4. Computations and Results

The primary aim of this project was the investigation of surfaces of small molecules. The method described in chapter 3 has been explicitly designed for that purpose. As a proof of concept 8 thermolysin inhibitors were investigated that were subject to an earlier surface similarity search performed by Cosgrove et. al. [37] with their SPAt program. This dataset was also used to validate the scoring algorithm against the ranking of a flexible alignment [82]. In addition another set of structures was assembled, which contains known active ligands of dihydrofolate reductase, to test the performance of different kinds of molecular surfaces. The effects of conformational flexibility were also tested with different conformations of ATP^{4-} and of a dihydrofolate reductase inhibitor.

During the project it was possible to apply the program to protein/protein and in particular to ligand binding site comparisons. With only little adjustments SURFCOMP achieved successful and illustrative alignments between the active site surfaces of different SH2-domains and phosphatases. These alignments helped elucidating important aspects in the differences and similarities between the binding sites of these proteins.

In the following sections the experimental details and results of the aforementioned experiments are presented. Unless otherwise noted the experiments were performed according to the methodologies described in the previous chapters.

4.1. Ligand Surfaces

4.1.1. Preparation of the Input Structures and Experimental Design

All the molecular surfaces investigated for the experiments described in this section were calculated from 3D atomic data. The 3D structures were extracted from crystallographic data of protein/ligand complexes available in the Brookhaven Protein Data Bank (PDB) [13]. To compare the overlays generated by the present method with the experimental alignments of the different ligands in the proteins' active sites, the complexes in the PDB were superimposed by the backbone atoms of corresponding amino acids in the binding sites, which was always possible with a very small RMS deviation. The structures of the ligands were extracted and hydrogen atoms were added with Sybyl 6.9 [2].

For each structure the solvent excluded surfaces (section 2.1.4) were computed. Electrostatic potential based on semi-empirical calculations (section 3.12), lipophilic potential (section 2.2.1) and two sets of canonical curvatures together with shape type indices for a cutoff range of 1.0 and 2.0 Å (section 2.2.3) were mapped onto the molecular surfaces. For proteins the cutoff ranges were 2.0, 4.0 and 6.0 Å.

4.1.2. SURFCOMP Validation: Comparison of 8 Thermolysin Inhibitors

Thermolysin (TLN, EC-number 3.4.24.27) is a thermostable extracellular metalloendopeptidase containing four calcium ions from *Bacillus thermoproteolyticus.* [70]. The active site of the enzyme (see Figure 4-1) consists of two subsites: a zinc ion complexed by two histidine residues and one glutamic acid representing the catalytic reaction center, and a hydrophobic cleft, formed between two α-helices, that contains the selective part of the site. The crystals of thermolysin contain a lysine-valine dipeptide in this pocket that seems to be the product of the cleavage of the C-terminus of another thermolysin molecule.

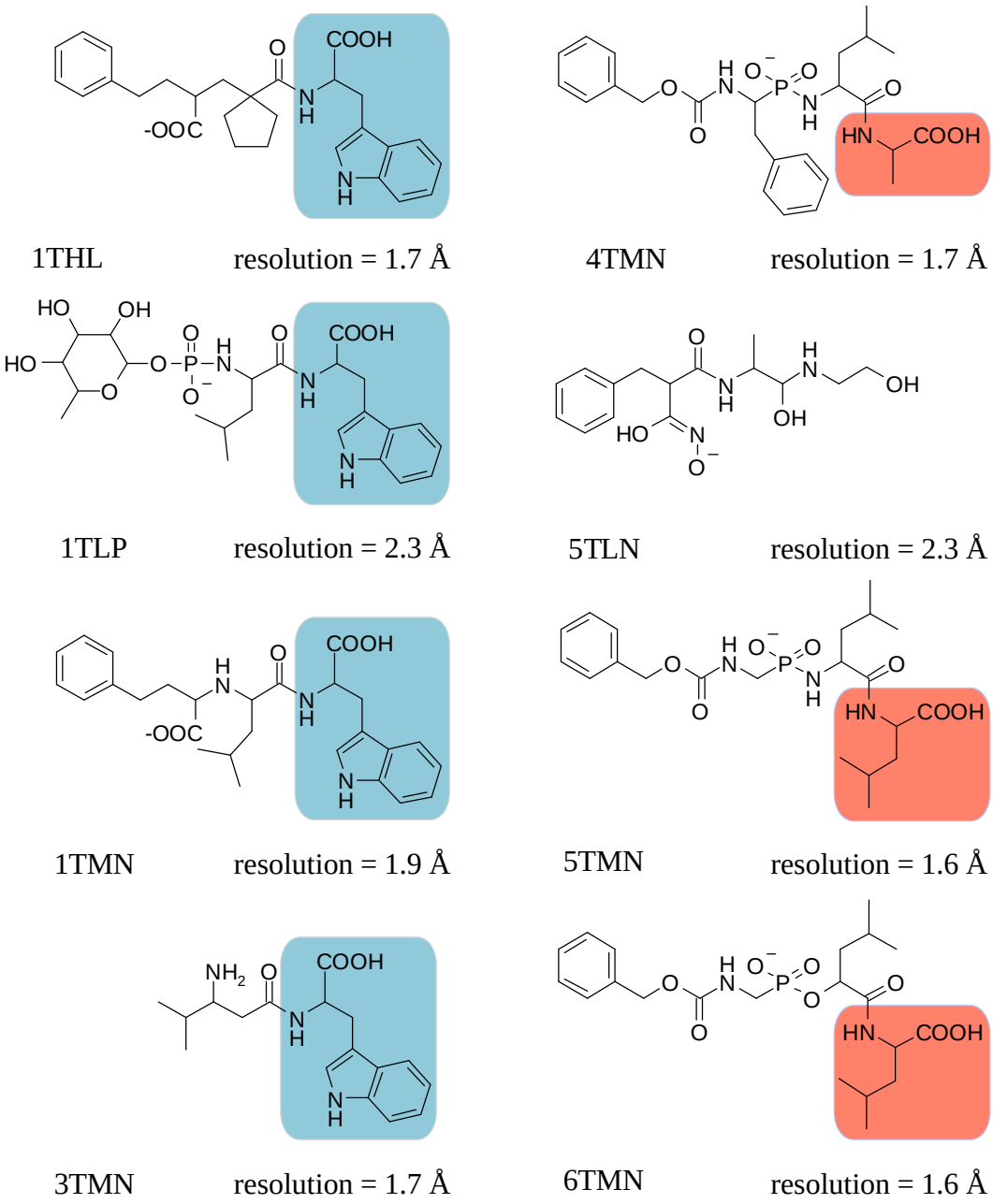


Chart 4-1: 2D structures of eight thermolysin inhibitors:
The structures are identified by the PDB entry name of the corresponding protein/ligand complex. The given resolution is for the complete protein/ligand complex in the X-ray data.

Several structures of TLN cocrystalized with different inhibitor compounds are available from the PDB. Cosgrove et. al. used a subset of 8 inhibitor structures to demonstrate the abilities of their molecular comparison software SPAt [37]. The same set was used to perform an exhaustive pairwise similarity search between the molecular surfaces and the results were compared with the results of the aforementioned publication to validate the program SURFCOMP.

The structures of the eight thermolysin inhibitors in Chart 4-1 were extracted from the PDB. All molecules except 3TMN and 5TLN are complexed via a negatively charged carboxyl- or phosphate-like group to the zinc ion in the active site of the protein. Thus a single negative formal charge was placed at these positions. 5TLN is also complexed to

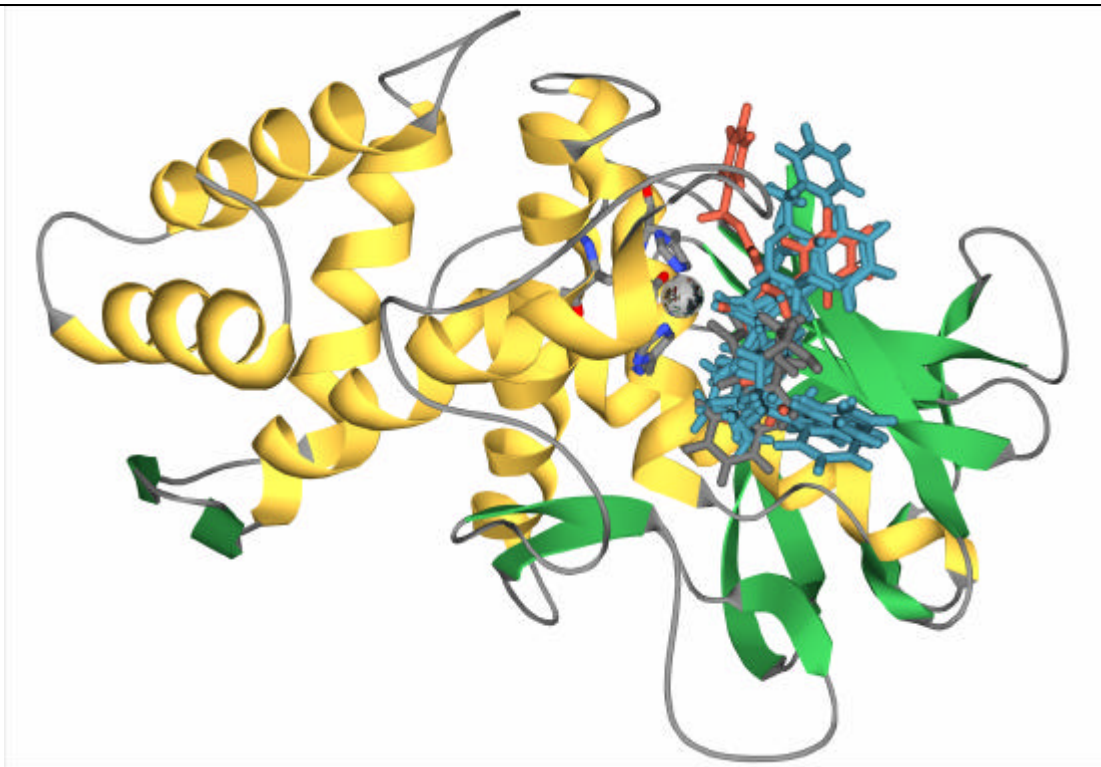


Figure 4-1: The 3D structure of thermolysin (TLN).
The 8 ligand structures of the set are superimposed in the active site. Ligands of the tryptophan class are colored in blue while the structures belonging to the valine and alanine class are shown in red. 5TLN, which does not belong to any class, is left grey. The metallic sphere represents the position of the complexed Zn ion.

the zinc ion but via a charged hydroxamic acid group. 3TMN does not show any complex binding to the ion at all and was left uncharged.

