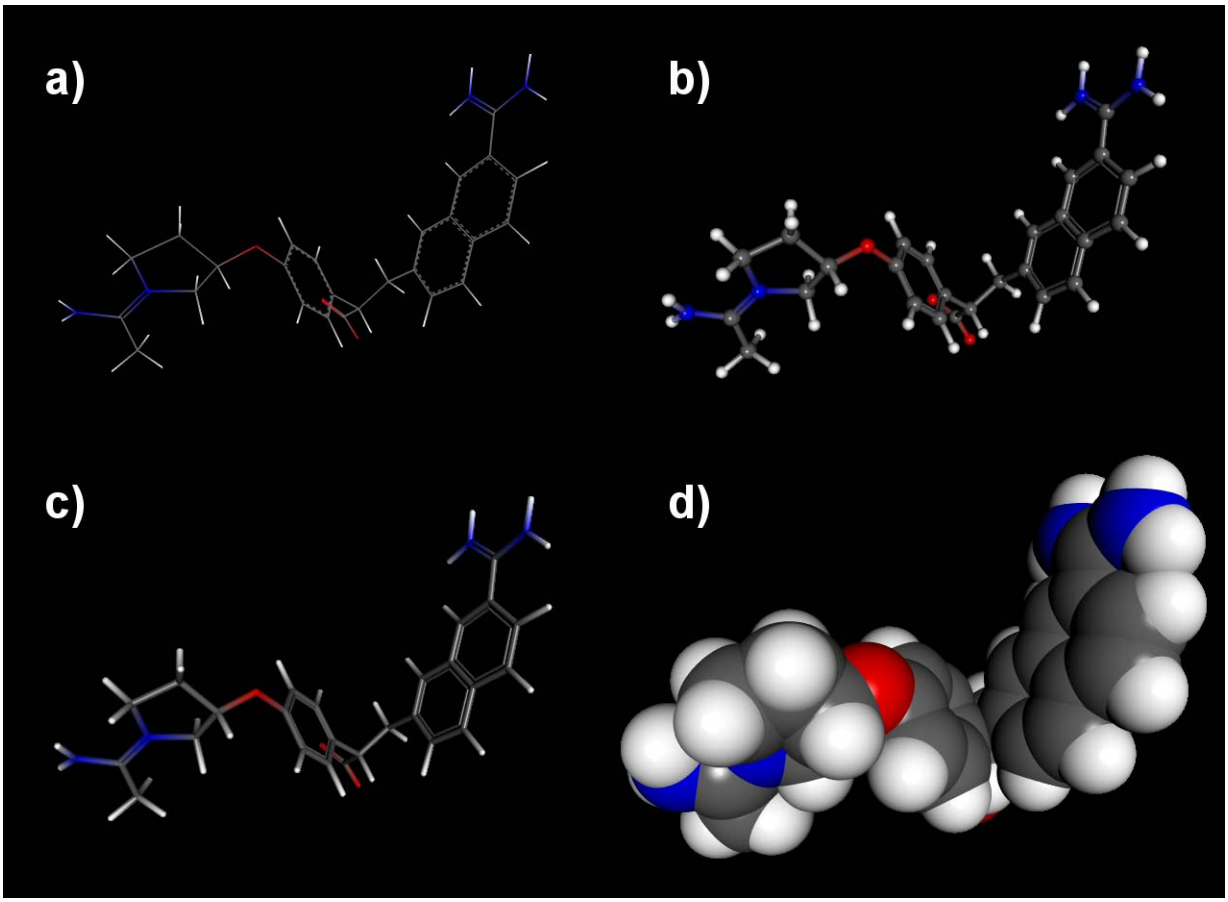

Figure 3. 3D models of the factor Xa inhibitor of Figure 2: (a) wire frame; (b) balls-and-sticks, (c) capped sticks; and (d) CPK model.

due to hidden-line removal (Figure 3). The latter also depicts the volume and the outer shape of the molecule in the hard-sphere approximation.

3. PROTEIN STRUCTURE PREDICTION

The first step in the rational, computer-aided drug design should be the careful characterization of the protein target. One key element for doing so is the knowledge of the arrangement of the atoms composing the macromolecule in three-dimensional space, the protein structure. Proteins are linear polymer strings of amino acids. The structure of proteins can be represented in several different levels of detail. The so-called primary structure of a protein describes the amino acid sequence of the peptide chains. Mediated by inter-residue hydrogen bonds, the linear chain of amino acids folds into discrete secondary structure motifs, like helices or strands, known as alpha helices and beta sheet strands. The three-dimensional arrangement of these often rigid and conserved

secondary structure elements and more flexible protein regions (coil or loop regions) form the tertiary protein structure. The formation of this overall fold is driven and stabilized by several different types of interaction such as the burial of hydrophobic residues inside the core, formation of hydrogen bonds, salt bridges, and disulfide bonds. Knowledge of the tertiary structure is important for studies of protein function.

Experimental determination of the three-dimensional protein structure (e.g. via X-ray diffraction or NMR experiments) is still a difficult process and much more demanding than determining the protein sequence (primary structure). This difference is evident in the ratio of known protein sequences versus three-dimensional structures (~5 million sequences in the UniProt database (19), ~47,000 structures in the Protein Data Bank (20)).

As outlined above the structure is important for functional studies. When it is not feasible to determine the structure experimentally, one can resort to a variety of computational methods, which try to predict and construct an atomic-resolution model on the basis of the amino acid sequence. Homology modeling (15,21,22), also known as comparative modeling, is such a technique. It is based on the observation that related proteins within a protein family have similar three dimensional folds (tertiary structure) even when their amino acid sequences show weak similarity (23). The homology modeling process can be divided into several distinct steps. First one or more known protein structures, which can be used as templates for the protein structure prediction, are identified. The sequences of these homologs are then aligned to the protein sequence to be modeled (known as query or target sequence) (see Figure 4). This alignment maps the residues in the query sequence to the residues in the template sequences. The next step then is to construct a core model using the structural features, which are conserved over the templates and the sequence alignment. This model of structural conserved regions (SCRs) is often composed of secondary structure elements like alpha helices and beta sheets. Next the structurally variable regions (SVRs or loops) connecting the secondary structure elements are modeled. A variety of algorithms including database searches, ab-initio modeling, or force field guided methods exist for this task. Depending on the details of the employed algorithm, the amino acid side-chains are added during the modeling of the structurally conserved regions and loops, or added in an additional step. The final steps of the model construction include optimization of the atomic structure with e.g. energy minimizations to relieve local stress and validation of the overall model quality by analyzing the protein geometry and comparing its properties, like dihedral angles etc., with distributions from experimental structures as it is done in Ramachandran plots.

During all steps of the homology modeling process visualization of the data is helpful and therefore important. A wide variety of different visualization techniques and their combinations can be used. In the alignment step the sequence information of the homologs and the target sequence are often shown as simple strings of letters, each letter representing a single amino acid (e.g. A = alanine, G = glycine, etc.). Additional information about e.g. the secondary structure can be added as color code or text style of the sequence string letters (see Figure 4). Homology programs also often link the sequence strings representing the primary structure to a three-dimensional view of the corresponding homolog or model structures, so that the researcher can combine the strength of the various visualizations types to improve the model building results. Additionally the three-dimensional representation of the homolog and model structures varies from step to step. At the structural alignment stage and SCR building step the display of just the protein backbone atoms or just a ribbon/tube rendering is favored (see Figure 5), while at the side-chain placement stage the positions of all side-chain atoms and their possible interactions like hydrogen bonds or atom clashes are important and shown. The visualization types used during the whole model building process have to fit the specific needs of each step.

3.1. Textbox 2. Simplified molecular representations and special features

The standard models of Textbox 1 based on an atomic representation are not well suited for biochemical applications because they become increasingly confusing if the number of atoms increases to more than a few hundreds. Here a reduction of complexity is necessary to generate better understanding. For proteins, DNA, and RNA this can be achieved by drawing only the backbone of the biopolymers as strands, ribbons, or tubes (Figure 6). Secondary structure elements can be highlighted in the so-called Richardson style (112) by arrows and cylinders. The simplified representations can be mixed with standard representations to generate detailed insights into particular parts of the scenario.

4. BINDING-SITE IDENTIFICATION AND CHARACTERIZATION

For the understanding of the function of a protein and to model ligands interfering with this function, the highly accurate characterization of the ligand- or protein-binding site of the specific pharmaceutical target is the next step towards successful structure-based drug design. However, sometime even the location of the site for direct or allosteric ligand binding - l*et al*one its specific properties - is not known. Additionally, finding ligand binding sites to modulate protein-protein interactions as well as additional binding sites, where appropriate targeting could result in new biological effects or new classes of inhibitors, is of increasing interest. Therefore, tools for the automatic, computer-based prediction of binding sites (25,26,27,28,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49) have become quite popular especially as front-ends to molecular docking (see next section). For these algorithms, the huge amount of data available from experimentally determined complex structures as stored e.g. in the Protein Data Bank (20) was used to analyze the properties of proteins in general and of binding sites in particular. All these investigations stress the large complementarity of the shape as well as the

```
                    10        20        30        40        50
query          IVGGQECKDGECPWQALLINEE-NEGFCGGTILSEFYILTAAHCLYQAKR
1kigh          IVGGRDCAEGECPWQALLVNEE-NEGFCGGTILNEFYVLTAAHCLH---Q
1fxya          IVGGYNCKDGEVPWQALLINEE-NEGFCGGTILSEFYILTAAHCLY---Q
1ppb           IVEGSDAEIGMSPWQVMLFRKSPQELLCGASLISDRWVLTAAHCLLYPPW
1bbr           IVEGQDAEVGLSPWQVMLFRKSPQELLCGASLISDRWVLTAAHCLLYPPW

                    60        70        80        90        100
query          --------FKVRVGDRNTEQ-EEGGEAVHEVEVVIKHNRFT-KETYDFDI
1kigh          -AK---R-FTVRVGDRNTEQ-EEGNEMAHEVEMTVKHSRFV-KETYDFDI
1fxya          -AK---R-FKVRVGDRNTEQ-EEGGEAVHEVEVVIKHNRFT-KETYDFDI
1ppb           DKNFTENDLLVRIGKHSRTRYERNIEKISMLEKIYIHPRYNWRENLDRDI
1bbr           DKNFTVDDLLVRIGKHSRTRYERKVEKISMLDKIYIHPRYNWKENLDRDI

                    110       120       130       140       150
query          AVLRLKTPITFRMNVAPACLPERDWAESTLMTQ--KTGIVSGFGRTHEKG
1kigh          AVLRLKTPIRFRRNVAPACLPEKDWAEATLM-T-QKTGIVSGFGRTHE--
1fxya          AVLRLKTPITFRMNVAPASLPTAP-P-A----TGT-KCLISGWGNTAS--
1ppb           ALMKLKKPVAFSDYIHPVCLPDRETA-ASLLQAGY-KGRVTGWGNLKETW
1bbr           ALLKLKRPIELSDYIHPVCLPDKQTA-AKLLHAGF-KGRVTGWGNRRETW

                    160       170       180       190       200
query          EQS------TRLKMLEVPYVDRNSCKLSSSF-IITQNMFCAGYDTKQE--
1kigh          -K--GR-LSSTLKMLEVPYVDRSTCKLSSSF-TITPNMFCAGYD-TQPE-
1fxya          -S--GADYPDELQCLDAPVLSQAKCEASYP-GKITSNMFCVGFL-EGGK-
1ppb           TANVGKGQPSVLQVVNLPIVERPVCKDSTRI-RITDNMFCAGYKPDEGKR
1bbr           TTSVAEVQPSVLQVVNLPLVERPVCKASTRI-RITDNMFCAGYKPGEGKR

                    210       220       230       240       250
query          -DACQGDSGGPHVTR--FKDTYFVTGIVSWGEGCARKGKYGIYTKVTAFL
1kigh          -DACQGDSGGPHVTRFK--DTYFVTGIVSWGEGCARKGKFGVYTKVSNFL
1fxya          -DSCQGDSGGPVVC--N--G--QLQGVVSWGDGCAQKNKPGVYTKVYNYV
1ppb           GDACEGDSGGPFVMKSPFNNRWYQMGIVSWGEGCDRDGKYGFYTHVFRLK
1bbr           GDACEGDSGGPFVMKSPYNNRWYQMGIVSWGEGCDRDGKYGFYTHVFRLK

                    260
query          KWIDRSMKTRGLPKAK
1kigh          KWIDKIMKARAGAAGS
1fxya          KWIKNTIAANS-----
1ppb           KWIQKVID-QFGE---
1bbr           KWIQKVID-RLGS---
```

Key

Joy Annotation:

alpha helix	red	x
beta strand	blue	x
3 - 10 helix	maroon	x

Figure 4. Multiple sequence alignment of a query sequence and related homologs as generated by the program FUGUE (24). The sequence characters of the homologs are color coded by the secondary structure classification.

complementarity of the molecular properties between the two molecular partners building a complex. They also show that binding sites share some common features but depend on the molecule to which the protein is bound. In this way, protein–protein interfaces are more hydrophobic than other parts of the protein surface, even if they are less

