Chapter 7

Summary and Concluding Remarks

Virtual drug design aims at designing new drugs *in silico*. One important step in virtual drug design is the identification of new lead structures that constitute starting points for the development of pharmaceutically active compounds. Large data bases exist that contain millions of known drugs and natural compounds. In order to screen these data bases efficiently, a so called *pharmacophore* is often used. A pharmacophore represents a reduced model carrying the most essential structural as well as physico-chemical features an active compound needs to have. There are different methods to generate a pharmacophore, depending on whether or not the target protein, e.g. an enzyme, is known. In the absence of the target protein, a pharmacophore can be identified by determining the common features of known active compounds, if such features exist. The identification of common features between several compounds requires a multiple alignment of the 3-dimensional structures of these molecules. In contrast to pairwise alignments, multiple alignments are generally based on a much smaller set of common features. Once multiple alignments have been found, conventional methods for pharmacophore elucidation and quantitative structure activity relationship (QSAR) analyses can be applied.

In the context of virtual drug design, where the functional similarity of drug-sized molecules is defined by their ability to bind to the active site of the target protein, the "outer shape" of the molecule is of particular importance. Consequently, in order to identify functional similarities, the alignment algorithm should be based on the molecule's "outer shape", which can be approximated by a specific kind of molecular surface, known as *solvent excluded surface*.

Contributions

In this thesis, we introduced a new framework for multiple surface alignment of drug-sized molecules. At the core of our approach lies a new flexible representation of molecular surface and physico-chemical properties by sets of points. We investigated the applicability of such a point-based representation to molecular surface alignment and made several

contributions, which can be partitioned into three major fields:

- *Distribution of points on triangular meshes.*
- *Pairwise surface alignment.*
- *Multiple alignment.*

Distribution of Points on Triangular Meshes. The central step of our approach is the generation of several point sets on the molecular surface, where each of these sets represents a single molecular property, such as shape, hydrogen bonding, and electrostatic potential. The idea is to distribute points on the solvent excluded surface as regular as possible with respect to a scalar field representing each molecular property. Here, the scalar defines a local point density within each surface region.

To obtain such point distributions, we developed a new approach consisting of two steps. In the first step, an initial point distribution is generated by applying a *multilevel k-way graph-partitioning* algorithm [92], due to which very satisfying initial point distributions are generated. During the second step, the points are iteratively relaxed using *centroidal Voronoi tesselation* [51], which we extended to triangular surface meshes [16]. Using this two-step approach, we obtained high-quality point distributions for constant as well as smoothly varying scalar fields.

The proposed point distribution algorithm is not restricted to molecular surfaces. Hence, it might also be of interest for other applications that require points to be distributed regularly on a triangular surface.

Pairwise Surface Alignment. For the pairwise surface alignment, we use the point-based representation described above. To align two molecular surfaces, we generate a large number of initial alignments, each of which is locally optimized using an efficient *point matching scheme* [97]. We extended this point matching scheme to meet the demands imposed by surface points [16] and developed a data structure speeding up the computation of point matchings by a factor of three.

Cosgrove et al. [35] and Hofbauer [81] also used surface points to compute molecular surface alignments. However, our approach differs in at least three respects. First, we use considerably more points, which allows to represent the molecular shape much better. Second, we represent each molecular property by a separate point set, while in their approaches, a single point combines several properties. This simplifies the computation of point matchings and allows an easier visual inspection of the results. Finally, while Cosgrove et al. and Hofbauer use clique detection to align the point sets, we only use clique detection to generate initial alignments. A comparison of the results obtained for a set of eight thermolysin inhibitors revealed, that our approach gives better results while being faster.

We also showed, that our algorithm can be used to align conformational densities representing metastable conformations. This is accomplished by aligning isodensity surfaces generated from the conformational densities.

Multiple Alignment. The discrete representation of molecular shape and properties using points allows a straight forward computation of multiple alignments from pairwise alignments by successive intersection of the pairwise matchings underlying the pairwise alignments. However, this straight forward approach is challenged by two problems. First, due to the large number of surface points, the number of multiple alignments might become very large. And second, due to the local optimization scheme used for pairwise alignment, several very similar pairwise alignments are generated, which entail a large number of very similar multiple matchings.

The former challenge is met by introducing size restrictions on the one hand, and by using an efficient data structure, called PATRICIA tree [128], on the other hand.

The latter problem is solved by removing similar pairwise alignments. This is accomplished by analyzing the *root mean square* (rms) distance of the transformed point sets in 3-dimensional space. We proved that it suffices to look at the translational part of the rigid body transformation to obtain a lower bound to the rms distance, which renders the determination of the true rms distance between most pairwise alignments unnecessary.

Concluding Remarks

We evaluated the quality of our proposed methods for pairwise and multiple surface alignment using two sets of molecules. The first set consisted of eight *thermolysin inhibitors*. The second set consisted of seven *HIV-1 protease inhibitors*. For both sets, the experimental complex structures were available, which allowed a quantitative evaluation of the obtained alignments.

Besides aligning the experimental conformers of the two sets of molecules, we also computed ensembles of conformers using conformation analysis tools. These ensembles were used to investigate, whether our surface alignment algorithm is able to identify active conformers. We compared these results with a previously published approach based on point matching using the molecules' atoms as points [15], i.e. atom alignment. This comparison revealed that the surface alignment algorithm performed much better than the atom alignment algorithm, in particular when comparing molecules with dissimilar scaffolds.

Multiple alignments are generally based on a rather small set of features common to all molecules participating in the multiple alignment. However, this small set of features often carries more information than common features obtained from pairwise alignment. We demonstrated, that by visual inspection of the multiple matchings, the user can already gain much insight into the system of active compounds. If more quantitative insight is needed, the generated multiple alignments may be used as input to a quantitative structure activity relationship (QSAR) analysis.

Finally, from the high quality of the results we can conclude that the proposed surface alignment algorithm represents a versatile tool for virtual drug design.