Two different experiments were performed, one with the electrostatic and one with the lipophilic potential mapped onto the molecular surfaces. The experimental details can be found in Table 4-1. Using the ESP the program could find good overlays for all structures, except for 5TLN, which is quite different in shape, especially in the most interesting region around the complex-building part. The rest of the molecules can be divided into two classes: structures with tryptophan (blue boxes in Chart 4-1) and structures with an aliphatic (alanine, leucine; red boxes in Chart 4-1) residue at the C-

filter parameter	**symbol**	**section**[a]	**value**	**property**[b]
curvature cut-off range	c_{CR}	2.2.3	2.0 Å	
neighbourhood radius	r_{CP}	3.2	2.0 Å	
fuzzy threshold	F	3.5	0.3	ESP or LP
shape threshold	R	3.6	0.6	STI
distance tolerance	T	3.7	1.0 Å	
minimum distance	d_{min}	3.7	0.5 Å	
angular tolerance	f_{tol}	3.8	15.0 °	

Table 4-1: Experimental conditions used in the thermolysin experiments.
[a)]the section in the text where the filter is described
[b)]the molecular surface property applied to the specific filter (ESP, electrostatic potential and LP, lipophilic potential).

terminal end.

The tryptophan structures could be overlaid with an RMS deviation between the experimental and calculated alignment of less than 0.6 Å. The only exception is 3TMN aligned to 1TMN which shows a slightly worse RMSD of 1.0 Å mainly due to differences in their electrostatic potential and to a different angle between the indole ring and the peptide backbone. The three structures with aliphatic residues show comparable, good overlays with RMSD all below 0.6 Å. A special case is the comparison of 5TMN and 6TMN because the molecules are almost the same except for one group. Consequently their shapes and electrostatic potential are also very similar which is

Molecules		ESP RMSD[a] [Å]			LP RMSD[a] [Å]			Molecules		ESP RMSD[a] [Å]			LP RMSD[a] [Å]		
A	B	points[b]	surf.	struct.	points[b]	surf.	struct.	A	B	points[b]	surf.	struct.	points[b]	surf.	struct.
1THL	1TLP	580	1.95	0.40	441	1.60	1.04	1TMN	4TMN	446	1.05	1.03	417	1.44	0.80
	1TMN	711	1.77	0.40	554	1.55	0.31		5TLN	145	0.99	5.14	205	1.61	6.25
	3TMN	366	1.04	0.33	368	1.12	0.55		5TMN	464	1.21	0.93	222	0.71	0.84
	4TMN	431	1.07	1.18	349	0.98	0.95		6TMN	610	1.26	0.99	426	1.49	0.86
	5TLN	227	1.93	5.68	181	1.78	5.11	3TMN	4TMN	255	1.36	1.42	339	2.07	5.45
	5TMN	336	1.04	1.20	169	0.98	7.08		5TLN	252	1.99	2.90	116	0.58	6.91
	6TMN	439	1.00	0.63	228	0.89	0.73		5TMN	254	1.18	1.51	363	1.68	1.18
1TLP	1TMN	630	1.73	0.53	309	1.35	1.39		6TMN	180	1.26	4.28	283	1.39	0.67
	3TMN	471	1.26	0.46	424	1.52	1.20	4TMN	5TLN	383	3.52	5.83	169	1.17	6.22
	4TMN	446	2.16	1.29	188	1.51	6.01		5TMN	320	0.75	0.43	168	0.52	0.54
	5TLN	342	2.13	7.00	335	2.50	1.27		6TMN	409	0.83	0.58	312	1.90	0.49
	5TMN	454	0.93	0.63	165	0.63	1.22	5TLN	5TMN	175	1.34	2.31	176	1.44	3.37
	6TMN	409	1.12	0.59	282	0.79	1.04		6TMN	153	1.77	5.78	188	1.55	1.18
1TMN	3TMN	193	0.93	1.00	393	2.50	0.75	5TMN	6TMN	975	0.51	0.08	965	0.55	0.05

Table 4-2: Surface overlays of different thermolysin inhibitors
The surface comparisons were performed with electrostatic potential (ESP) and lipophilic potential (LP)
[a]root mean square deviation
[b]specifies the number of all surface points in the patches that were used to calculate the surface alignment. This number indicates the size of the similar surface region (higher number: larger region).

reflected by the small RMS deviation of 0.05 Å and the nearly one-to-one match of the surfaces.

As expected, the overlays between the two classes were not as good as the within-class results but the general orientation and the important similar surface regions were detected correctly with RMSD values around 1.0 Å. The only exception is again 3TMN which shows rather poor alignments with the structures of the second group. This is due to the different total charge which shifts the ESP values and to the fact that 3TMN does not have the complexing group and the latter do not have the indole ring system.

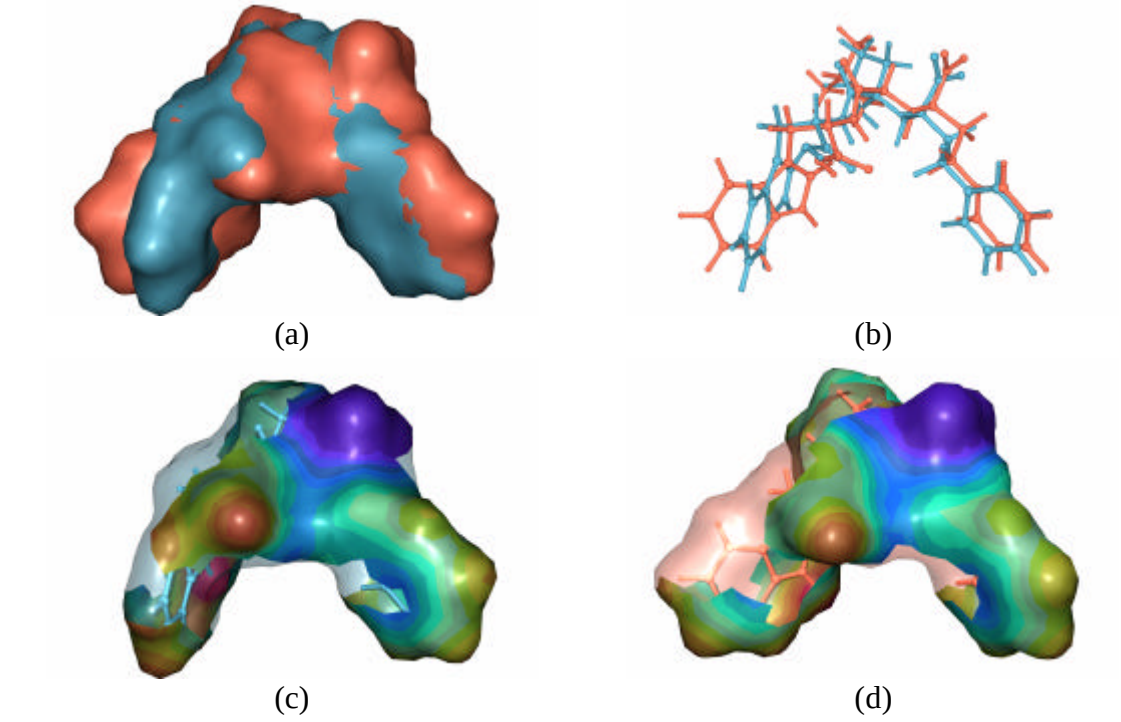


Figure 4-3: Surface alignment of 1THL (blue) and 1TMN (red).
(a) and (b) display the alignment of the molecular surfaces and structures respectively based on the detected surface similarity. (c) and (d) show the similar surface regions of 1THL and 1TMN color coded by the electrostatic potential to illustrate their size and physicochemical similarity.