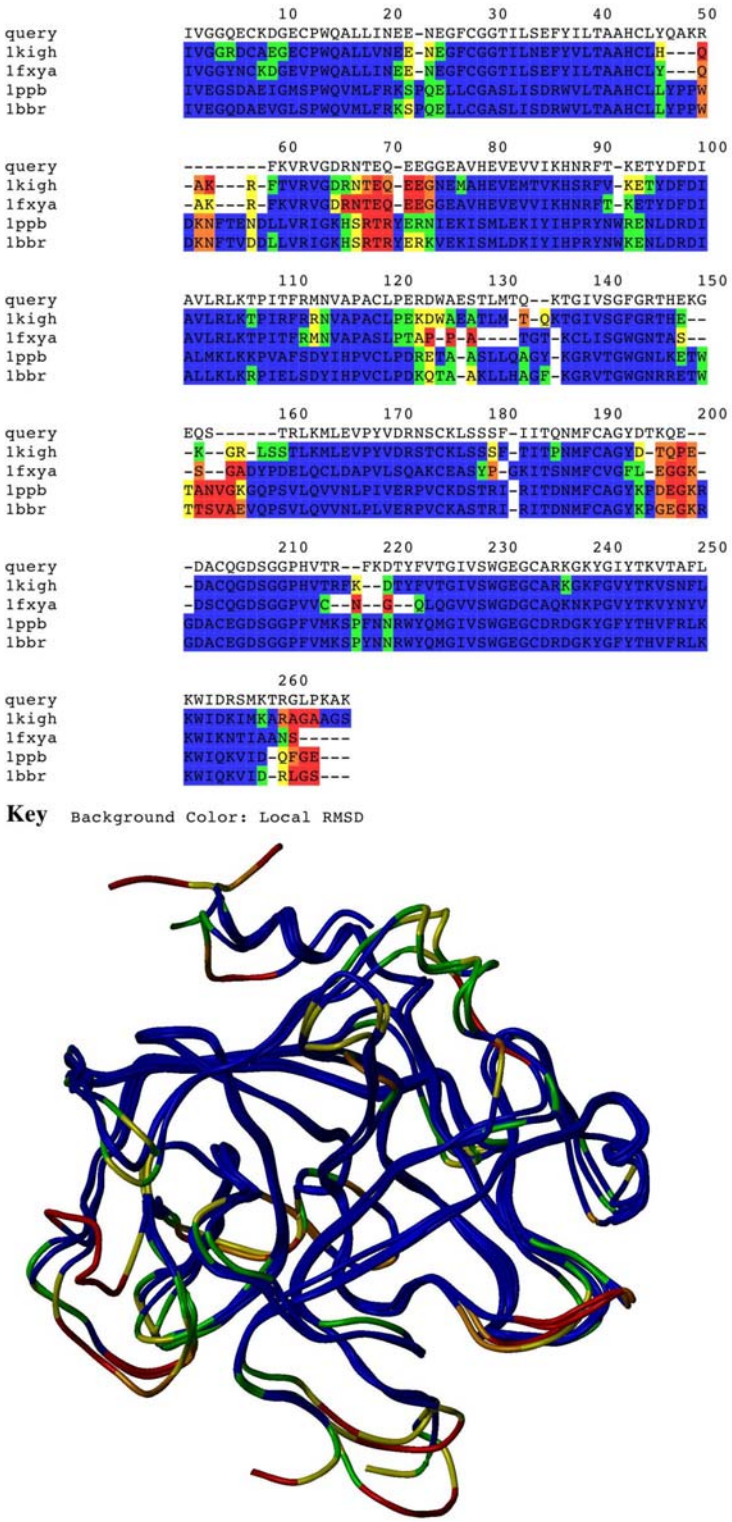

Figure 5. Sequence and structural alignment of the homologs color coded by the local RMSD of the aligned residues. The alignments were generated with the program BATON which is part of ORCHESTRAR inside the SYBYL molecular modeling suite (12). Blue indicates that the C-alpha atom of the residues in one column of the sequence alignment deviates less then 1 Å from the mean position (green: <2 Å, yellow: <3Å, red: >3 Å). This visualization type enables the user to judge the quality of the sequence and structural alignment. Red regions indicate either structural variable regions like loops which are acceptable or a misalignment, which should be investigated in order to achieve the best possible protein modeling result.

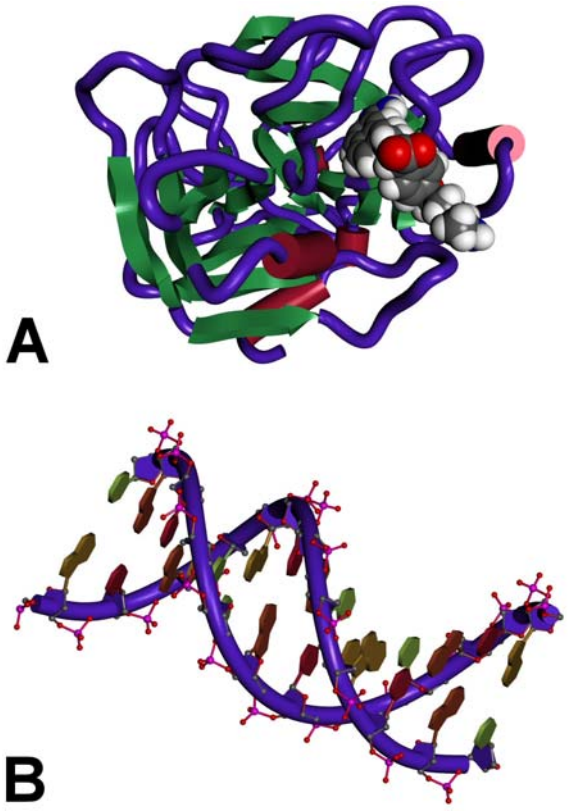

Figure 6. Simplified molecular representation of the factor Xa – inhibitor complex and a B-DNA dodecamer. The backbones are represented as tubes. α-helices are highlighted as red cylinders and β-sheets as green arrows. The inhibitor is shown as a CPK model and the DNA bases as pentagons and hexagons.

hydrophobic than the core region (31,33,37,39). Thus, on one hand, hydrophobic interactions are important not only in folding - resulting in the tertiary structure - but also in complex formation resulting in the quaternary structure. On the other hand, the number of charged amino acids is significantly increased in binding sites compared to the core region. Elcock *et al.* (35) related this to the fact that, while charged groups destabilize the hydrophobic core of a protein, these groups can stabilize complexes by building salt bridges across the binding interface. E.g. to bind to a DNA molecule, the binding site of a protein has to be positive polarized to favor interactions with the negative charges of the phosphate backbone (42). Thus, positive charged amino acids, like arginine, occur more often in DNA binding sites. In addition, the number of hydrogen bond donors is also increased. Binding sites for small molecules are mainly characterized by their concave shape. They are located in deep cavities, like bags and clefts of the molecular surface (49,32,34,46,48,44,47).

These special features open the possibility to find binding sites by comparing the molecular properties at different parts of the protein surfaces and locate those that differ from the average. For small molecule-binding sites often the identification of large clefts or bags in the protein surface are sufficient (49,32,34,46,48,44,47). The inclusion of additional features of the surface, like lipophilicity, electrostatic potential, hydrogen-bond acceptor / donor profiles, and amino-acid composition may be needed to increase the accuracy or even make a prediction possible in the first place (especially for protein-protein interactions). The visualization of these features or a combined score of these points the researcher directly to the important parts of the surface. E.g. Keil *et al.* (30) analyzed the electrostatic potential, local lipophilicity, hydrogen bond donor and hydrogen bond acceptor density, surface topography, and cavity depth of the molecular surface of the target protein and combined them by a neuronal network to identify protein-ligand, protein-protein, and protein-DNA binding sites. By visual inspection, binding sites can then be easily identified as large regions with a high probability of belonging to a specific binding type (Figure 7). Finally, even more sophisticated force-field-based methods can be applied to calculate explicit interactions of the protein with probes (50,29,34,51,52,53,54,55), representing e.g. hydrophobic or water molecules, molecular fragments, or small organic molecules to identify interaction hot spots belonging to druggable binding regions.

Another interesting question related to pharmacokinetics is how a ligand can access the active site. For proteins, where the binding site is located at the molecular surface, this is straight forward. But there exist also catalytic sites, which are buried deep inside the protein and can only be reached trought a large channel. To predict the path of the ligand, the displacement of water molecules occupying the channel, and possible gating mechanisms, these channels have to be detected. One famous example is the family of cytochrome P450 playing a central role in most phase I metabolisms by oxidizing a variety of substrates, of both endogenous and exogenous origin. Here, the active site is composed out of a heme group inside the protein. The specificity of the different subfamilies for substrates as well as inhibitors is probably highly related to the size and the physicochemical properties of the channel leading to the extremely active iron (see Figure 8).

If the active site is identified, it can be further characterized using various visualization techniques. As a first example, we demonstrate this with the p53 tumor suppressor protein-DNA complex (Figure 9). Additional examples will be shown in the following sections. The p53 protein controls the cell cycle checkpoint responsible for maintaining the integrity of the genome. When DNA is damaged, the p53 level increases stopping the cell cycle and allowing DNA repair followed by normal cell growth or inducting apoptosis (57). The inactivation of p53 by single point mutations is found in almost half of human cancers (58). By projecting the mutation hotspots onto the molecular surface, the influences of these point mutations on the structure as well as the physicochemical properties can be analyzed (59). E.g. the mutation R248W shown in Figure 9 is too bulky to fit into the minor grove of the DNA and, additionally, removes the positive charge interacting favorably with the DNA backbone

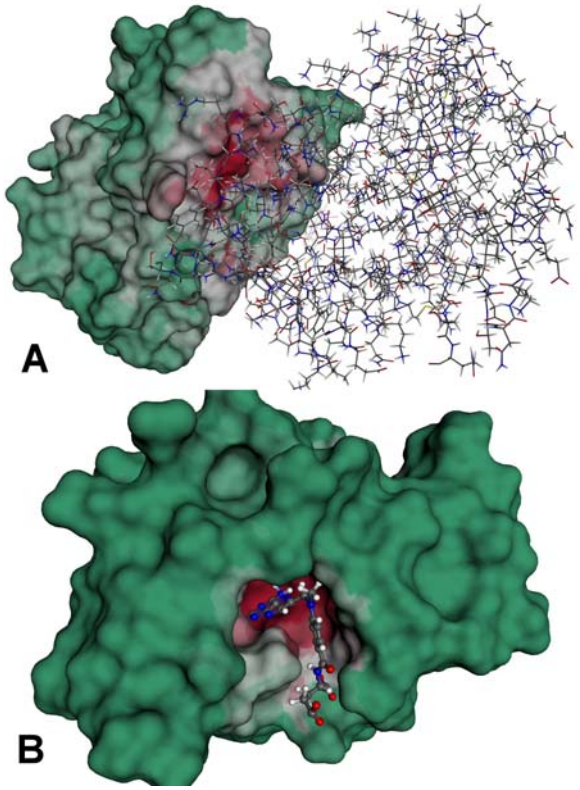

Figure 7. Binding site of the thymidilate kinase dimer (A) and the dihydrofolate reductase–Methotrexate (B) complex. The property describing the ability to form the specific complex type is shown color-coded on the molecular surface of the target protein (30) (red, high probability for complex formation; green, low probability). The second protein and the ligand are represented as a stick and balls-and-sticks models, respectively.

4.1. Textbox 3. Molecular surfaces and mapped properties

Atoms and molecules do not have any surface such as macroscopic objects have. They are composed out of nuclei and electrons and are, following quantum mechanics, best described by an electron (or even nuclear) distribution function or density occupying the total 3D space. Due to the higher probability to find an electron near the nuclei, isovalue surfaces (see Textbox 4), connecting all points in space with the same value of the electron density, can be used for the visualization of the shape of the molecule. This physics-based approach to generate a molecular surface is mainly suitable for small molecules, for which quantum mechanical calculations are feasible. In the life sciences, the definition of molecular surfaces introduced by Richards (113) describing a molecular envelope accessible by a solvent molecule is more commonly and very successfully used. A smooth surface based on a hard sphere model can be generated by rolling another hard sphere model particle (e.g. a water molecule with an effective sphere radius of r = 1.4. Å) over the CPK model of a molecule. This procedure was first applied by Richards (113) and Connolly (114,115) and forms as so-called “Connolly surface” a reference standard in many molecular modeling packages (Figure 10).

The surfaces are not only used for imaging the "bulkiness" and shape of molecules, they also serve as screens for the visualization of higher dimensional information. A popular method of displaying scalar values on molecular surfaces is achieved by the use of color codes (2,1). 3D properties like the electrostatic potential can be calculated at the position of the surface and atom-based properties like the molecular lipophilicity can be projected onto the surface allowing an immediate identification of important molecular regions (2,1). A useful application of multiple property mapping is the introduction of transparency for encoding an additional chemical property. Such filter methods enable the scientist even more to distinguish between important and irrelevant information. Another example of 2D-texture mapping is using a normal color ramp for one property and saturation of these colors for the second property.

5. PROTEIN-LIGAND DOCKING

Protein-Ligand Docking, as the term implies, means the placement of small chemical entities (ligands) into the active site of a protein (receptor) in a valid pose and the assessment of the binding affinity between the ligand and the receptor. Therefore, the ligand under consideration must fulfill two major criteria. First of all it must obey a topological fit citerion, i.e. its size and shape must be compatible with the size and shape of the protein’s binding site. Secondly, the ligand’s molecular features must complement the features found in the protein’s binding site in order to allow for the binding of the ligand. Following these considerations most docking programs work in a two-step approach.