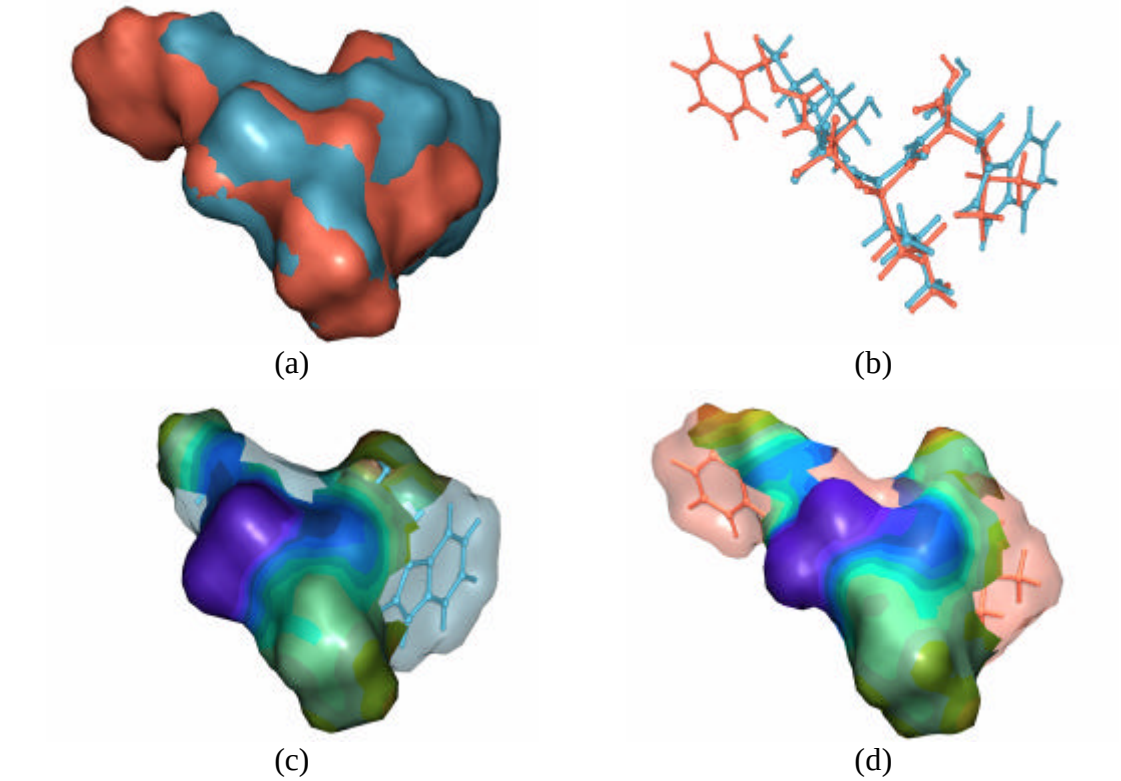


Figure 4-2: Surface alignment of 1TLP (blue) and 6TMN (red).
(a) and (b) display the alignment of the molecular surfaces and structures respectively based on the detected surface similarity. (c) and (d) show the similar surface regions of 1TLP and 6TMN color coded by the electrostatic potential to illustrate their size and physicochemical similarity.

The overlays found by the surface matching conducted with the lipophilic potential as the chemical filter were in general not as good as the results obtained with ESP. The main reason is that regions of the molecules that are quite close to each other in the active site, like the fructose residue of 1TLP and the phenyl ring of 1THL or 1TMN, show different lipophilicities. However the fact that the LP overlays of 3TMN on 1TMN, 5TMN and 6TMN are significantly better than the ESP overlays is due to the strong hydrophobic similarity between the alanine, tryptophan and leucine side chains. The results of both experiments are presented in Table 4-2 and example alignments are displayed in Figures 4-2 and 4-3. These results together with a description of the method have been published [65].

Besides the structure alignment based on the surfaces, the SURFCOMP program also provides a detailed picture of the surface similarities that were found between the molecules. If the results of the comparison between 1THL and the rest of the dataset are lined up (excluding 5TLN which does not show any reasonable surface similarity), one can see that the similar surface regions contain some recurring patterns (see Figure 4-4 and Figure 4-3 for 1TMN). The most common motif between them seems to be the

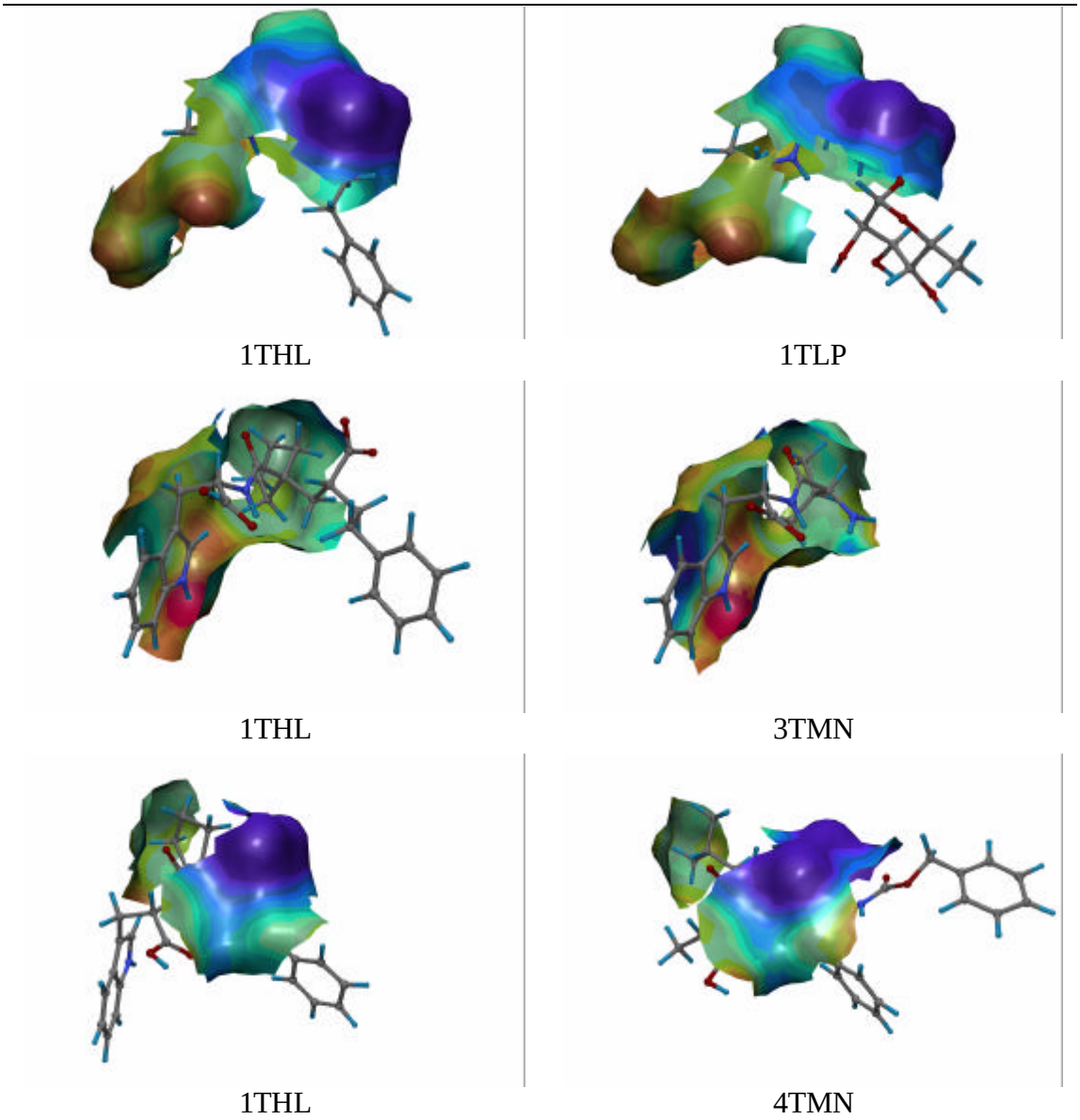


Figure 4-4: continued on p. 46

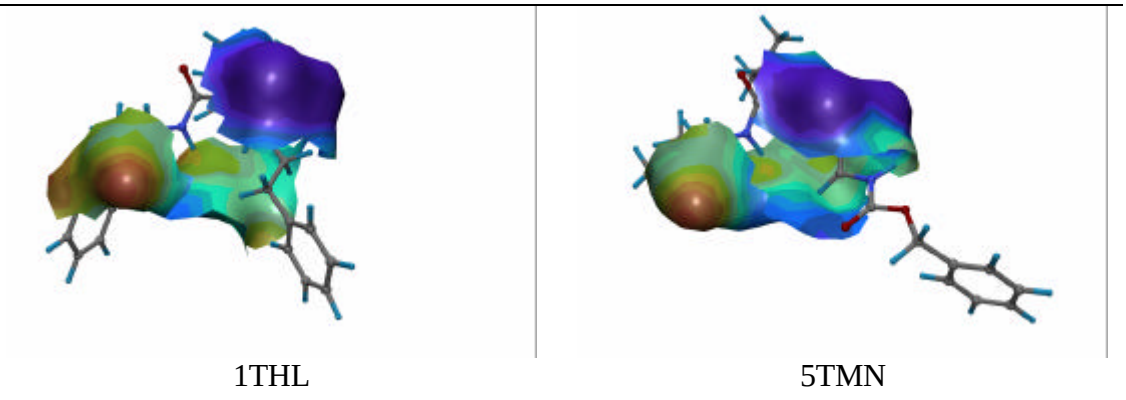


Figure 4-4: Similar surface regions between 1THL (left) and other molecules (right).
The similar patches are color-coded by their electrostatic potential, where blue represents negative and red positive patches.

negatively charged surface region around the phosphate or carboxyl groups in the center of the molecules. The only exception is 3TMN, where that group is not present. In 1TMN, 4TMN and 6TMN the valley between this group and the C-terminal carboxylic acid is included, while in 1TMN, 1TLP and 5TMN the terminal carboxylic group itself is part of the similarity region. A strong similarity is also detected between the indole ring systems of 1THL, 1TLP, 1TMN and 3TMN where the center of that pattern is located around the nitrogen atom. Another interesting, but rather small pattern can be identified around the aliphatic sidechains upstream of the C-terminal end of the molecules. It was detected in all comparisons except for 5TMN.

The results agree with the alignments published earlier by Cosgrove et al [37] for the same dataset. The result of their SPAt program is an overlap graph, an acyclic graph that describes the best way to produce a consensus overlay between the surfaces of the dataset. In the case of the thermolysin structures, this graph consists of two connected components, which can be considered as some kind of arbitrary classification, although this is not the intention of the SPAt software. The dataset is divided into one large group containing 1THL, 1TLP, 1TMN, 3TMN, 4TMN and 5TLN and a smaller group that consists of 5TMN and 6TMN. The edges of the graph are weighted by the fraction of points of one surface that are placed within 1.0 Å of any point of the other surface by the given alignment. This evaluation of the surface similarities is different to the one used in the present experiment, because it is sensitive to differences in the size of the two compared molecules, while SURFCOMP considers only the RMSD of the similar patches. However, if the results are compared with all the data published for the SPAt calculation of the thermolysin dataset, it can be demonstrated that SURFCOMP produces comparable alignments and performs better if the size of the similar surface patches is small compared to the rest of the surfaces.

4.1.3. Ranking of Surface Alignments

If one takes a look at the number of possible alignments that are found by the SURFCOMP program in the thermolysin example (Table 4-3, below), it is obvious that a fast evaluation of the results is necessary to process a large set of surface similarity searches. In the case of the thermolysin data set the RMSD between the alignments found by the SURFCOMP program and the actual positions in the X-ray data provides a good basis for the selection of the best surface similarities. This is possible because all the structures are complexed to the same protein conformation. If, as described above, the different crystal structures are superimposed according to the backbone atoms of the thermolysin protein, the ligand molecules are brought into a natural alignment. Any superposition of two molecular structures that is based on their surface similarity can be compared to that alignment.

However, one does not always have this opportunity, especially if the 3D structures are taken from different contexts. In section 3.10 a consensus scoring scheme is described that was designed to enable a fast filtering of a native SURFCOMP result list. It produces a scoring based on the goodness of the shape fit, the size of the surface similarity and the chemical correlation, and it should be able to distinguish between promising and poor surface similarity clusters. Two different kinds of ranking are of particular interest:

(1) the identification of promising clusters within a single surface comparison and

(2) the ranking of a set of surfaces based on their similarity to a template surface.

In the first case the similarity, which reproduces the natural alignment best, should be close to or at the top of the ranking. This would guarantee that only a few clusters need to be inspected manually to find the optimal solution. The latter ranking type, also known as comparative scoring, must ensure that among the combined clusters of different experiments those surfaces are scored best that show the closest similarity to the template.

Identification of Promising Clusters. To find out, if the comparative scoring scheme is appropriate or not, the rankings produced for the similarity searches of the thermolysin data set were investigated. In Table 4-4 the ranks of the alignments which are closest to the experimental situations (I, closest cluster) compared with the ranks of the best scored clusters when ordered by the RMSD to the natural alignments (II, top cluster). The detailed results of this investigation revealed that the difference in the RMSD between the top clusters and the closest clusters were small for almost all cases where a reasonable similarity between the two surfaces exists. If no similarities could be established, the difference in the rankings and RMSD became larger. This was the case in the comparisons of 5TLN with all the members of the set, because of the totally divergent shape of its surface, and in some of the pairwise comparisons of 3TMN, particularly with 6TMN, where no satisfying similarity between the surfaces could be established.

In two cases the top cluster was the closest cluster (1TMN-3TMN, and 3TMN-4TMN). It is interesting to point out that these comparisons detected only weak or small alignments with an RMSD to the X-ray data above 1 Å. It is possible, that there are only a few acceptable alignments in such situations and the closest cluster does not face much competition from other candidate clusters. For instance the program produced only 44 different clusters when comparing 1TMN and 3TMN which is the third-lowest number among the calculations of the thermolysin dataset.

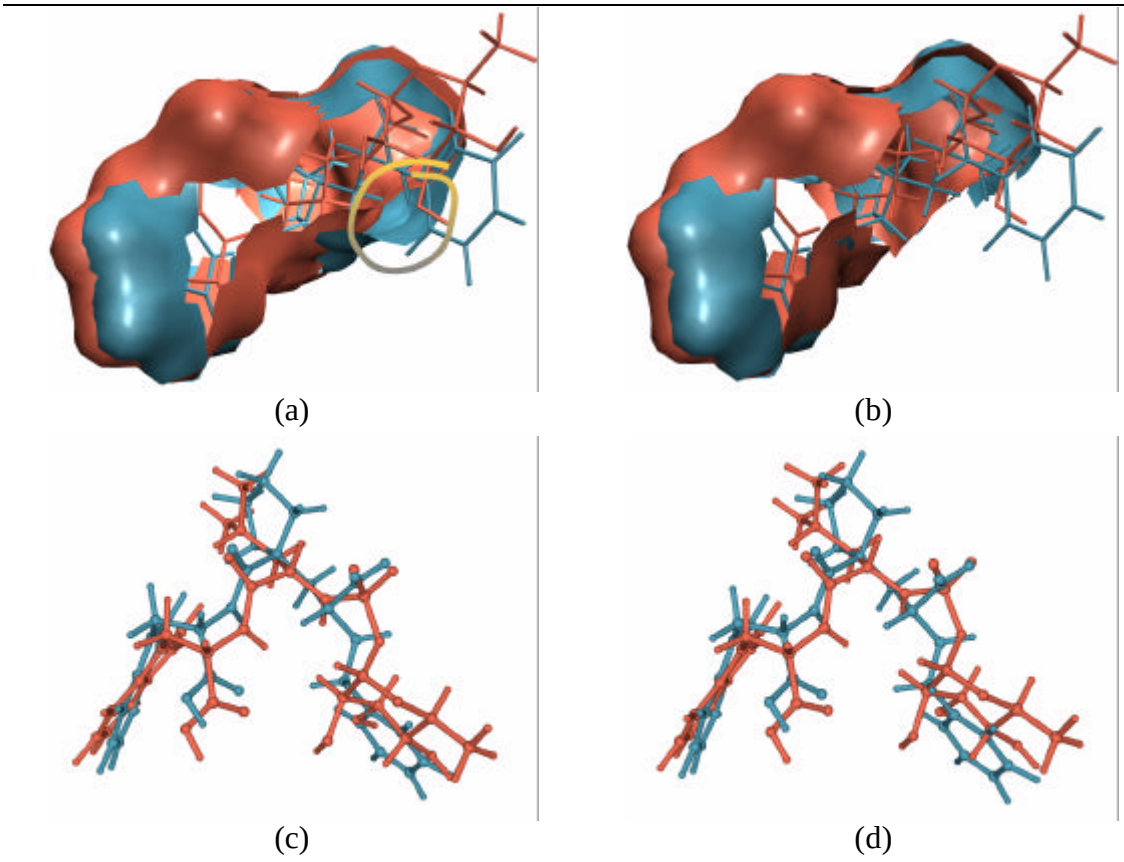


Figure 4-5: Comparison between the top and closest clusters for 1THL and 1TLP.
The left column shows the closest and the right the top cluster of 1THL (blue) and 1TLP (red). (a) and (b) line up the actual similar surface patches to focus on the difference between them and (c) and (d) give a snapshot of the atomic superposition viewed from the top of the molecules.

The bad ranking of the closest cluster in the comparison between 1THL and 1TLP also deserves attention. Although the difference between the top and the closest cluster is within the range of many other pairs (0.18 Å), it was ranked only at position 42 by the consensus scoring method. The main reason for that is a rather bad alignment between the corresponding surface points which is expressed by the high RMSD value of 1.94 Å. The cause of this bad alignment is a single patch pair between both molecules that could only be superimposed with a relatively large gap (see also emphasized region in Figure 4-5). The top cluster, however, does not include this patch pair and the superposition between its corresponding points is much better, although it is only the fifth best reproduction of the natural alignment (but with a very small difference). Further investigation reveals that the top cluster is a subset of the closest cluster except for the single bad matching patch pair.

Molecule				Molecule			
A	**B**	**top-level**	**total**	**A**	**B**	**top-level**	**total**
1THL	1TLP	19	205	1TMN	4TMN	18	132
	1TMN	15	101		5TLN	9	48
	3TMN	11	71		5TMN	14	90
	4TMN	16	131		6TMN	14	106
	5TLN	13	83	3TMN	4TMN	12	94
	5TMN	10	60		5TLN	10	64
	6TMN	10	76		5TMN	7	39
1TLP	1TMN	17	137		6TMN	9	35
	3TMN	14	92	4TMN	5TLN	18	126
	4TMN	25	201		5TMN	17	151
	5TLN	15	99		6TMN	18	160
	5TMN	16	138	5TLN	5TMN	18	52
	6TMN	22	162		6TMN	17	64
1TMN	3TMN	10	44	5TMN	6TMN	18	190

Table 4-3: The number of top level and total alignments
These clusters were found by the SURFCOMP program during the surface similarity searches in the thermolysin dataset.

Molecule					Molecule				
A	**B**	**I**	**II**	D**RMSD**	**A**	**B**	**I**	**II**	D**RMSD**
1THL	1TLP	42	5	0.18	1TMN	4TMN	2	16	0.76
	1TMN	5	6	0.14		5TLN	9	46	6.60
	3TMN	4	2	0.04		5TMN	8	3	0.36
	4TMN	16	22	1.21		6TMN	10	7	0.88
	5TLN	84	80	4.75	3TMN	4TMN	1	1	0.00
	5TMN	10	2	0.30		5TLN	61	24	4.29
	6TMN	8	4	0.12		5TMN	3	4	0.43
1TLP	1TMN	7	10	0.21		6TMN	21	3	1.13
	3TMN	8	5	0.15	4TMN	5TLN	46	62	3.93
	4TMN	11	6	0.52		5TMN	4	4	0.20
	5TLN	80	75	3.63		6TMN	5	3	0.20
	5TMN	7	7	0.15	5TLN	5TMN	5	19	6.16
	6TMN	4	6	0.07		6TMN	6	33	5.94
1TMN	3TMN	1	1	0.00	5TMN	6TMN	15	3	0.00

Table 4-4: Comparison of closest and top clusters.
A comparison between the ranks of the clusters that are closest to the natural alignment sorted by consensus scoring method (I) and the best scoring clusters when sorted by the RMSD to the natural alignment based on the binding site (II). In addition, the ΔRMSD between the two clusters is shown in the third column.