In the first step, the ligand is placed into the binding site considering only the shape and size of the ligand. Because, even relatively small ligands are quite flexible, the shape and size of a given ligand can vary tremendously. It is therefore of paramount importance to explore the configurational and conformational degrees of freedom of the ligand under consideration sufficiently in order to find the optimum solution. This can be accomplished by either pre-generating all possible conformations and configurations like e.g. tautomers and storing them in a database for a subsequent docking using a rigid docking approach or it can be accomplished on the fly during the placement of the ligand. In practice, often a combination of both methods is used. While the conformational space is sampled during the placement of the ligand, the configurational space is sampled by pre-generating all possible and meaningful tautomers and stereoisomers (61). A number of different algorithms for the exploration of conformational space are implemented in the different docking programs using systematic and stochastic approaches as well as deterministic methods. For a detailed discussion of these approaches see (62). Deterministic approaches however, are mainly used in single ligand docking due to their computational expensiveness. In virtual high-throughput docking experiments the faster and computational cheaper systematic and stochastic search methods are applied

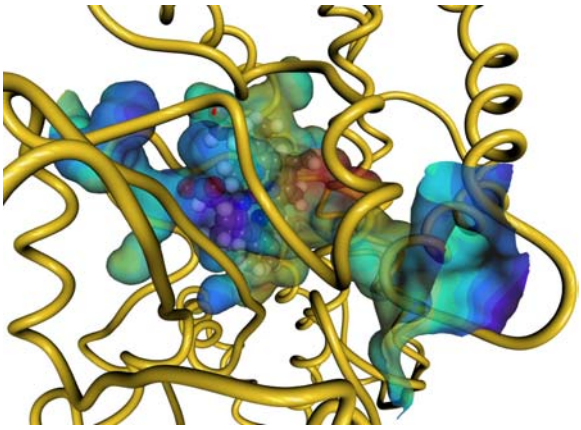

Figure 8. Channel leading to the active site of cytochrome $P450_{cam}$ (56). The molecular surface of the channel is color-coded according to the local lipophilicity. The large entrance on the right side is leading directly to the iron ion of the heme group. On the left, another opening can be seen, which is possibly an exit for the water molecules replaced by the substrate.

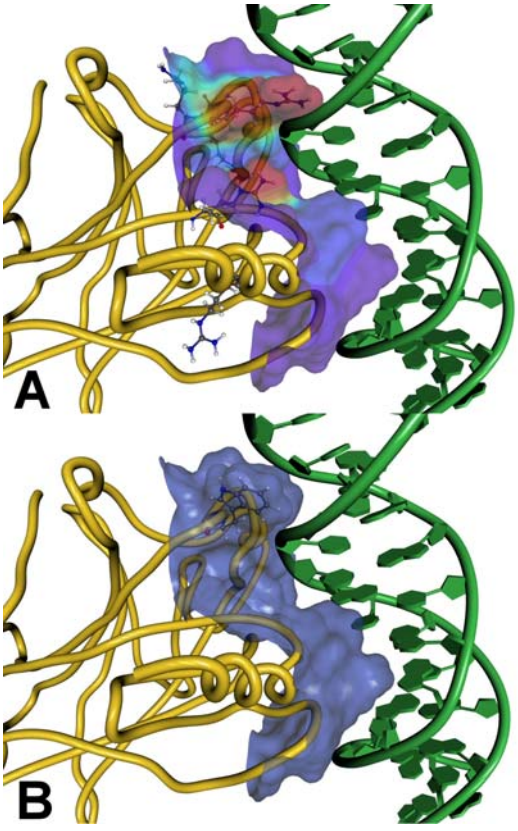

Figure 9. (A) the wild type p53 protein-DNA complex (pdb entry 1tup) is shown taken from the crystal structure of Cho *et al.* (60). The mutation hot spots are shown in balls-and-sticks representation (59). They are also color coded on the molecular surface of the interface (red=high and blue=low mutation rate). (B) the R248W mutation and its influence on the interface are exemplarily visualized.

Once the ligand is placed in the protein's binding site, its binding pose must be assessed and transformed into a binding affinity. For this purpose a number of scoring functions have been developed in recent years. Scoring functions allow to estimate the binding energy of the complex on the basis of individual interactions of the ligand and the protein residues. The available scoring functions can be classified according to Brooijmans and Kuntz (62) into four main categories: (i) first-principles methods, (ii) semi-empirical scoring functions, (iii) empirical methods, and (iv) knowledge-based methods. First-principles methods rely on functional terms for explicit interactions between atoms in the two molecules. Interaction parameters are taken directly from force fields without modulating them with empirically derived parameters (63,64). Solvent effects caused by the surrounding solvent molecules are taken into account implicitly by solving the Poisson-Boltzmann equation or estimating the desolvation term of the Born equation (65,66). Semiempirical scoring functions use similar terms like first-principles methods but weigh these terms in order to tune the calculated energies towards the experimentally derived values. A common semiempirical scoring function, is the GOLD Scoring Function (67). Empirical scoring functions are directly derived from a training set of known complexes and their binding energies. They are constructed as a sum of energy contributions which account for the different interaction types. Each summand is penalized by a factor accounting for derivations from the respective ideal geometry. A common example for an empirical scoring function is implemented in the docking program FlexX (68). The last category – knowledge-based scoring functions – are derived from statistical analyses of the frequency of occurrence of atom-atom interactions observed in known complexes. The frequencies of occurrence can also be converted into free energies resulting in Potentials of Mean Force (PMF) which can also be used as scoring functions. In contrast to empirical scoring functions, the complexes' structures are sufficient for the determination of knowledge-based scoring function, no measured binding energies are necessary. A modern knowledge-based scoring function is DrugScore, developed by Gohlke *et al.* (69). The latter two categories are considered to be more general than regression-based scoring functions (semi-empirical scoring functions) because they implicitly account also for effects not understood so far (62). Even if visualization is not very often used in the development of scoring functions, it can nevertheless give helpful hints about the scoring function space. For examples, scoring functions values for specific problematic complexes can be visualized on a regular grid around the target by centering a rigid (or even flexible) ligand on each grid point and calculate the relative rotation (and conformation) with the strongest binding energy (Figure 11). In this way, in the case of a scoring function failure, the wrong global minima, not corresponding to the experimental structure, can be easily identified and then analyzed in detail hopefully leading to reasons for the failure. Additionally, special features of the scoring function like extreme roughness resulting in convergence problems can also be recognized by many adjacent local minima.

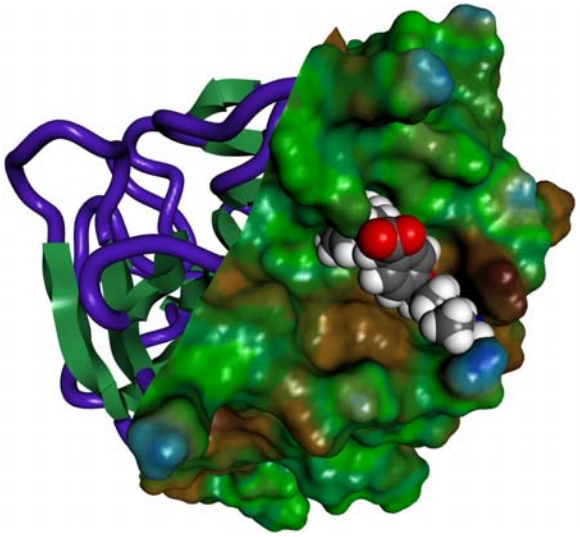

Figure 10. Molecular surface of the factor Xa – inhibitor complex. The representation is the same as in Figure 6. Additionally, the Connolly surface is partly shown color-coded by the local lipophilicity (blue: almost hydrophilic, brown: almost lipophilic). A cutting plane is used to hide the surface on the left side.

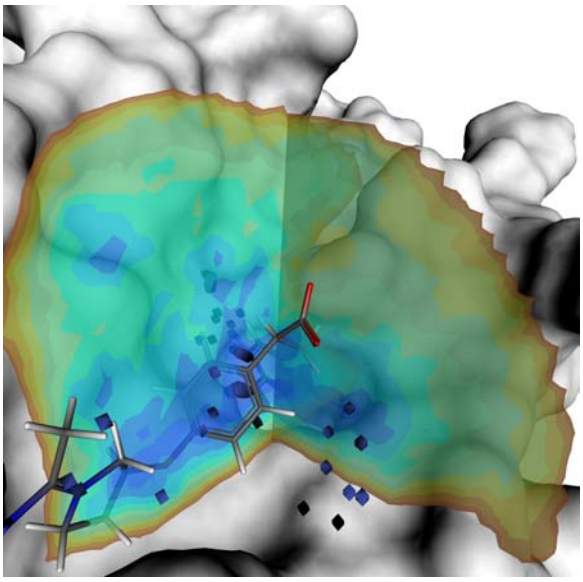

Figure 11. Fitness landscape for docking an inhibitor to coagulation factor Xa (PDB-entry 1fax) (70). The two slicing planes are color-coded from blue (low scoring function values) to red (high scoring function values). Additionally, blue isosurfaces mark low-scoring regions throughout the whole binding site.

Because, the best scored and highest ranked docking solution may not be the optimal one, which is due to the inaccuracy of the scoring functions, often a limited number of docking solutions is stored. The question is now, how to find the optimal solution in this set of possible solutions. Quite often, the researcher has already a clear idea of the desired binding mode, e.g. hydrogen bonds between the ligand and specific protein residues or the occupation of specific binding site pockets by the ligand. While it would be possible to analyze the results numerically the interpretation of the results is difficult. At this point intelligent visualization techniques are extremely helpful. Modern molecular visualization programs allow to sift through the different docking solutions and to inspect them visually. Hydrogen bonds and other important interactions are detected and visualized on the fly, which allows the researcher to quickly find the docking solutions which comply with the desired binding mode (see Figure 12a). Moreover, this technique allows to identify additional beneficial interactions, which probably had remained unnoticed when the results had only been analyzed numerically. However, not only the three-dimensional representation is used for the analysis of docking solutions but also a two dimensional representation of the protein-ligand contacts (71,72) can be of great use (Figure 12b). In this representation only the ligand and the protein residues bound by hydrogen bonds are depicted in 2D molecule representation. Residues in hydrophobic contact to the ligand are drawn as symbols and identified with their three-letter abbreviation and their residue number. Because the human eye is very good in identifying patterns in different images, the 2D interaction representations are well suited to identify docking solutions complying with a special binding mode. Moreover, they are especially used in reports and publications to explain the actual binding mode.

For more complicated active sites, the visualization of shape complementarity using standard techniques like molecular surfaces can be, on one hand, little informative or, on the other hand, very confusing. Even if these surfaces (if prepared carefully) highlight unfilled cavities or problematic regions with steric overlap, there is no quantitative information on the strength of the steric overlap or the size of the void space. In contrast, specially designed surfaces, lying exactly between the ligand and the protein (e.g. separating surface (76) (see Figure 13) and intersurf (77)) can be of great advantage in this respect, because e.g. the shortest distance between the two molecules has to be calculated during the generation anyways and, thus, can be directly color coded on them. Differences in both complex partners, e.g. variations in substituents of the ligand, dissimilarities in different homolog targets or even in structures determined by different experimental methods, will show up in the shape of the surface and the properties mapped onto them.

5.1. Textbox 4. Three-dimensional data fields

Physicochemical properties often depend on the position in 3D space in respect to the system under investigation at which they are measured or calculated. A common example is the electrostatic potential, which describes the interaction of a positive charge with the system at each point in space. The visualization of these 3D data sets has always been a challenge in interactive computer graphics applications. Several methods are used. Possibilities, which are capable of displaying 3D data directly, are voxel-based volume rendering or 3D texture mapping. In these representations, the data is visualized color-coded on a regular 3D grid (Figure 14)

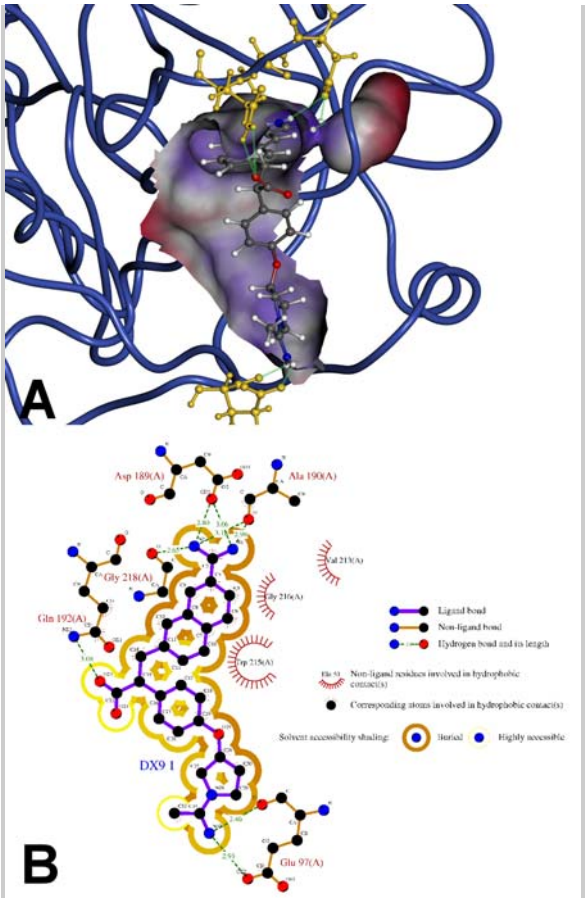

Figure 12. 3D (A) and 2D (B) visualization of the complex of coagulation factor Xa with inhibitor DX-9065a (73). In the 3D model the local lipophilicity is shown color coded on the molecular surface of the active site. Additionally, important residues of the target and hydrogen bond between the target and the ligand are highlighted in yellow and green, respectively. The 2D model was generated and visualized with the program LIGPLOT (71). Hydrogen bonds and solvent accessibilities were calculated with HBPLUS (74) and NACCESS (75), respectively.