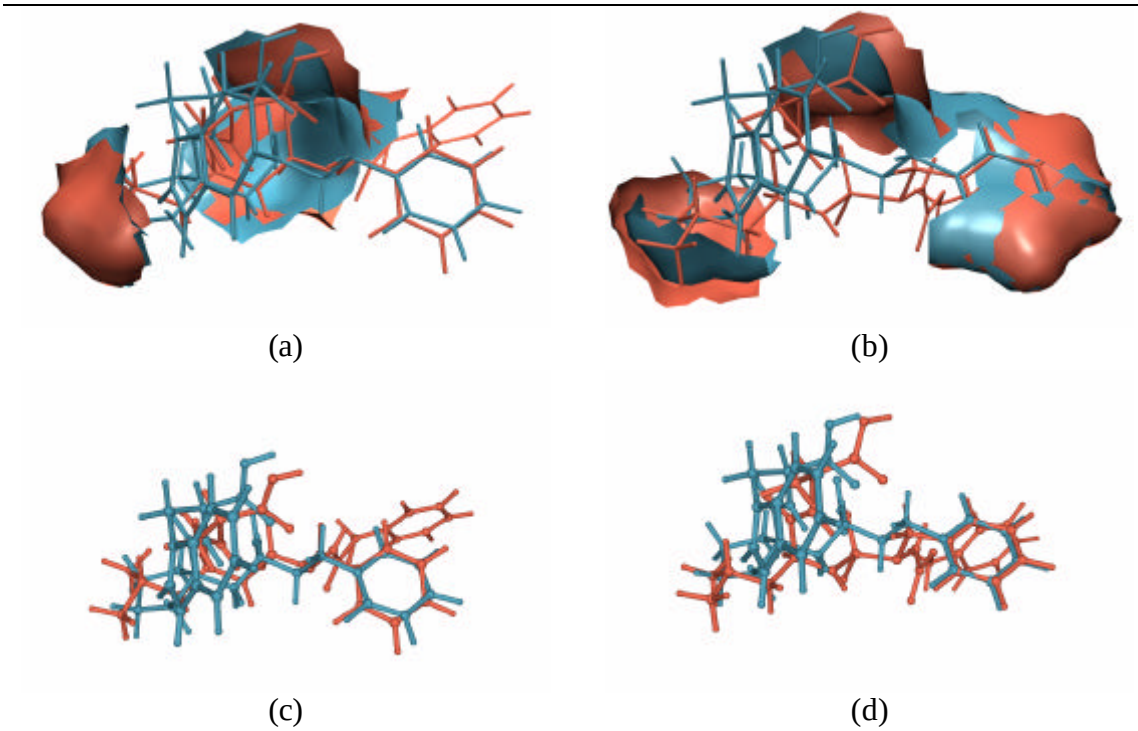


Figure 4-6: Comparison between the top and closest clusters of 1THL and 4TMN.
The left column shows the closest and the right the top cluster of 1THL (blue) and 4TMN (red). (a) and (b) line up the actual similar surface patches to focus on the difference between them and (c) and (d) give a snapshot of the atomic superposition viewed from the top of the molecules.

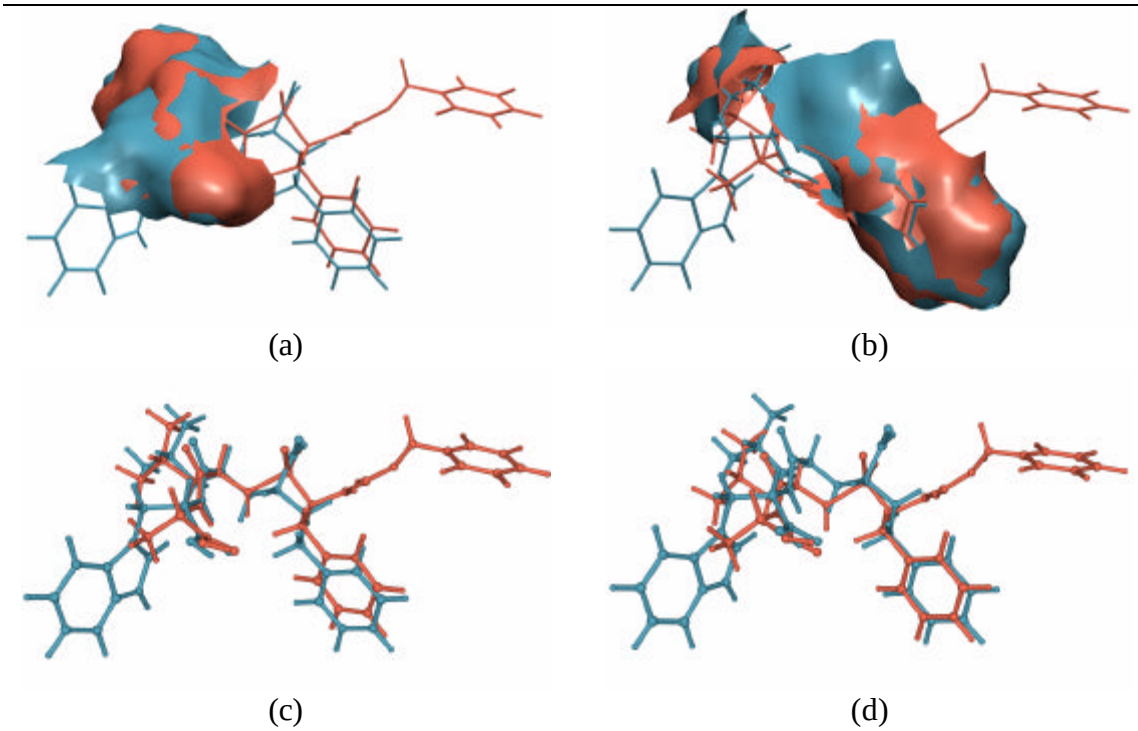


Figure 4-7: Comparison between the top and closest clusters of 1TMN and 4TMN.
The left column shows the closest and the right the top cluster of 1TMN (blue) and 4TMN (red). (a) and (b) line up the actual similar surface patches to focus on the difference between them and (c) and (d) give a snapshot of the atomic superposition viewed from the top of the molecules.

Two calculations resulted in a very large ΔRMSD between the top and the closest cluster. These were the comparisons between 1THL and 4TMN as well as 1TMN and 4TMN. All three molecules have a phenyl ring attached via a two atom bridge to the core of the structure. This resulted in a very similar surface region at that end of the molecules, which was detected by the program and used to create the atomic superpositions. Unfortunately these rings are not aligned in the active site of the proteins, which causes a displacement that is emphasized even more by the fact that the rings are placed at the perimeter of the molecules. The superpositions of the X-ray structures, the top and the closest clusters are shown in Figure 4-6 and Figure 4-7.

Finally it should be mentioned that the ΔRMSD between the top and closest clusters as well as the RMSD values between the natural alignment and the closest clusters are well below the resolution of the X-ray structures for those cases where surface similarity could be established.

Comparative Scoring. The second scoring task is the identification or ranking of the most similar surfaces compared to a template. This is similar to the evaluation of docking results, where the docked conformations of the ligand dataset are ranked to identify the most promising compounds. To accomplish this not only the alternative clusters of a single surface comparison, but the complete results of all comparisons between a set of surfaces and the given template surface are ranked by the scoring algorithm. The single surfaces of the dataset are finally sorted according to their best ranked cluster in that evaluation.

The proof of concept for that procedure can be provided by any other technique that can rank different surfaces according to their similarity to a specific template. Unfortunately, to the author's best knowledge, there is no method available that can rank molecules according to their surface similarity. Therefore the software FlexS [82] was used to validate the comparative scoring scheme of SURFCOMP, because it uses a volumetric technique to generate good flexible alignments between different molecules.

FlexS is closely related to the flexible docking program FlexX [112]. It uses various forms of possible intermolecular interactions as well as different property distributions (such as partial atomic charges or the H-bonding potential) to generate a flexible superposition between a template and a query molecule. Interaction centers and geometries or pairwise intermolecular interactions are used to evaluate the coincidence of H-bonds, salt bridges or lipophilic interactions in an alignment between the two structures. To check the similarity between certain molecular properties, Gaussian

filter parameter	**symbol**	**section**[a]	**value**	**property**[b]
curvature cut-off range	c_{CR}	2.2.3	2.0 Å	
neighbourhood radius	r_{CP}	3.2	2.0 Å	
fuzzy threshold	F	3.5	0.4	ESP
shape threshold	R	3.6	0.5	STI
distance tolerance	T	3.7	1.0 Å	
minimum distance	d_{min}	3.7	0.5 Å	
angular tolerance	f_{tol}	3.8	15.0 °	

Table 4-5: Experimental conditions used in the comparative ranking experiments.
[a)] the section in the text where the filter is described
[b)] the molecular surface property applied to the specific filter (ESP, electrostatic potential).