These volume rendering techniques are well suited to obtain a general overview of the extrema of the property. But for more complex scenarios and for in-depth investigations of specific regions, theses become, on the one hand, very demanding in respect to the graphics hardware and, on the other hand, more and more confusing and meaningless due to the overlapping features. In such a case it is preferred to apply an intermediate step to derive a representation on a geometric object, e.g. a plane or a solid surface. So-called slicing planes sample the contents of the volume as if it were exposed by cutting the object with a knife (Figure 15). Multiple slicing planes or the interactive moving and rotating of one slicing plane can be used to get a feeling of the total 3D property.

As described above in Textbox 3, the molecular surfaces can be used as screens for the 3D data as well (Figure 16). Another approach is the generation of isovalue surfaces or short isosurfaces (Figure 16 and 17). These connect points in 3D space with the same value of the specific property. Prominent examples for the use of isosurfaces are the visualization of electron densities or of individual atomic or molecular orbitals.

6. LIGAND-BASED DESIGN

The methods described so far are dependent on the receptor's 3D structure. This 3D structure can either be derived experimentally by X-ray crystallography or high-resolution nuclear magnetic resonance spectroscopy or it can be derived by comparative homology modeling. Many interesting target proteins, however, e.g. GPCRs (G-protein-coupled receptors) or ion channels, cannot be crystallized and are not accessible to NMR-based structure elucidation and, therefore, receptor-based drug design methods are unsuited. At this point ligand-based drug design methods come into play. Common to all ligand-based methods is, that they derive information on the binding mode from the molecular structure of a number of known ligands. One such method is pharmacophore modeling and screening. The pharmacophore is the spatial arrangement of the functional groups of a ligand which are necessary for binding to its receptor (78) or the framework of molecular features that carries the pharmacological activity of the compound (79). Ligands that have a common binding mode, therefore, must share a set of molecular features. The idea behind pharmacophore modeling and screening is thus, to derive the common set of molecular features and their spatial arrangement from a set of known ligands, and to use this set of features to identify compounds in a compound database, which possess a similar set of features in a similar arrangement.

Although, pharmacophore models consider the spatial arrangement of the features, the bond network between the features is neglected. In addition, usually features are defined very generically, e.g. instead of an hydroxy-group (-OH) at a given position a hydrogen bond donor feature is placed, which can be satisfied by either a hydroxy-group or an amino-group ($-NH_2$). The method is therefore ideally suited for scaffold hopping. Scaffold hopping means the replacement of a given chemical core structure by a new core structure belonging to a different chemical class.

The most critical step in pharmacophore modeling is to create a hypothesis for how the ligands bind to the protein, because the energetically favorable conformation and configuration of a small chemical compound resulting from a conformational search must not necessarily be the biologically active one (80). Aligning the ligands in the wrong, i.e. the biologically inactive conformation or configuration would thus lead to an incorrect pharmacophore hypothesis. Sufficient conformational and configurational exploration is, therefore, at least equally important than it is in receptor-based methods, like molecular docking. Once the ligands are correctly aligned, common molecular features can be derived and translated into a pharmacophore hypothesis. In a second step, the pharmacophore hypothesis can be used to

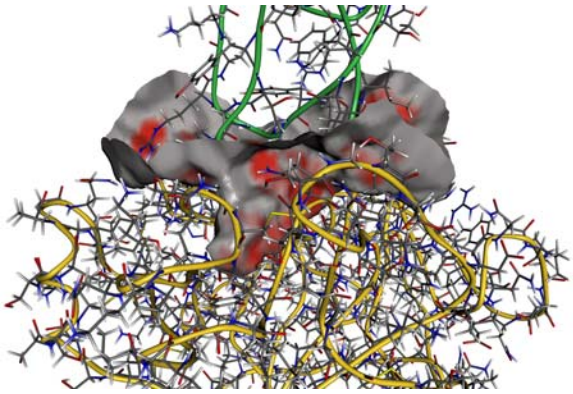

Figure 13. Separating surface between the structures of the Trypsin/BPTI complex (76). The surface shows the form of the active site and how well the inhibitor fits into the binding pocket. The surface is color coded by the distance between the atoms of both binding partners (red: close distance). No large bumping regions are visible, the red spots can be attributed to atoms within hydrogen bond contact distance.

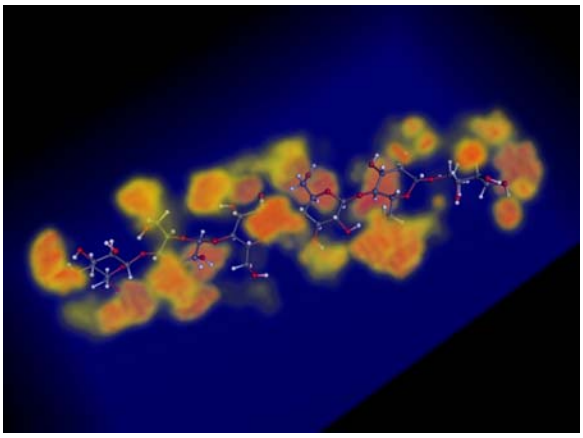

Figure 14. Volume rendering representation of the N,N,N-trimethylamine-N-oxide (TMAO) density distribution around cellohexaose in a 1:1 TMAO/water mixture (116) using 3D texture mapping. Large and low densities are shown in red and blue, respectively.

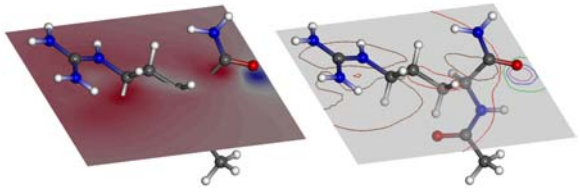

Figure 15. Electrostatic potential of the amino acid argentine shown on a slicing plane cutting through the molecule. On the left side, the values are color-coded from red (positive) to blue (negative). On the right side, only some isolines, connecting points with the same value (like contour line in geographical maps), are shown.

identify potential ligands in a compound database. Each individual compound is fitted onto the pharmacophore hypothesis and assessed according to the degree of overlap. Due to the three dimensional nature of pharmacophore models, the compounds in the databases must first be converted into 3D representations before the database can be screened. At this point ligand flexibility comes into play. The conformer generation is thus as crucial for this screening method as the pharmacophore search itself.

Almost every integrated molecular design package nowadays offers a program for ligand-based pharmacophor modeling. Tripos' SYBYL module DISCOtech (81) was recently supplemented by Galahad, which uses a newly developed genetic algorithm (82). Accelrys included HipHop (83) and HypoGen into its molecular design suite Catalyst and also the modeling package MOE (Chemical Computing Group, Inc.) has a pharmacophore modeling module included. While pharmacophore model building is generally based on the molecular features of the ligands, some of these programs also can take activity data into account.

Like other methods, pharmacophore screening has a couple of limitations. Quite often, not only one pharmacophore exists, but different pharmacophore models can be derived from subsets of the set of known ligands. In this case, one has to be careful not to derive a too general and therefore worthless model. For an extensive review of the pharmacophore screening approach and its limitations see Nicklaus (79) and Horvath *et al.* (84).

Pharmacophore modeling is a highly interactive process. Very often quite a few different pharmacophore models can be built from the same set of known ligands. It is, therefore, important to choose a model, which is sufficiently specific to identify potential inhibitors in a compound database, but which is not too specific, so that no novel compounds or only compounds from the same chemotype are identified. In addition, the quality of the model must be assessed by judging the topological fit between the model's features and the features of each of the ligands the model is build from. The topological fit is easiest judged by visual inspection, because it allows the researcher to include his chemical intuition into the process. The results from the pharmacophore screening are judged in the same way. The pharmacophore model is displayed as a set of spheres, cones or other geometric primitives and the potential inhibitors identified by the screening are fitted to the model as depicted in Figure 18. Analogous to the analysis process of molecular docking results, one can then sift through the different solutions and identify the most promising ones.

The major advantage of pharmacophore screening is its speed compared with virtual screening using molecular docking. However, usually pharmacophore screening does result in relatively large hit sets, which must be processed further to enter the biological assays. These hit sets can either be screened again using additional pharmacophore queries or can be submitted to a cluster

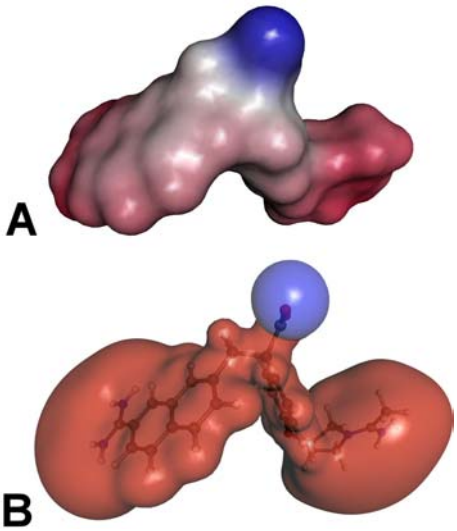

Figure 16. Electrostatic potential of the factor Xa inhibitor of Figure 2. In the upper half (A), the values are color-coded from red (positive) to blue (negative) on the molecular surface. In the lower half (B), two isosurfaces are shown (red: 0.1. a.u and blue:-0.1. a.u.).

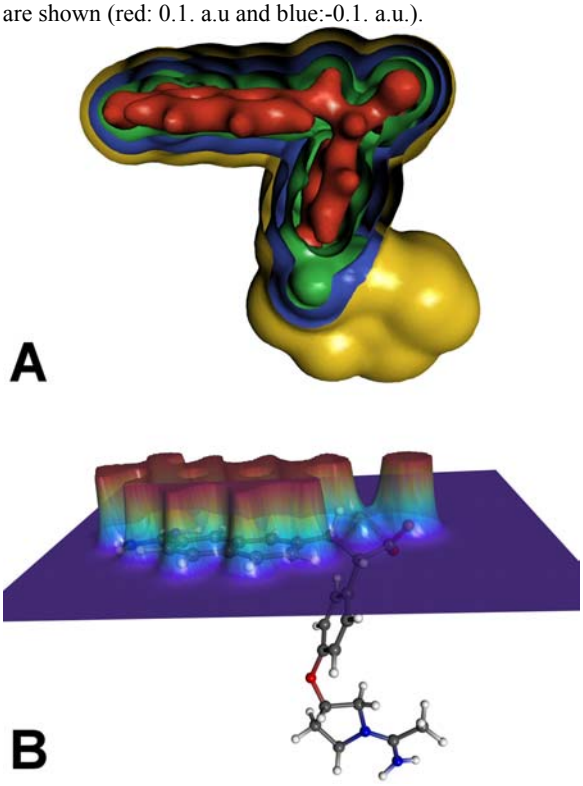

Figure 17. Electron density of the factor Xa inhibitor of Figure 2. Multiple isosurfaces are shown in (A) with decreasing isovalue (red: 0.1. a.u., green: 0.0.1 a.u., blue: 0.0.01 a.u., and yellow: 0.0.001 a.u.). To see the isosurfaces with higher isovalue, a cutting plane removes parts of the surfaces in the upper part. A more unconventional but nevertheless informative visualization of the electron density is shown in (B). It uses a cutting plane like Figure 15 but the values are shown as height in the third dimension as well as color coded. The very high electron cusps at the nuclear positions are truncated for clarity.

analysis in order to reduce the size of the hit set. The quality of pharmacophore screening varies with respect to the amount of information that is available about the active ligands. In cases, where enough information is available its performance is comparable to molecular docking (86).

Besides using pharmacophore screening as a stand-alone method it is often, due to its speed, used as a pre-filter method for virtual screening using molecular docking. In this case, the knowledge about the 3D structure can even be used within the pharmacophore concept (87,88). E.g. LigandScout (87) is a fully automated method to create a pharmacophore from known complex structures. In this way, the advantages of both worlds, structure-based and ligand-based screening are combined by (i) guaranteeing that the bioactive conformations of the inhibitors are used for model building, (ii) encoding more information about the complexes in the models, and (iii) allowing very high throughput due to the efficient point-matching algorithms used in the pharmacophore screens. Point (i) is accomplished by aligning the active sites in all complex structures. In this way, the ligands are placed so that the functional groups, which interact with the same groups in the protein, are aligned or at least are lying very close in space. Other possible combinations of functional groups leading to wrong models will not be considered making the pharmacophore much more reliable. An example for point (ii) is the inclusion of so-called excluded-volume spheres, which mark regions in space occupied by the protein (Figure 19). Ligands overlapping with these spheres are too large for the binding site and are energetically penalized by the pharmacophore model. Additionally, by overlaying the pharmacophore model with the experimental complex structure, it can be checked if each feature of the model is really paired with a corresponding functional group of the target protein. E.g a hydrogen-donor feature is only meaningful if a hydrogen-acceptor group of the protein is in close proximity. Otherwise it should be removed because it was generated just by chance and would guide the database search into the wrong direction.

All pharmacophore models described so far are based on an atomistic representation of the ligands. All atoms are assigned to classes of functional groups and combinations of atoms belonging to the same group are aligned to define a pharmacophore feature. But the physical fundamental of molecular recognition are not functional groups but the interaction field generated by the whole molecule, i.e. the overall sum of protein-ligand interactions in solution. This field is normally approximated by dividing it into different contributions, like the electrostatic potential, steric interactions, lipophilic interactions (defined as interactions with a CH_3 group), or solvation/desolvation (interactions with a water molecule). This is the foundation for 3D quantitative structure-activity relationships (3D-QSAR) using comparative molecular field analysis (CoMFA) (89,90,91). This procedure involves statistical methods to compute the contributions of e.g. steric or electrostatic molecular fields, providing physical parameters which can then be correlated to specific

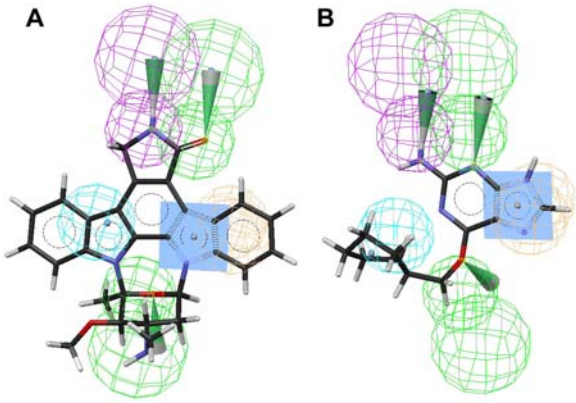

Figure 18. Pharmacophore model for CDK-inhibitors generated by the program Catalyst (85). The molecular features are depicted as geometrical bodies, H-bond donors are represented as a set of two green spheres and a cone indicating the direction of the H-bond. H-bond acceptors are represented in the same way, but with purple spheres. Blue spheres represent hydrophobic features and ocre-colored spheres stand for aromatic, hydrophobic features. The blue square represents the plane of the aromatic system. The two inhibitors Staurosporine (A) and Purvalanol (B) are mapped into the pharmacophore model. Although, both inhibitors match the pharmacophore hypothesis, the agreement of the molecular features and the model is worse for Purvalanol than it is for Staurosporine, which is indicated by the length and direction of the H-bond cones and the small solid blue sphere, representing the location of the hydrophobic feature.

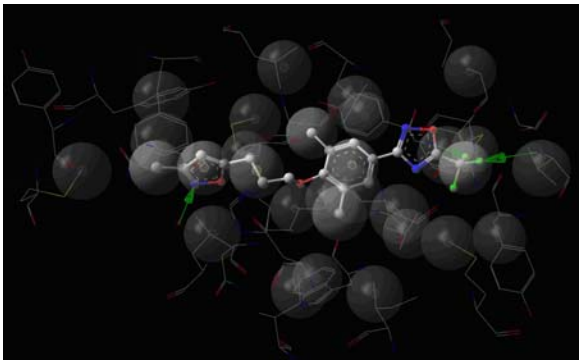

Figure 19. Pharmacophore model of the PDB entry 1ncr aligned with Pleconaril and the rhinovirus protein hull as automatically generated by LigandScout (87). The excluded-volume spheres are shown in dark gray. Reproduced in part with permission from (87). Copyright (2005) American Chemical Society.

biological properties of the molecules. The ideal case would be to align the molecules based on these molecular fields. But then the fields must be calculated for each conformation of the possibly very flexible molecules and a similarity measure must identify the best overlap of these fields for each relative translation and rotation of the molecules. This is by far too time consuming so that in the standard procedure the molecules are first aligned on the atomistic level and only the molecular fields for this alignment are correlated. Visualization of the results may then improve the understanding of the interactions and help in designing compounds with improved activity without the bias of a specific structure generating the changes in the molecular fields. This will be demonstrated with two examples taken from the literature (Figure 20 and Figure 21). Beside the illustration of the method, these Figure s also show the huge improvement of clarity which can be obtained with raster graphics (Figure 21) in comparison to older vector graphics (Figure 20).

6.1. Textbox 5. Three-dimensional vector fields

For 3D vector fields, like the electric field surrounding a molecule or a reaction field determining the most favorable direction of an incoming reaction partner, yet additional visualization techniques are needed to capture the size as well as the direction of the field at a certain point. The vectors can be visualized using small cones or arrows positioned on a regular grid (Figure 22). The size and/or the color indicate the field strength and the orientation the field direction. Another method to visualize vector fields uses field lines (sometimes also called “lines of force”), which are drawn as curves so that the tangent line to the curve at an arbitrary point is directed along the vector of the electric field at this point, and the density of lines is directly proportional to the magnitude of the field (Figure 22).

7. LIGAND OPTIMIZATION

As described in the OVERVIEW OF THE DRUG DESIGN PROCESS section, after the identification of lead structures using virtual screening followed by experimental screening, these structures are chemically modified to optimize their activity as well as other required characteristics like selectivity and ADME-Tox parameters. This is the most complicated part of the rational design, because an extreme amount of expert knowledge on the experimental as well as the theoretical side is needed. The structures are looped through a number of design steps building the lead optimization cycle. First, possible regions of the ligand, in which modifications could increase the needed characteristics, are identified. Different substituents are then added and tested applying chemoinformatics methods, like pharmacophore modeling, molecular docking, QSAR analyses etc. As mentioned, not only the binding affinity is important at this stage but also all other properties characterizing the drug-likeness of a substance. The most promising modifications (and sometimes also some of the less potent ones in order to check a hypothesis) are synthesized and experimentally tested. Using the best candidates as lead structures the cycle is then started again until no further improvements can be obtained. Last but definitely not least, the drug candidate will go into the pre-clinical and clinical test phase. A failure at this clinical stage can be disastrous for a pharmaceutical company due to the high cost of these tests. It is extremely important that the chance for success of a drug candidate is already maximized, which is tried to guarantee by large scale experimental and computational testing during the whole

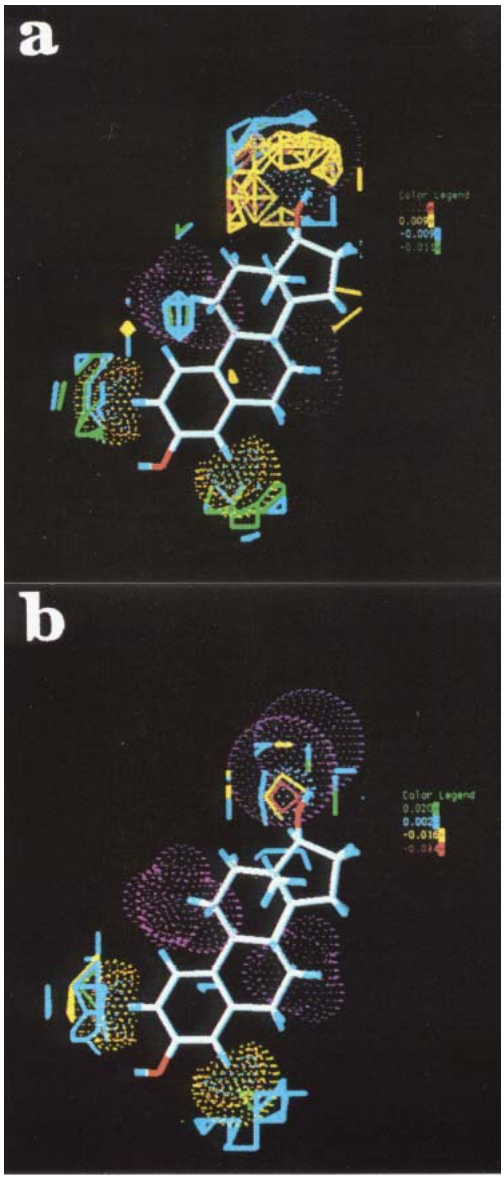

Figure 20. (a) Visualization of CoMFA steric field contours of receptor-binding properties of halogenated estradiol derivates. The green and cyan polyhedra indicate regions where lower steric interaction would increase the binding affinity. The red and yellow contours surround regions where higher steric interactions would increase binding affinity. Dotted clouds represent vdW surfaces of the 21 iodine atoms with 20E- (magenta) and 20Z- (purple) configurations; (b) Visualization of the CoMFA electrostatic field contour. The red and yellow contours indicate regions where addition of negative charge would increase binding affinity. Cyan and green contours indicate regions where addition of positive charge would increase the binding affinity. Dotted clouds represent vdW surfaces as indicated in (a). Reproduced in part with permission from (92). Copyright (1994) American Chemical Society.

design process leading up to these test phases. Because there are so many fields involved especially in lead optimization covered in this section, we cannot describe all of them here in detail. Therefore, we will concentrate on a few examples, in which visualization can be a surplus.

Theoretical methods are at the moment not reliable enough to be a suitable replacement for the experiment. Compared to the identification of lead compounds this is even more the case during lead optimization, because here pharmacokinetics, i.e. how is the drug taken up by, distributed throughout, and removed from the body, must be considered as well. The human being is such a complex system that no theoretical model will exist for it in the foreseeable future. Therefore, computational methods can only concentrate on specific parts. Nevertheless, good progress has been made in the last decades. Many of the techniques described so far are also used in lead optimization. Because a specific class of compounds was already identified as lead, more computational time can be invested for describing their binding properties or their interactions with other key players in the life cycle of a drug, for examples cytochromes P450 involved in almost all routes of drug metabolism.

The first example is dealing with possible improvements of the description of the complex between the target and the ligand. The model of the protein used in molecular docking is very basic. It is treated as a rigid body witout any solvents surrounding it. In some complexes the interaction of the ligand with the protein is mediated through a water molecule. Therefore, newer docking algorithms can include these essential water molecules and remove them on the fly, if they are replaced by a part of the ligand (94,95,96,97). Additionally, induced fit can be simulated in more advanced algorithms. This is mainly done by making specific side chains flexible but sometimes also backbone flexibility is considered. The easiest way to Figure out which side chains adapt to the incoming ligand, is to overlay different complex structures of the same target with different ligands (holo forms) and/or the uncomplexed structure (apo form). Overlaps between ligands of one complex structure and side chains of another one are strong evidence that all known ligands can only be docked correctly into the binding site if a flexible approach is used (see Figure 23).

The scoring functions used in docking algorithms are at the moment the main reason for failures to predict the correct complex structure. But the improvement of these general-purpose functions is not an easy task due to the two goals which have to be achieved at the same time: accuracy and computational efficiency. In lead optimization, however, one is not interested in a general-purpose docking approach, but in the best performing algorithm for the specific target. Weak to medium binders are already known for the target from the lead finding stage and experimental structures for the binders are solved in most cases. All this information can be used to generate a tailored scoring function, i.e. a function which is trained to best reproduce all known experimentally determined complex structures of a specific target (98,99,100,101,102). The training procedure is equivalent to those of general-purpose

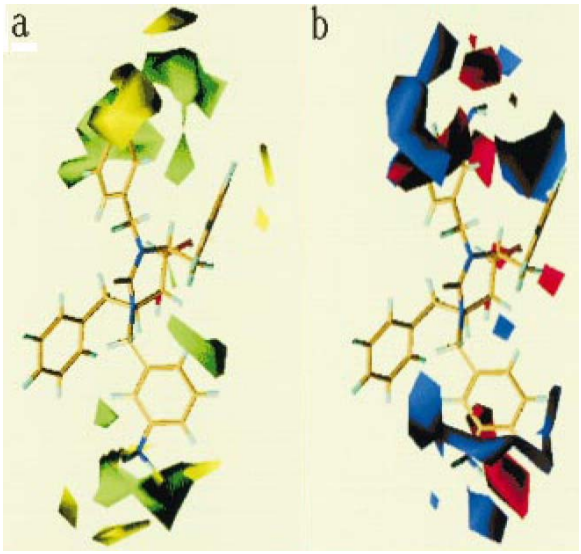

Figure 21. Visualization of the CoMFA contour plots of (a) steric fields taken from a 3D-QSAR study on cyclic urea derivatives as HIV-1 protease inhibitors, the green polyhedra (80% contribution) represent areas where bulkier groups may enhance activity, whereas yellow polyhedra (20% contribution) indicate areas where bulk may have detrimental effects on activity; (b) electrostatic fields, the favored electrostatic areas with positive charges are indicated by blue polyhedra (80% contribution), whereas the favored electrostatic areas with the negative charges are indicated by red polyhedra (20% contribution). One of the most active compounds, DMP450, is shown as the reference compound. Reproduced in part with permission from (93). Copyright (1999) American Chemical Society.

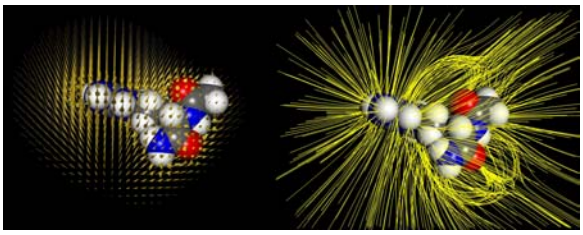

Figure 22. Electric field around the amino acid arginine shown as small cones (left) and as field lines (right).

unctions only the training set is not chosen as diverse but as specific as possible. If not enough experimental complex structures are known for the target, very similar targets should be taken into account to minimize the problem of overfitting. After the training, the new function can be applied for a more focused screening. Additionally, the visualization of the differences in the general-purpose and tailored scoring functions (see Figure 24**)** will highlight the important interactions between the protein and the ligands, on which following optimization cycles can concentrate.

If a single or a small number of complexes should be explored even more accurately, molecular dynamics simulations can be used (see related articles in this issue). Full protein flexibility and an explicit or implicit solvent model will increase the accuracy of the calculations but also increase the computational demand drastically. In favorable cases, even quantitative binding free energies can be obtained using free-energy perturbation ((17) and references therein), thermodynamic integration ((17) and references therein), or the molecular mechanics – Poisson Boltzmann / surface area (MM-PBSA) approach (103).