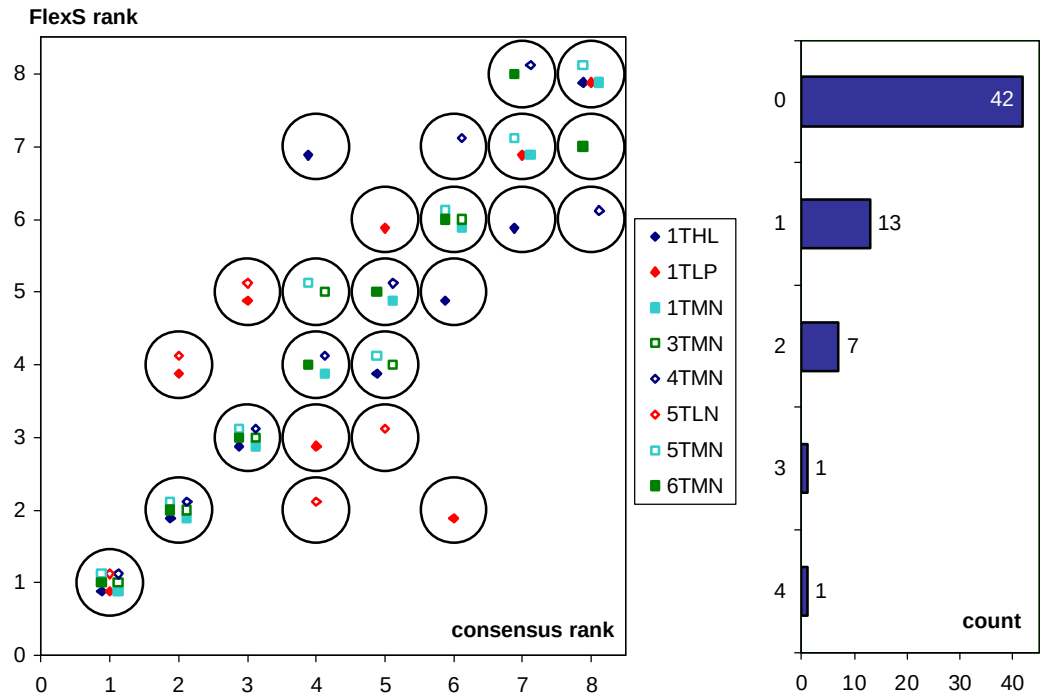


Figure 4-8: Results of the comparative ranking
(left) Mapping between the SURFCOMP consensus ranking and the FlexS ranks of all comparative ranking experiments. Each circle represents a distinct mapping between the two rankings that occurs at least once in the calculations. All correct matches appear in the diagonal of the graph. (right) A histogram of the mismatches (0 indicates a correct match).

functions that model the respective densities are used.

The experiment was designed as follows: For each structure in the thermolysin dataset, a flexible alignment with all the other structures in the set was generated. The conformations that produced the best alignment with the current template structure were taken to form the data for the surface similarity searches. Solvent excluded surfaces were generated for all structures in that set and compared to the surface of the template molecule with SURFCOMP. The resulting tables of alternative clusters were combined into one table for each template molecule and ranked by the consensus scoring approach. From this cluster scoring a ranking of the molecules of the data set was assembled based on the first occurrence of the best cluster of each molecule. The parameters for these experiments are summarized in Table 4-5.

In Figure 4-8 the results of all 8 comparative ranking experiments are summarized and the details are given in Table 4-6. Overall the agreement between the ranking based on FlexS' total score and the consensus scoring of the SURFCOMP program is very good. More than 65% of the structures were assigned the same rank by both methods and another 20% showed only a ranking difference of 1. Furthermore, many of the mismatches are still in a correct relative order. For example, the flexible superposition against 1THL ranks the molecule 3TMN at the next to last position, because it can only cover a part of the template molecule. The surface similarity ranking, however, does not take that into account, because it considers only the local similarities and does not consider the fraction of the covered template surface. Therefore 3TMN is ranked much higher by SURFCOMP because the absolute size of the similar patches is comparable to other similar molecules like 1TLP and 1TMN. It should also be mentioned that the agreement between the two scoring methods is in general better for the high and low

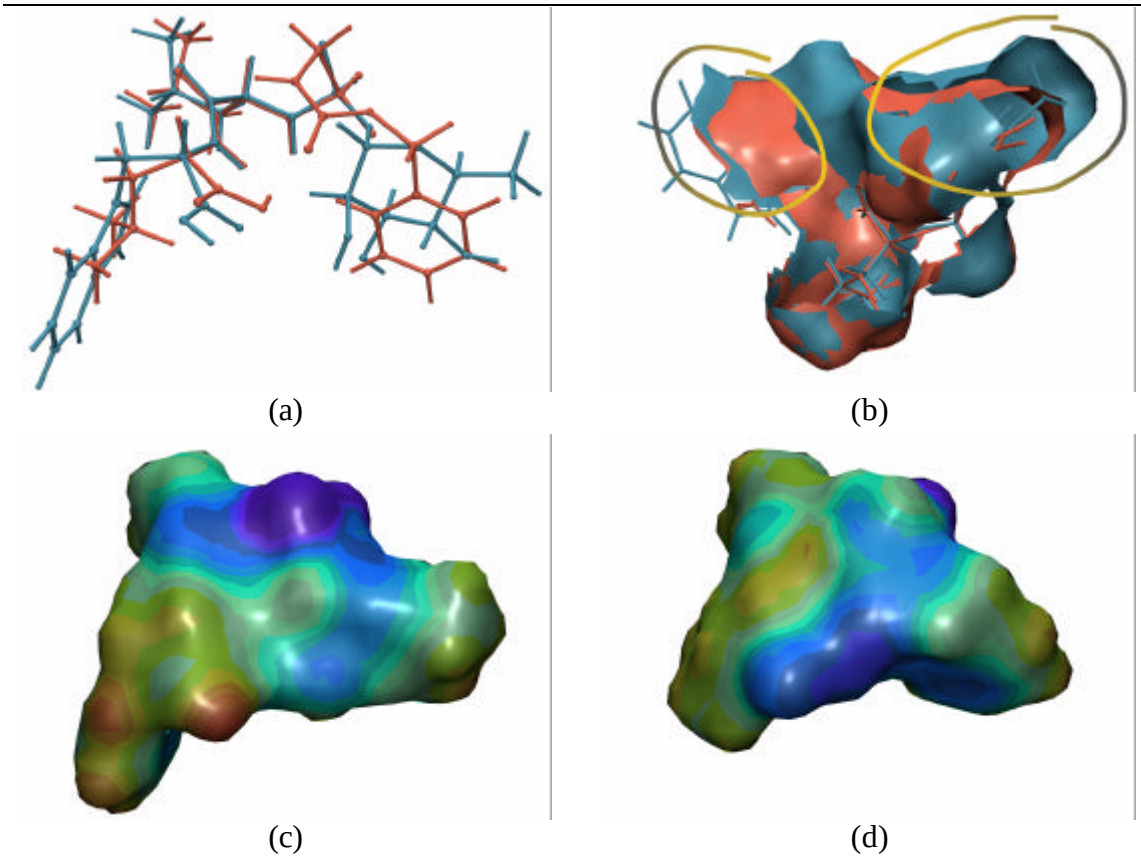


Figure 4-9: The superposition of 1TLP (blue) and 5TMN (red).
(a) The FlexS program aligns the C-terminal residues as well as the fructose and phenyl residues respectively. The SURFCOMP cluster that covers most of the surface similarities (b) has a very large surface RMSD although it represents the original superposition best. In general the two surfaces look very similar, but they have nevertheless a different ESP distribution (c and d).

ranks. While the top ranking molecules are the same for all experiments the ranks 4 and 5 are most dispersed while the exact matches increase again at the bottom of the list.

The larger differences were mainly caused by the comparative scoring experiments against 1TLP and 5TLN. As mentioned before, 5TLN does not have any significant surface similarities with any of the other molecules, which makes a reasonable ranking based on that criterion most unlikely. The situation with 1TLP is more difficult to explain, but the main reason for the bad correlation between the FlexS and SURFCOMP rankings is the fructose residue of 1TLP and the way the rest of the molecules are superimposed to that structural feature. A good example for these effects is the behavior of 5TMN in that experiment: The superposition algorithm aligned the phenyl ring of 5TMN with the fructose moiety of 1TLP and the valine side chain with the indole ring system (Figure 4-9a). These conformational changes make the surfaces of both molecules look very similar (Figure 4-9c, d). However, a surface similarity, which is in a good agreement with the superposition found by FlexS, can only be established with a high RMSD of approx. 2.7 Å between the surface patches due to the large differences between the valine and tryptophan surface and the fructose and phenyl residues (Figure 4-9b). The best ranking cluster is a subset of the closest one, where these different parts are excluded, but it is smaller and therefore ranked after the best clusters of other molecules which are sufficiently larger (e.g. 1TMN or 4TMN).

molecules		SURFCOMP consensus scoring			FlexS	
A	B	rank.	first cluster	score	rank	total score
1THL	1THL	1	1	154.67	1	-1171.90
	1TMN	2	18	313.67	2	-1129.50
	1TLP	3	749	1037.67	3	-969.10
	3TMN	4	1170	1573.67	7	-717.10
	5TMN	5	1495	2125.33	4	-928.50
	6TMN	6	1512	2163.67	5	-872.00
	4TMN	7	1593	2376.67	6	-869.40
	5TLN	8	1677	2585.33	8	-656.00
1TLP	1TLP	1	1	227.00	1	-1425.01
	1TMN	2	18	399.67	4	-1116.86
	6TMN	3	95	603.00	5	-991.06
	4TMN	4	99	605.33	3	-1146.32
	1THL	5	270	1122.67	6	-965.99
	5TMN	6	424	1614.67	2	-1263.97
	3TMN	7	678	2275.67	7	-618.78
	5TNL	8	733	2398.67	8	-605.56
1TMN	1TMN	1	1	215.67	1	-1206.58
	1TLP	2	1008	1104.00	2	-1049.43
	1THL	3	1137	1257.67	3	-1039.12
	5TMN	4	1327	1514.33	4	-1021.14
	6TMN	5	1737	2347.33	5	-964.34
	4TMN	6	1785	2442.33	6	-908.82
	3TMN	7	2061	3067.33	7	-678.71
	5TLN	8	2326	3559.33	8	-654.66
3TMN	3TMN	1	1	22.67	1	-861.91
	1TMN	2	92	131.00	2	-773.78
	1THL	3	111	184.67	3	-758.26
	5TLN	4	157	323.67	5	-603.77
	6TMN	5	162	339.00	4	-614.55
	5TMN	6	190	402.67	6	-586.36
4TMN	4TMN	1	1	135.67	1	-1425.60
	5TMN	2	307	539.00	2	-1264.84
	6TMN	3	583	6693.33	3	-1123.36
	1TLP	4	803	1830.67	4	-991.51
	1TMN	5	1013	2471.67	5	-824.99
	5TLN	6	1034	2547.67	7	-702.75
	3TMN	7	1118	2791.00	8	-557.61
	1THL	8	1140	2846.67	6	-735.21
5TLN	5TLN	1	1	93.67	1	-739.67
	6TMN	2	13	177.33	4	-632.14
	3TMN	3	24	279.33	5	-486.10