The second example for visualization in lead optimization is dealing with another important property of a drug: specificity. If a ligand is binding strongly to one target, the chance that it also binds and inhibits other similar proteins is very high resulting in side effects. There is no drug without side effects but the minimization of these is highly desirable and can seal the fate of a drug. One prominent example are kinases, which transfer phosphate groups from high-energy donor molecules, such as ATP, to specific substrates. At least 500 distinct kinases, which can be grouped into roughly 20 known families on the basis of structural relatedness, have been sequenced in the human genome (104). Many of them are excellent targets for developing new drug candidates and treatment strategies for major diseases like cancer, autoimmune disorders, vascular diseases and degenerative brain diseases. This has led to the fact that at the moment kinases are beside GPCRs the most prominent target family in pharmaceutical research despite many associated concerns: The high intracellular ATP concentrations versus ATP site-directed inhibitors; a common catalytic mechanism across the many families of kinases; structural similarity of other features of kinase enzyme active centers; and the importance of kinase activities to many, totally unrelated physiological processes (104). These concerns create obligations for highly specific inhibitors. It is not possible to examine the interactions of a drug candidate with all possible kinases, but the visualization of the differences of the target kinase with some highly homologues ones can identify regions of the active site, in which interactions could lead to higher specificity (see Figure 25). For doing so, one starts with a superposition of the various enzymes. Interaction fields, like GRID (50) or FLOG (105) maps, are calculated around each individual enzyme on a common grid for all homologs. E.g. FLOGTV (105) uses five probe types by default: donor/cation, acceptor/anion, polar, hydrophobic, and all other atoms feeling only van der Waals interactions. So-called trend vectors are then calculated as a weighted sum of the interaction energy with a specific probe multiplied by the normalized activity, capturing the differences in map space between desirable and undesirable enzymes. Large positive values correspond to regions, in which an atom of the specific probe type should be placed for higher selectivity. By contrast, a placement of the atoms of this type should be avoided in regions with large negative values. Therefore, contouring of the maps with the help of isosurfaces will directly draw the attention to the regions important for specificity, which are very hard to determine from looking at the structure alone (see Figure 25).

These are only two examples subjectively chosen by the authors. Many others could be given especially

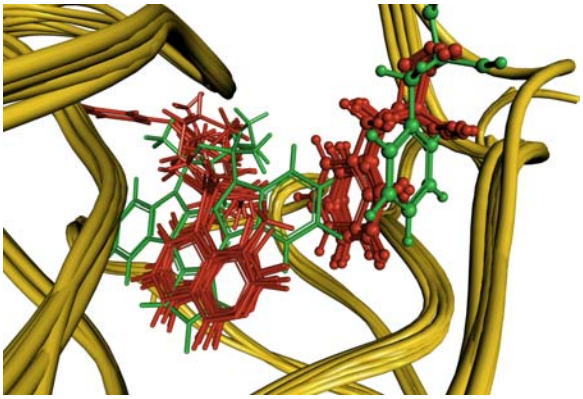

Figure 23. Superimposition of the eight protein kinase A structures (the protein backbones are illustrated as ribbon) (70). The side-chain conformations of residue PHE327 for each protein structure are shown as green and red balls-and-sticks (for the sake of clarity all other residues are hidden). Ligand staurosporine (shown as green wire frame in the middle along with all other ligands shown in red) clashes with this residue in all non-native structures

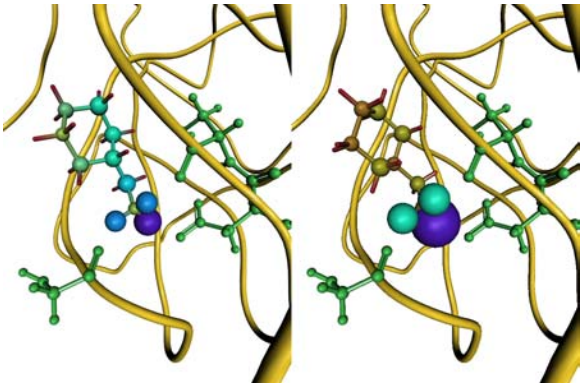

Figure 24. Per-atom contributions to the total scoring function value of a general purpose scoring function (left) and a target-specific scoring function (right) (70). Bigger spheres contribute more than smaller ones to the total scoring function value. The score values per atom are additionally color-coded from blue (low scoring function values) to red (high scoring function values). The trypsin protein structure (PDB-entry 1tnh) is represented as tubes. Key-residues interacting with the ligand are shown in green.

demonstrating the advantages of modern visualization techniques for the communications between computational and experimentally working medicinal chemists (see Textbox 6). Combined with high-performance computer clusters, more and more exact but also demanding methods like the already mentioned molecular dynamics simulations or ab initio quantum mechanical calculations on larger systems will enter into the rational drug development process. At some point of time, it will hopefully be possible to perform the calculations on computer clusters on the fly and then visualize them according to the needs of an ongoing discussion. Interactive docking experiments were already performed in a cave, in which the computer gives feedback about the docking pose using haptic devices such as force-feedback joysticks and graphical effects (106,107,108).

7.1. Textbox 6. Interactive modeling

Images are well suited for the presentation of final results in publications or at conferences and meetings. These results are worth that much time can be invested into the right presentation. The author can choose the right view and abstraction of the molecular data to highlight his findings. But in ongoing research, it is not clear what is important to look at. Therefore, the molecular scenario must be inspected from many different angles and with different resolutions and representations to generate new ideas. Today's graphic systems are fast enough to generate all representations described in this paper from existing, pre-calculated data in real time, which makes interactive modeling possible without annoying waiting periods. Even most of the calculations for the advanced representations, like ribbons, molecular surfaces and slicing planes, can be done one the fly. In this way, discussions of groups of researches can be supported directly with the interactive visualization of the data at hand. Different representations can be chosen to fit the needs of each individual researcher. This describes a scenario in which the researchers are in a common room using the same computer display. But in many cases, collaborations are established between groups a long way apart from each other sometimes distributed all over the world. Then it is important to transport the visual information from one site to the other. This can be done by images or short videos or animated gifs, which can be automatically produced by many visualization programs. Even if multiple images or animations can pinpoint the chain of reasoning of the researcher producing the visualization, they are still limited to the ideas of the producer, because they cannot be changed afterwards. Nevertheless, they are especially helpful for dynamic data as generated e.g. during molecular dynamics calculations (see related articles in this issue).

Scripting is another possible way to transmit visualization data. Many programs (RasMol (117) as prominent long existing examples) have the ability to access their functionality, besides the graphical interface, within a text console. In this way, multiple commandos can be combined in a script and then executed one after each other. On one hand, complex visualization and analysis procedures can be automated in this way. E.g. in the program VMD (10) the scripting capability is so powerful that even totally new functionality can be included in the program using the Tcl or Python scripting language. On the other hand, the scripts can be transmitted with the underlying data to another research site. At the beginning of a joint discussion, researchers at different sites can load the scene in exactly the same orientation and representation but can then manipulate it to their own needs. Using plug-ins to standard web-browsers the communication can be done directly over the intra- or internet. In the early stages, the Virtual Reality Modeling Language (VRML) developed for web-based visualization of all kinds of 3D scenes was

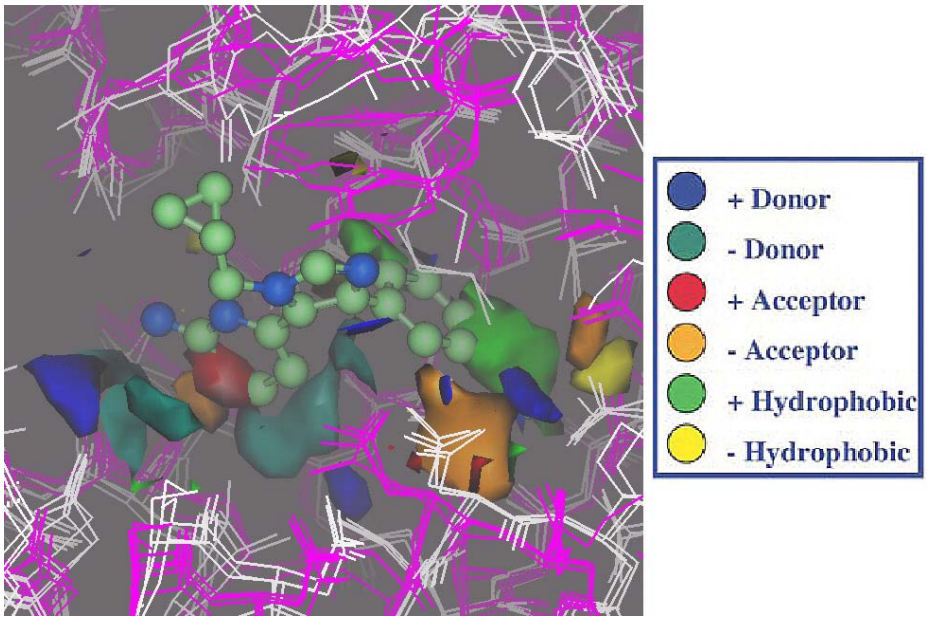

Figure 25. Trend vector maps of the comparison of P38 vs. ERK protein kinase as calculated by the program FLOGTV (105). P38 kinase are shown as purple wire, ERK kinase as white wire. The inhibitor SB-218655 is shown as ball and stick. Positive contours are where the specified probe would selectively bind to P38: kinase: donor/cation, blue; acceptor/anion, red; hydrophobic, green. Negative contours are where the probe would selectively bind to ERK kinase: donor/cation, blue–gray; acceptor/anion, orange; hydrophobic, yellow. Reproduced in part with permission from (105). Copyright (2002) Elsevier Limited.

used (2,59) (Figure 26). The advantage was that almost all features of a visualization software, including surfaces, slicing planes, etc., can directly be translated into a VRML scene. Problems are that the scenes are complicated to generate and cannot be manipulated afterwards (expect the normal rotation, translation and zoom) as well as the inconsistencies between different VRML players and standards. In the last time, more and more plug-ins specially designed for transferring chemical information are developed. The probably best know one is MDL Chime. Chime masters all standard representations and simplified models for bio-macromolecules as well as Van der Waals surfaces and hydrogen bonds. Another example is the Chem3D plug-in from CambridgeSoft. Additionally, the platform-independent programming language JAVA has let to a large number of chemical applets with various capacities (118,119), e.g. JME, Marvin Sketch/View, Jmol, JChemPaint, and WebMol to name just a few.

Some programs have the capability to automatically log all actions performed by the user using their scripting language. These log files can be used to go back to a specific time of the modeling session or restore the scene at a later occasion or after a crash of the system. Additionally, these automatically generated scripts can be used in joint discussion as described above. Then, even the manipulations done at one site can be reproduced at the other by transferring the output of the logging and executing all logged commands. An even more advanced version of this technique is based on a server-client application. Here, all commands are sent to the server, which distributes them to all the clients. In this way, users at all sites can manipulate the scene and view the results from these manipulations. The server-client architecture is also needed for visualizations beyond the normal computer screen. The probably most impressive visualization hardware is the CAVE virtual reality system. The images are projected onto three walls, the ceiling, and/or the floor of a small room build out of rear-projection screens. Each of the projectors is controlled by a client getting the needed data for the specific viewing angle from the server process. The user will go inside the CAVE wearing special glasses creating the 3D impression of the images. This is accomplished by rendering two images with a little offset in the viewing angle, which are drawn in turn, one for the right and one for the left eye, and by the glasses preventing the other eye from seeing the image not ment for it. In this way, our human brain is tricked to see a three-dimensional image in which one can walk around. To interact with the scene, special input devices like spaceballs, joy sticks, or the CAVE wand are used. With a lower budget, the 3D glasses, e.g. shutter glasses, can also be combined with a standard monitor (or one or multiple projectors for classrooms or tiled-display theaters).

8. CONCLUSION

Everyday new information is generated which can be visualized using existing or new visualization techniques. This review shows with a number of examples, taken from the rational drug design field, how well-designed presentation of molecular data using state-of-the-art techniques can help to judge numerical results and outputs of modeling programs and to discuss experimental

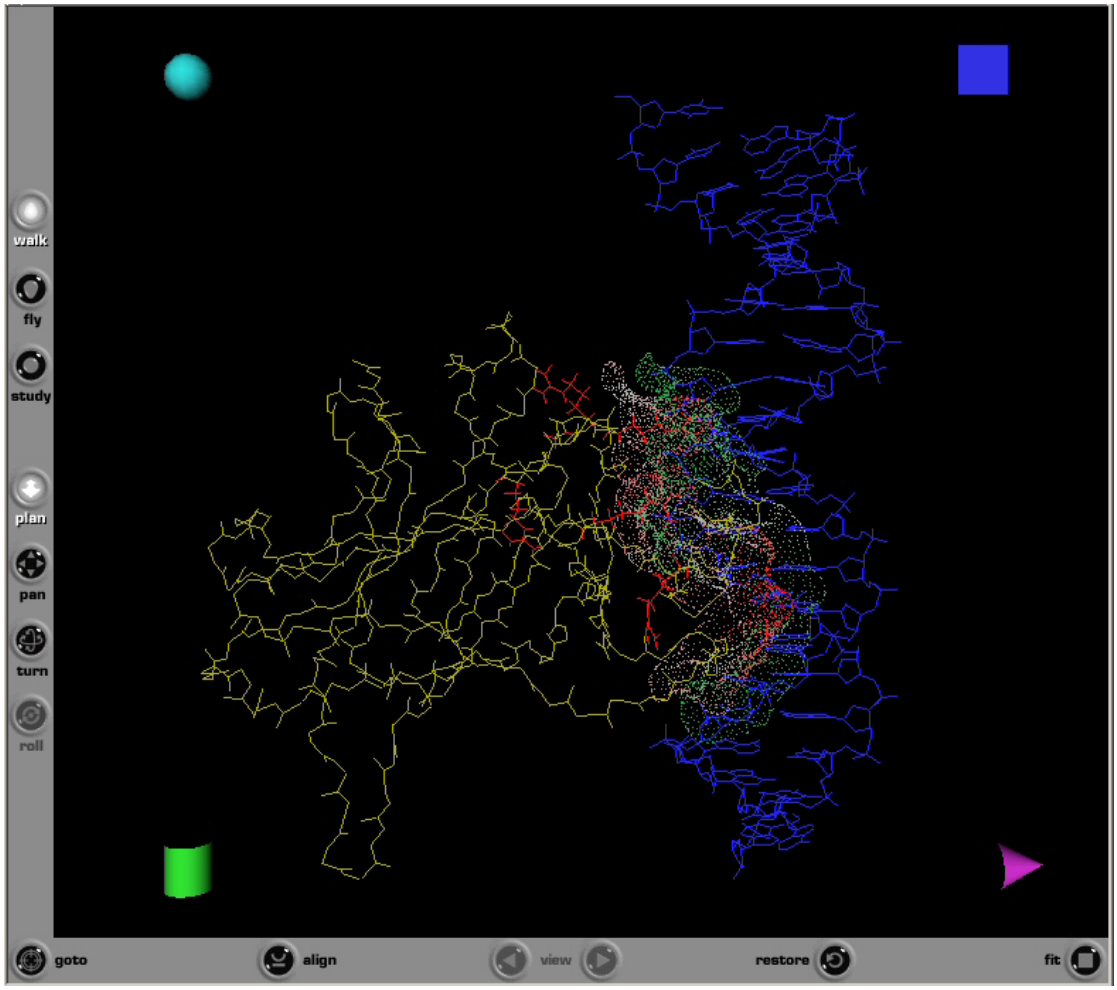

Figure 26. VRML scene of the p53 protein-DNA complex (pdb entry 1tup) (60) displayed with the Cortona VRML Client from ParallelGraphics.

and theoretical findings. Computer Graphics and its applications in molecular science have developed and hopefully will develop further very rapidly in the near future. However, this can not be achieved alone by advances in the hardware and software development. Only if one focuses on what should be represented in order to obtain a maximum of information from the underlying data and to get an optimal insight from the image, new ideas can be generated from and communicated with (beautiful and appealing) images. For doing so, on the one hand, new graphical representation forms have to be found in order to optimize the man-machine communication and also the communication between humans over globally accessible networks. The direct, interactive experience with three-dimensional objects by visualization and manual interaction has and will increase the efficiency of this communication. On the other hand, the preprocessing of the data becomes more and more important especially if the scenes are getting increasingly complex. In almost all examples given here, not the primary data is directly visualized, but it is first analyzed with more or less time-consuming algorithms specially optimized for generating secondary data for the visualization of the most important findings. Especially here, only the interplay between many disciplines, like computer science, theoretical, computational but also experimental chemistry, biology, and above all arts, will lead to further improvements of the field.

9. ACKNOWLEDGEMENT

We would like to thank Gerhard Wolber and Robert P. Sheridan for providing the original images from their publications.

10. REFERENCES

1. Brickmann J, M. Keil, T. E. Exner, R. Marhöfer: Molecular Graphics - Trends and Perspectives. *J Mol Model* 6, 328-340 (2000)

2. Brickmann J, M. Keil, T. E. Exner, R. Marhöfer, G. Moeckel: Molecular Models: Visualization. In: The Encyclopedia of Computational Chemistry. Eds.: Schleyer, P. v. R., Allinger, N. C., Clark, T., Gasteiger, J., Kollmann,

P. A., Schaefer III, H. F., and Schreiner, P. R., *John Wiley & Sons*, Chichester, (1998)

3. Keil M, T. Borosch, T. E. Exner, J. Brickmann: Computer Visualization - Tools for the Man-Machine-Communication in Molecular Science –. In: Chemoinformatics - From Data to Knowledge. Eds.: Gasteiger, J. and Engel, T., *Wiley-VCH*, Weinheim, (2003)

4. Atwood J L, L. J. Barbour: Molecular Graphics: From Science to Art. *Cryst Growth Des* 3, 3-8 (2002)

5. Ferrin T E, T. E. Klein: Computer Graphics and Molecular Modeling. In: The Encyclopedia of Computational Chemistry. Eds.: Schleyer, P. v. R., Allinger, N. C., Clark, T., Gasteiger, J., Kollmann, P. A., Schaefer III, H. F., and Schreiner, P. R., *John Wiley & Sons*, Chichester, (1998)

6. Goodsell D S: Looking at Molecules - An Essay on Art and Science. *ChemBioChem* 4, 1293-1298 (2003)

7. Ihlenfeldt W-D: Virtual Reality in Chemistry. *J Mol Model* 3, 386-402 (1997)

8. Olson A J, M. E. Pique: Visualizing the Future of Molecular Graphics. *SAR QSAR Environ Res* 8, 233-247 (1998)

9. Olson A J: Molecular Graphics and Animation. In: International Tables for Crystallography, Vol. F. Eds.: Rossmann, F. M. and Arnold, E., *Kluwer Academic Press*, Dordrecht, The Netherlands, (2001)

10. Humphrey W, A. Dalke, K. Schulten: VMD: Visual Molecular Dynamics. *J Mol Graphics* 14, 33-38 (1996)

11. Waldherr-Teschner M, T. Goetze, W. Heiden, M. Knoblauch, H. Vollhardt, J. Brickmann: MOLCAD - Computer Aides Visualization and Manipulation of Models in Molecular Science. In: Second Eurographics Workshop on Visualization in Scientific Computing, Delft, Netherlands. (1991)

12. SYBYL 7.3. *Tripos International,* 1699 South Hanley Road, St. Louis, MO, USA, (2007)

13. Watson J, F. H. CRICK: Molecular Structure of Nucleic Acids; a Structure for Deoxyribose Nucleic Acid. *Nature* 171, 737-738 (1953)

14. Klebe G: Virtual Ligand Screening: Strategies, Perspectives and Limitations. *Drug Discov Today* 11, 580-594 (2006)

15. Hillisch A, R. Hilgenfeld: The Role of Protein 3D-Structures in the Drug Discovery Process. In: Modern Methods of Drug Discovery. Eds.: Hillisch, A. and Hilgenfeld, R., *Birkhäuser*, Basel, Switzerland, (2003)

16. Fara D C, T. I. Oprea, E. R. Prossnitz, C. G. Bologa, B. S. Edwards, L. A. Sklar: Integration of Virtual and Physical Screening. *Drug Discov Today: Technologies* 3, 377-385 (2006)

17. Pearlman D A, B. G. Rao: Free Energy Calculations: Methods and Applications. In: The Encyclopedia of Computational Chemistry. Eds.: Schleyer, P. v. R., Allinger, N. C., Clark, T., Gasteiger, J., Kollmann, P. A., Schaefer III, H. F., and Schreiner, P. R., *John Wiley & Sons*, Chichester, (1998)

18. Muegge I, S. Oloff: Advances in Virtual Screening. *Drug Discov Today: Technologies* 3, 405-411 (2006)

19. The UniProt Consortium: The Universal Protein Resource (UniProt). *Nucleic Acids Res* 35, D193-D197 (2007)

20. Berman H M, J. Westbrook, Z. Feng, G. Filliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne: The Protein Data Bank. *Nucleic Acids Res* 28, 235-242 (2000)

21. Tramontano A: Protein Structure Prediction – Concepts and Applications. *Wiley-VCH*, Weinheim (2006)

22. Hilisch A, L. F. Pineda, R. Hilgenfeld: Utility of Homology Models in the Drug Discovery Process. *Drug Discov Today* 9, 659-669 (2004)

23. Lesk A M, C. Chotia: The Response of Protein Structures to Amino-Acid Changes. *Philos Trans R Sc Lond B Biol Sci* 317, 345-365 (1986)

24. Shi J, T. L. Blundell, K. Mizuguchi: FUGUE: Sequence-Structure Homology Recognition Using Environment-Specific Substitution Tables and Structure-Dependent Gap Penalties. *J Mol Biol* 310, 243-257 (2001)

25. Greaves R, J. Warwicker: Active Site Identification Through Geometry-Based and Sequence Profile-Based Calculations: Bruial of Catalytic Clefts. *J Mol Biol* 349, 547-557 (2005)

26. Halgren T: New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chem Biol Drug Des* 69, 146-148 (2007)

27. Konc J, D. Janezic: Protein-Protein Binding-Site Prediction by Protein Surface Structure Conservation. *J Chem Inf Model* 47, 940-944 (2007)

28. Soga S, H. Shirai, M. Kobori, N. Hirayama: Use of Amino Acid Composition to Predict Ligand-Binding Sites. *J Chem Inf Model* 47, 400-406 (2007)

29. Landon M R, D. R. Jr. Lancia, J. Yu, S. C. Thiel, S. Vajda: Identification of Hot Spots Within Druggable Binding Regions by Computational Solvent Mapping of Proteins. *J Med Chem* 50, 1231-1240 (2007)

30. Keil M, T. E. Exner, J. Brickmann: Pattern Recognition Strategies for Molecular Surfaces III. Binding Site Prediction With a Neural Network. *J Comp Chem* 25, 779-789 (2003)

31. Bogan A A, K. S. Thorn: Anatomy of Hot Spots in Protein Interfaces. *J Mol Biol* 280, 1-9 (1998)

32. Brady G P, P. F. W. Stouten: Fast Prediction and Visualization of Protein Binding Pockets With PASS. *J Comput -Aided Mol Des* 14, 383-401 (2000)

33. Chothia C, J. Janin: Principles of Protein-Protein Recognition. *Nature* 708(1975)

34. Danziger D J, P. M. Dean: Automated Site-Directed Drug Design: a Gerneral Algorithm for Knowledge Acquisition About Hydrogen-Binding Regions at Protein Surfaces. *Proc Roy Soc Lond B Bio* 236, 101-113 (1989)

35. Elcock A H, D. Sept, J. A. McCammon: Computer Simulation of Protein-Protein Interfaces. *J Phys Chem B* 105, 1504-1518 (2001)

36. Janin J, S. Miller, C. Chothia: Surface, Subunit Interfaces and Interior of Oligomeric Proteins. *J Mol Biol* 204, 155-164 (1988)

37. Janin J, C. Chothia: The Structure of Protein-Protein Recongnition Sites. *J Biol Chem* 265, 16027-16030 (1990)

38. Jones S, J. M. Thornton: Protein-Protein Interactions: a Review of Protein Dimer Structures. *Prog Biophys Mol Biol* 63, 31-65 (1995)

39. Jones S, J. M. Thornton: Principles of Protein-Protein Interactions. *Proc Natl Acad Sci USA* 93, 13-20 (1996)

40. Jones S, J. M. Thornton: Analysis of Protein-Protein Interaction Sites Using Surface Patches. *J Mol Biol* 272, 121-132 (1997)

41. Jones S, J. M. Thornton: Prediction of Protein-Protein Interaction Using Patch Analysis. *J Mol Biol* 272, 133-143 (1997)

42. Jones S, P. van Heyningen, H. M. Berman, J. M. Thornton: Protein-DNA Interactions: A Structural Analysis. *J Mol Biol* 287, 877-896 (1999)

43. Laskowski R A: SURFNET: A Program for Visualizing Molecular Surfaces, Cavities, and Intermolecular Interactions. *J Mol Graphics* 13, 323-330 (1995)

44. Laskowski R A, N. M. Luscombe, M. B. Swindells, J. M. Thornton: Protein Clefts in Molecular Recognition and Function. *Protein Sci* 5, 2438-2452 (1996)

45. Laskowski R A, J. M. Thornton, C. Humblet, J. Singh: X-SITE: Use of Empirical Derived Atomic Packing Preferences to Identify Favourable Interaction Regions in the Binding Sites of Proteins. *J Mol Biol* 259, 175-201 (1996)

46. Peters K P, J. Fauck, C. Frömmel: The Automatic Search for Ligand Binding Sites in Proteins of Known Three-Dimensional Structure Using Only Geometric Criteria. *J Mol Biol* 256, 201-213 (1996)

47. Ruppert J, W. Welch, A. N. Jain: Automatic Identification and Representation of Protein Binding Sites for Molecular Docking. *Protein Sci* 6, 524-533 (1997)

48. Verdonk M L, J. C. Cole, P. Watson, V. Gillet, P. Willett: SuperStar: Improved Knowledge-Bases Interaction Fields for Protein Binding Sites. *J Mol Biol* 307, 841-859 (2001)

49. Wang R, L. Liu, L. Lai, N. Ben Tal, Y. Tang: A New Empirical Method for Estimationg the Binding Affinity of Protein-Ligand Complexes. *J Mol Model* 4, 379-394 (1998)

50. Goodford P J: A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J Med Chem* 28, 849-857 (1985)

51. Delaney J S: Finding and Filling Protein Cavities Using Cellular Logic Operations. *J Mol Graphics* 10, 174-177 (1992)

52. Kellogg G E, S. F. Semus, D. J. Abraham: HINT: A New Method of Empirical Hydrophobic Field Calculation for CoMFA. *J Comput -Aided Mol Des* 5, 545-552 (1991)

53. Klebe G: The Use of Composite Crystal-Field Environments in Molecular Recognition and the De Novo Design of Protein Ligands. *J Mol Biol* 237, 212-235 (1994)

54. Laskowski R A, J. M. Thornton, C. Humblet, J. Singh: X-SITE: Use of Empirical Derived Atomic Packing Preferences to Identify Favourable Interaction Regions in the Binding Sites of Proteins. *J Mol Biol* 259, 175-201 (1996)

55. Laurie A T, R. M. Jackson: Q-SiteFinder: An Energy-Based Method for the Prediction of Protein-Ligand Binding Sites. *Bioinformatics* 21, 1908-1916 (2005)

56. Exner T E, M. Keil, G. Moeckel, J. Brickmann: Identification of Substrate Channels and Protein Cavities. *J Mol Model* 4, 340-343 (1998)