molecules		SURFCOMP consensus scoring			FlexS	
A	B	rank.	first cluster	score	rank	total score
	1TLP	4	38	359.67	2	-680.51
	1THL	5	47	383.67	3	-676.75
5TMN	5TMN	1	1	167.00	1	-1499.36
	4TMN	2	262	681.33	2	-1289.80
	6TMN	3	814	1205.33	3	-1225.49
	1TMN	4	1154	1539.33	5	-938.25
	1TLP	5	1521	1999.67	4	-1157.74
	1THL	6	2384	3658.33	6	-873.37
	5TLN	7	2394	3676.67	7	-577.19
	3TMN	8	3023	4875.33	8	-547.60
6TMN	5TMN	1	1	436.67	1	-1466.22
	6TMN	2	218	693.00	2	-1304.29
	4TMN	3	1531	1746.33	3	-1248.28
	1TLP	4	2380	2589.67	4	-1099.71
	1TMN	5	2687	3187.33	5	-883.13
	1THL	6	3016	3983.67	6	-847.09
	3TMN	7	3079	4125.67	8	-517.30
	5TLN	8	3121	4643.33	7	-568.19

Table 4-6: Comparative rankings of all molecules of the thermolysin dataset.
For the SURFCOMP ranking the comparative rank, the appearance of the first cluster of that molecule and the consensus scoring value are given. The FlexS rankings are described by the comparative rank and the total score.

4.1.4. Evaluation of Different Surface Types: Comparing DHFR ligands

The enzyme dihydrofolate reductase (DHFR, EC 1.5.1.3) plays a key role in the folate metabolism of eukaryotic and prokaryotic cells [16]. It is responsible for the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate which is required for DNA, RNA and protein synthesis. Inhibition of DHFR has been a target in drug discovery since many years and different antagonists have been developed. Methotrexate (MTX) has been successfully applied in cancer therapy and trimethoprim (TMP) is a useful drug for the treatment of various infections [121]. The triazine WR99210 is an inhibitor of malarial DHFR but shows some side effects [61]. Because of the presence of DHFR in almost any species selective dihydrofolate antagonists can be antibiotic agents as described by Li et. al. [84] for *Mycobacterium tuberculosis.* Partly because of its pharmaceutical relevance DHFR and the various folate antagonists have become a reference system for molecular modeling. Especially the DHFR/methotrexate complex is a common standard for the validation of docking algorithms [27;51;106;122;134;137;140].

For the present investigation, a set of three folate antagonists together with dihydrofolate (Chart 4-2) was assembled. All data were taken from X-ray structures of complexes with the DHFR enzyme, which were published by Li et. al. [84] (MTX, TMP and Br-WR99210, a derivative of WR99210, henceforth referred to as WRB) and Davies et. al. [38] (folic acid, abbreviated FOL in the sequel) and involved DHFR from *Mycobacterium tuberculosis* and human cells respectively. The antagonists (MTX, TMP

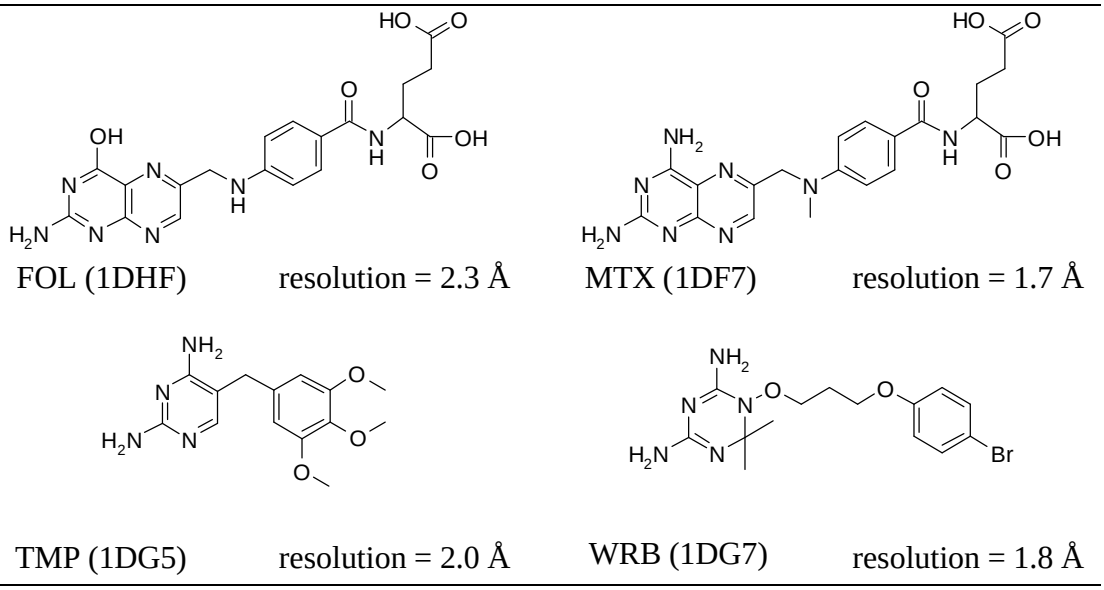


Chart 4-2: DHFR inhibitors
2D structures of folic acid (FOL), methotrexate (MTX), trimethoprim (TOP) and Br-WR99210 (WBR). The codes in parentheses are the identifiers of the corresponding DHFR/ligand complex structures in the PDB database. The given resolution is for the complete protein/ligand complex in the X-ray data.

and WRB) were measured in a ternary complex with the enzyme and one molecule of NADPH bound to its natural binding site, which is not present in the complex of FOL with DHFR. The backbone atoms of the three complexes with the enzyme from *M. tuberculosis* (containing MTX, TMP and WRB) were aligned with an excellent RMSD of about 0.3 Å and the human protein complex with FOL could also be matched to the other protein structures with an error of about 1.0 Å. Hence the four structures could be superimposed within the binding sites of the proteins by the procedure given in section 4.1.1 on page 40.

The common feature of all four structures is a nitrogen-containing heterocycle (pyrimidine, pteridine or triazine) substituted with either one amino and one hydroxyl group or two amino groups. The remaining parts of the molecules are rather different except for MTX and FOL which have the same skeleton. When bound to the proteins the heterocycles are buried in the cleft of the active site. Several hydrogen bonds are formed

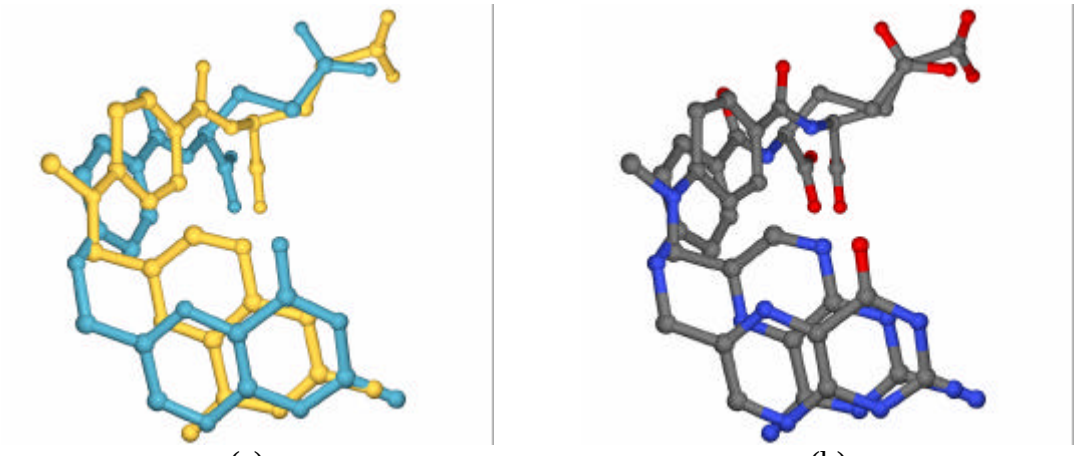


Figure 4-10: Alignment of methotrexate and dihydrofolate in the pocket of DHFR.
On the left side it can be seen clearly that the two pteridine ring systems are not in perfect superposition but are rotated against each other by 60°. The right side displays the two molecules in CPK colors to show which groups are in close contact.