57. Arrowsmith C, P. Morin: New Insights into P53 Function From Structural Studies. *Oncogene* 12, 1379-1385 (1996)

58. Hollstein M, D. Sidransky, B. Vogelstein, C. Harris: P53 Mutations in Human Cancers. *Science* 253, 49-53 (1991)

59. Moeckel G, M. Keil, T. E. Exner, J. Brickmann: Molecular Modeling Information Transfer With VRML: From Small Molecules to Large Systems in Bioscience. In: Proceedings of the Pacific Symposium on Biocomputing '98. Eds.: Altman RB, Dunker AK, Hunter L, and Klein TE, (1998)

60. Cho Y, S. Gorina, P. D. Jeffrey, N. P. Pavletich: Crystal Structure of a P53 Tumor Suppressor-DNA Complex: Understanding Tumorigenic Mutations. *Science* 265, 346-355 (1994)

61. Oellien F, J. Cramer, C. Beyer, W. D. Ihlenfeldt, P. M. Selzer: The Impact of Tautomer Forms on Pharmacophore-Based Virtual Screening. *J Chem Inf Model* 46, 2342-2354 (2006)

62. Brooijmans N, I. D. Kuntz: Molecular Recognition and Docking Algorithms. *Annu Rev Biophys Biomol Struct* 32, 335-373 (2003)

63. Goodsell D S, A. J. Olson: Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins* 8, 195-202 (1990)

64. Meng E C, B. K. Schoichet, I. D. Kuntz: Automated Docking With Grid-Based Energy Evaluation. *J Comput Chem* 13, 505-524 (1992)

65. Rashin A A: Hydration Phenomena, Classical Electrostatics and the Boundary Element Method. *J Phys Chem* 94, 1725-1733 (1990)

66. Shoichet B K, A. R. Leach, I. D. Kuntz: Ligand Solvation in Molecular Docking. *Proteins* 34, 4-16 (1-1-1999)

67. Jones G, P. Willett, R. C. Glen: Molecular Recognition of Receptor Sites Using a Genetic Algorithm With a Description of Desolvation. *J Mol Biol* 245, 43-53 (6-1-1995)

68. Rarey M, B. Kramer, T. Lengauer, G. Klebe: A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *J Mol Biol* 261, 470-489 (23-8-1996)

69. Gohlke H, M. Hendlich, G. Klebe: Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions. *J Mol Biol* 295, 337-356 (2000)

70. Korb O, T. Stützle, T. E. Exner: An Ant Colony Optimization Approach to Flexible Protein–Ligand Docking. Swarm Intell 1, 115-134 (2007)

71. Clark A M, P. Labute: 2D Depiction of Protein-Ligand Complexes. *J Chem Inf Model* 47, 1933-1944 (2007)

72. Wallace A C, R. A. Laskowski, J. M. Thornton: LIGPLOT: A Program to Generate Schematic Diagrams of Protein-Ligand Interactions. *Prot Engineering* 8, 127-134 (1995)

73. Brandstetter H, A. Kuhne, W. Bode, R. Huber, W. van der Sall, k. Wirthensohn, R. A. Engh: X-Ray Structure of Active Site-Inhibited Clotting Factor Xa. Implications for Drug Design and Substrate Recognition. *J Biol Chem* 271, 29988-29992 (1996)

74. McDonald I K, J. M. Thornton: Satisfying Hydrogen Bonding Potential in Proteins. *J Mol Biol* 238, 777-793 (1994)

75. Hubbard S J, J. M. Thornton: NACCESS. *Computer Program, Department of Biochemistry and Molecular Biology, University College London,* (1993)

76. Keil M, T. E. Exner, J. Brickmann: Characterisation of Protein-Ligand Interfaces: Separating Surface. *J Mol Model* 4, 335-339 (1998)

77. Ray N, X. Cavin, J.-C. Paul, B. Maigret: Intersurf: Dynamic Interface Between Proteins. *J Mol Graph Model* 23, 347-354 (2005)

78. Böhm H-J, G. Klebe, H. Kubinyi: Wirkstoffdesign. *Spektrum Akademischer Verlag*, Heidelberg, Berlin, Oxford, (1996)

79. Nicklaus M C: Pharmacophore and Drug Discovery. In: Handbook of Chemoinformatics. Eds.: Gasteiger, J., *Wiley-VCH*, Weinheim, Germany, (2003)

80. Jain A N: Ligand-Based Structural Hypotheses for Virtual Screening. *J Med Chem* 47, 947-961 (2004)

81. Martin Y C: Distance Comparisons (DISCO): A New Strategy for Examining 3D Struture-Activity Relationships. In: Classical and 3D QSAR in Agrochemistry. Eds.: Hansch, C. and Fujita, T., *American Chemical Society*, Washington, DC, (1995)

82. Richmond N J, C. A. Abrams, P. R. N. Wolohan, E. Abrahamian, P. Willett, R. D. Clark: GALAHAD: 1. Pharmacophore Identification by Hypermolecular Alignment of Ligands in 3D. *J Comput-Aided Mol Des* 20, 567-587 (2006)

83. Barnum D, J. Greene, A. Smellie, P. Sprague: Identification of Common Functional Configurations Among Molecules. *J Chem Inf Comput Sci* 36, 563-571 (1996)

84. Horvarth D, B. mao, R. Gozalbes, F. Barbosa, S. L. Rogalski: Strengths and Limitations of Pharmacophore-Based Virtual Screening. In: Methods and Principles in Medicinal Chemistry. *Wiley-VCH*, Weinheim, Germany, (2005)

85. Catalyst 4.11. *Accelrys Software Inc.,* 10188 Telesis Court, Suite 100, San Diego, CA 92121, USA, (2007)

86. Evers A, G. Hessler, H. Matter, T. Klabunde: Virtual Screening of Biogenic Amine-Binding G-Protein Coupled Receptors: Comparative Evaluation of Protein- and Ligand-Based Virtual Screening Protocols. *J Med Chem* 48, 5448-5465 (2005)

87. Wolber G, T. Langer: LigandScout: 3-D Pharmacophores Derived From Protein-Bound Ligands and Their Use As Virtual Screening Filters. *J Chem Inf Model* 45, 160-169 (2005)

88. Bordas B, I. Belai, A. Lopata, Z. Szanto: Interpretation of Scoring Functions Using 3D Molecular Fields . Mapping the Diacyl-Hydrazine-Binding Pocket of an Insect Ecdysone Receptor. *J Chem Inf Model* 47, 176-185 (2007)

89. Cramer R D I, D. E. Patterson, J. D. Bunce: Comparative Molecular Field Analysis (CoMFA): 1. Effect

on Binding of Steroids to Carrier Proteins. *J Am Chem Soc* 110, 5959-5967 (1988)

90. Marshall G R, R. D. I. Cramer: Three-Dimensional Structure-Activity Relationships. *Trends Pharmacol Sci* 9, 285-289 (1988)

91. Clark M, R. D. I. Cramer, D. M. Jones, D. E. Patterson, D. E. Simeroth: Comparative Molecular Field Analysis (CoMFA): 2. Toward Its Use With 3D-Structural Databases. *Tetrahedron Comput Methodol* 3, 47-59 (1990)

92. Gantchev T G, H. Ali, J. E. van Lier: Quantitative Structure-Activity Relationships/Comparative Molecular Field Analysis (QSAR/CoMFA) for Receptor-Binding Properties of Halogenated Estradiol Derivatives. *J Med Chem* 37, 4164-4176 (1994)

93. Debnath A K: Three-Dimensional Quantitative Structure-Activity Relationship Study on Cyclic Urea Derivatives As HIV-1 Protease Inhibitors: Application of Comparative Molecular Field Analysis. *J Med Chem* 42, 249-259 (1999)

94. Corbeil C R, P. Englebienne, N. Moitessier: Docking Ligands into Flexible and Solvated Macromolecules. 1. Development and Validation of FITTED 1.0. *J Chem Inf Model* 47, 435-449 (2007)

95. van Dijk A D J, A. M. J. J. Bonvin: Solvated Docking: Introducing Water into the Modelling of Biomolecular Complexes. *Bioinformatics* 22, 2340-2347 (2006)

96. Friesner R A, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, D. T. Mainz: Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J Med Chem* 49, 6177-6196 (2006)

97. Verdonk M L, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. Nissink, R. D. Taylor, R. Taylor: Modeling Water Molecules in Protein-Ligand Docking Using GOLD. *J Med Chem* 48, 6504-6515 (2005)

98. Catana C, P. F. W. Stouten: Novel, Customizable Scoring Functions, Parameterized Using N-PLS, for Structure-Based Drug Discovery. *J Chem Inf Model* 47, 85-91 (2007)

99. Gohlke H, S. Radestock, B. Breu: Better Knowledge - Better Scoring: Tailoring DrugScore to a Particular Protein Yields Improved Pose and Affinity Predictions and Allows to Incorporate Conformational Variability. In: Abstracts of Papers, 231st ACS National Meeting, Atlanta, GA, United States, March 26-30, 2006. (2006)

100. Zsoldos Z: Ehits: Exhaustive Flexible Ligand Docking With Customizable Scoring Function Tailored to Protein Families. In: Abstracts of Papers, 228th ACS National Meeting, Philadelphia, PA, United States, August 22-26, 2004. (2004)

101. Gohlke H, G. Klebe: DrugScore Meets CoMFA: Adaptation of Fields for Molecular Comparison (AFMoC) or How to Tailor Knowledge-Based Pair-Potentials to a Particular Protein. *J Med Chem* 45, 4153-4170 (2002)

102. Logean A, A. Sette, D. Rognan: Customized Versus Universal Scoring Functions Application to Class I MHC-Peptide Binding Free Energy Predictions. *Bioorg Med Chem Lett* 11, 675-679 (2001)

103. Srinivasan J, T. E. I. Cheatham, P. A. Kollmann, D. A. Case: Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J Am Chem Soc* 120, 9401-9409 (1998)

104. Dancey J, E. A. Sausville: Issues and Progress With Protein Kinase Inhibitors for Cancer Treatment. *Nature Rev Drug Discov* 2, 296-313 (2003)

105. Sheridan R P, M. K. Holloway, G. McGaughey, R. T. Mosley, S. B. Singh: A Simple Method for Visualizing the Differences Between Related Receptor Sites. *J Mol Graph Model* 21, 71-79 (2002)

106. Anderson A, Z. Weng: VRDD: Applying Virtual Reality Visualization to Protein Docking and Design. *J Mol Graph Model* 17, 180-186 (1999)

107. Levine D, M. Facello, P. Hallstrom, G. Reeder, B. Walenz, F. Stevens: Stalk: an Interactive System for Virtual Molecular Docking. *IEEE Computational Science and Engineering* 4, 55-65 (1997)

108. Ouh-young M, M. Pique, J. Hughes, N. Srinivasan, F. P. Jr. Brooks: Using a Manipulator for Force Display in Molecular Docking. *IEEE Robotics and Automation* 3, 1824-1829 (1988)

109. Lewis G N: The Atom and the Molecule. *J Am Chem Soc* 38, 762-785 (1916)

110. Dreiding A S: Simple Molecular Models. *Helv Chim Acta* 42, 1339-1344 (1959)

111. Feldmann R J, C. E. Bing, B. C. Furie, B. Furie: Interactive Computer Surface Graphics Approach to Study of the Active Site of Bovine Trypsin. *Proc Natl Acad Sci USA* 75, 5409-5412 (1978)

112. Richardson J S, D. C. Richardson: The De Novo Design of Protein Structures. *Trends Biochem Sci* 14, 304-309 (1989)

113. Richards F M: Areas, Volumes, Packing and Protein Structure. *Annu Rev Biophys Bioeng* 6, 151-176 (1977)

114. Connolly M L: Solvent Accessible Surfaces of Proteins and Nucleic Acids. *Science* 221, 709-713 (1983)

115. Connolly M L: Analytical Molecular Surface Calculation. *J Appl Cryst* 16, 548-558 (1983)

116. Marhofer R J, K. M. Kast, B. Schilling, H. J. Bar, S. M. Kast, J. Brickmann: Molecular Dynamics Simulations of Tertiary Systems of Cellohexaose/Aliphatic N-Oxide/Water. *Macromol Chem Phys* 201, 2003-2007 (2000)

117. Sayle R A, E. J. Milner-White: RASMOL: Biomolecular Graphics for All. *Trends Biochem Sci* 20, 374(1995)

118. Bohne-Lang A, E. Lang: 3D - Molecule Visualization on the Internet- Based Java Applets. *BIOspektrum* 10, 167-169 (2004)

119. Oellien F, T. Engel, M. C. Hemmer: Chemical Visualization on the WorldWide Web. *Nachrichten aus der Chemie* 48, 1507-1510 (2000)

Abbreviations: 3D: three-dimensional, ADME-Tox: absorption, distribution, metabolism, excretion, and toxicity, ATP: adenosine triphosphate, CoMFA: comparative molecular field analysis, CPK model: Corey-Pauling-Koltun model, GPCR: G-protein-coupled receptor, HTS: high-throughput screening, MM-PBSA: molecular mechanics – Poisson Boltzmann / surface area, NMR: nuclear magnetic resonance, PDB: protein data bank, PMF: potential of mean force, QSAR: quantitative structure-activity relationship, SCR: structural conserved region, SVR: structurally variable region, VRML: virtual reality modeling language

Key Words: Molecular Visualization, Computer Graphics, Rational Drug Design, Molecular Modeling, Homology Modeling, Protein-Ligand Docking, Pharmacophore Modeling, Virtual Screening, Review

Send correspondence to: Thomas E. Exner, Fachbereich Chemie, Universitat Konstanz, 78457 Konstanz, Germany, Tel: 49 (0)7531 882015, Fax: +49 (0)7531 883587, E-mail: thomas.exner@uni-konstanz.de

http://www.bioscience.org/current/vol14.htm