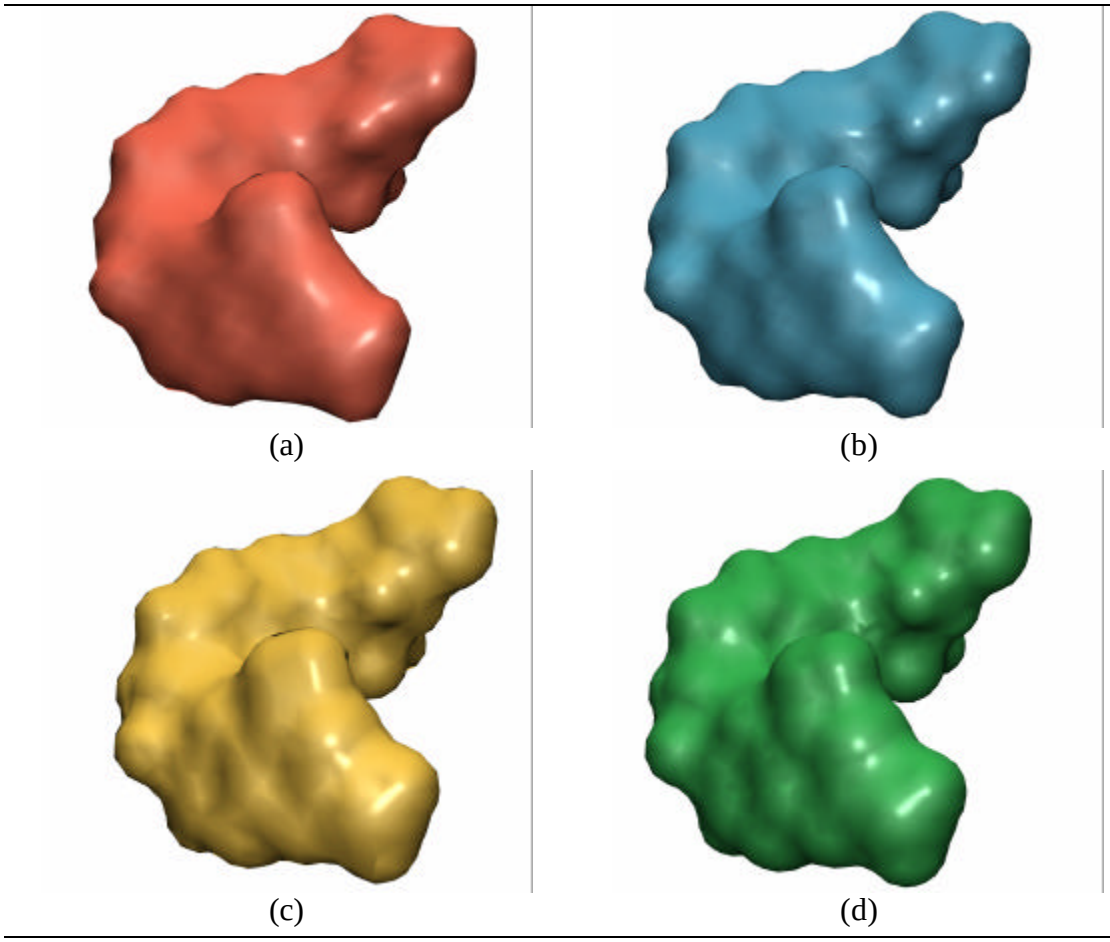


Figure 4-11: Four different molecular surfaces of the folic acid.
Fast-Connolly surfaces generated with MOLCAD with (a) 3 points per Å^2 and (b) 6 points per Å^2; molecular surface generated by Connolly's MS program with (c) 3 points per Å^2 and (d) 6 points per Å^2.

between the nitrogen atoms in the ring systems, the amino or hydroxyl groups of the ligands and different residues of the protein (especially ASP 27 ILE 5) or the NADPH molecule. The other molecules are forming different hydrophobic interactions with various amino acids of DHFR. An interesting difference in the binding modes can be observed between methotrexate and dihydrofolate. Although these two molecules have only two different functional groups (one amine is replaced by a hydroxyl group and a methyl group is added to the nitrogen that connects the pteridine with the phenyl ring), the orientation of their heterocycles is completely different. The two pteridine rings are aligned in a way that the 4-amino group of MTX is aligned with the 2-amino group of FOL. The consequence of this is that the fused rings are rotated by approximately 60 degrees against each other, while the central phenyl rings and the glutamic acids are still in a good superposition (Figure 4-10). One would not expect this constellation by comparing just the 2D molecular structures and it is a challenge for the program not to get confused by the similar looking shapes of the heterocycles.

The three different nitrogen heterocycles that form the common basis of the dataset are posing another problem to surface comparison: due to the planar character of the aromatic or conjugated systems the molecular surfaces around those parts of the compounds have only a few features that can be used as critical points in the SURFCOMP algorithm. In the case of dihydrofolate, only the amino or hydroxyl groups are responsible for a few clear peaks (see Figure 4-11a) and in the other molecules those

features are even symmetric and can lead to upside-down alignments. Consequently the first preliminary experiments did not perform very well (see below). To improve the results for this dataset the features had to be enhanced, especially around the heterocycles.

One possible solution to that problem is to increase the number of points that describe the molecular surface. Increasing the point density will decrease the triangle sizes and will allow the identification of smaller features on the surface. Usually the surfaces were created with 3 points per Å^2, which corresponds to an average triangle area of 0.18 Å^2. To obtain a finer representation a set of surfaces with a point density of 6 points per Å^2 was created. Other factors that control the resolution of a surface are the placement of the surface points and the triangulation process. These parameters are usually fixed for a specific surface generation algorithm, therefore not only MOLCAD's Fast-Connolly surfaces [24] but also the output of Connolly's original MS program [32] was used, which takes longer to compute, but produces a better feature resolution.

To investigate the influence of the different surface types and resolutions on the results of the experiments four different surfaces for each molecule in the DHFR dataset were created: Fast-Connolly surfaces with (a) 3 and (b) 6 points per Å^2 and original Connolly surfaces with (c) 3 and (d) 6 points per Å^2. In Figure 4-11 the different surfaces of dihydrofolate are given as an example for the complete sets. The experimental parameters are summarized in Table 4-7. The results, which are summarized in Table 4-8, show that a significant improvement in the surface alignments as well as in the reproduction of the experimental situations can be obtained if the resolution is increased from 3 to 6 points per Å^2 or if a Connolly surface is used instead of the Fast-Connolly type.

In the initial setup, 3 points per Å^2 Fast-Connolly surfaces, FOL could only be aligned properly with MTX and WRB especially around the heterocycles, but the alignment with TMP was poorer although the amino groups at the heterocycles were aligned correctly. The detected similarities between MTX, TMP and WRB did not cover everything that could be compared and the MTX vs. WRB alignment was completely wrong because the surface of the 2-amino group of MTX was assigned to the surface of the 4-amino group of WRB and vice versa. The surface similarity between TMP and WRB was correct but could not reproduce the experimental data well because of a rather unsimilar critical point pair that was positioned over a methoxy group of TMP and the ether bridge of WRB.

The same calculations performed with high resolution Fast-Connolly surfaces lead to

filter parameter	**symbol**	**section**[a]	**value**	**property**[b]
Curvature cut-off range	c_{CR}	2.2.3	1.0 Å	
neighbourhood radius	r_{CP}	3.2	2.0 Å	
fuzzy threshold	F	3.5	0.3	ESP
shape threshold	R	3.6	0.6	STI
distance tolerance	T	3.7	1.0 Å	
Minimum distance	d_{min}	3.7	0.5 Å	
angular tolerance	f_{tol}	3.8	15.0 °	

Table 4-7: Experimental conditions used in the DHFR ligand dataset experiments.
[a)] the section in the text where the filter is described
[b)] the molecular surface property applied to the specific filter (ESP, electrostatic potential).

Molecules			RMSD [Å]		Molecules			RMSD [Å]	
A	B	points	surf.	struct.	A	B	points	surf.	struct.
(a) MOLCAD surface 3 points per Å²					*(c) Connolly surface 3 points per Å²*				
FOL	MTX	449	1.36	1.23	FOL	MTX	372	0.85	0.73
	TMP	215	1.32	1.90		TMP	359	1.69	0.68
	WRB	257	0.99	1.26		WRB	337	0.99	1.46
MTX	TMP	216	0.64	1.63	MTX	TMP	273	1.14	0.97
	WRB	181	1.11	5.82		WRB	199	0.69	1.56
TMP	WRB	312	1.28	1.74	TMP	WRB	318	0.72	0.51
(b) MOLCAD surface 6 points per Å²					*(d) Connolly surface 6 points per Å²*				
FOL	MTX	890	0.76	1.61	FOL	MTX	954	0.99	1.13
	TMP	377	0.8	0.97		TMP	624	1.83	0.88
	WRB	629	1.02	1.36		WRB	721	0.84	1.58
MTX	TMP	595	1.04	0.87	MTX	TMP	581	0.53	1.37
	WRB	885	1.54	0.74		WRB	739	1.13	0.98
TMP	WRB	396	0.55	0.53	TMP	WRB	470	0.64	0.7

Table 4-8: Results obtained for the surface comparison with different surface types.
Under a-d are the best alignments, identified by visual inspection, for the MOLCAD and Connolly surfaces with 3 and 6 points per Å^2.

better results. For every pair except FOL and WRB, which gave the same quality, either the surface RMSD values or the displacements from the X-ray data dropped significantly. The algorithm could now find a correct alignment between MTX and WRB and the similarities between MTX and TMP were detected more completely. This usually increases the RMSD of the surface superposition, because more points are involved, but improves the fit to the experimental data. For other pairs like TMP and WRB or FOL and TMP the size of the detected surface similarities decreased because the higher resolution supported a better distinction between unsimilar pairs and therefore the representation of the X-Ray data also improved. A drawback of the increased resolution was that the calculation took up to four times longer because of the larger point sets and produced much more alternative clusters than the smaller 3 points per Å^2 surfaces of the initial setup.

An alternative solution, which does not necessarily increase the number points, is the use of a more accurate surface type. The results obtained by the set of Connolly surfaces with 3 points per Å^2 revealed surface similarities comparable to the high resolution Fast-Connolly surfaces. In this surface type the points are placed more carefully to give a better representation of small surface features with the same number of primitives. All pairs were aligned correctly and the symmetry of the amino groups attached to the heterocycles did not cause any problems, as opposed to the case with the low resolution Fast-Connolly surfaces. The quality of the surface superposition and experimental alignment was similar to the high resolution comparisons but the patches were usually larger. Therefore some of the RMSD increased but were nevertheless of the same quality because of the increase in patch size.

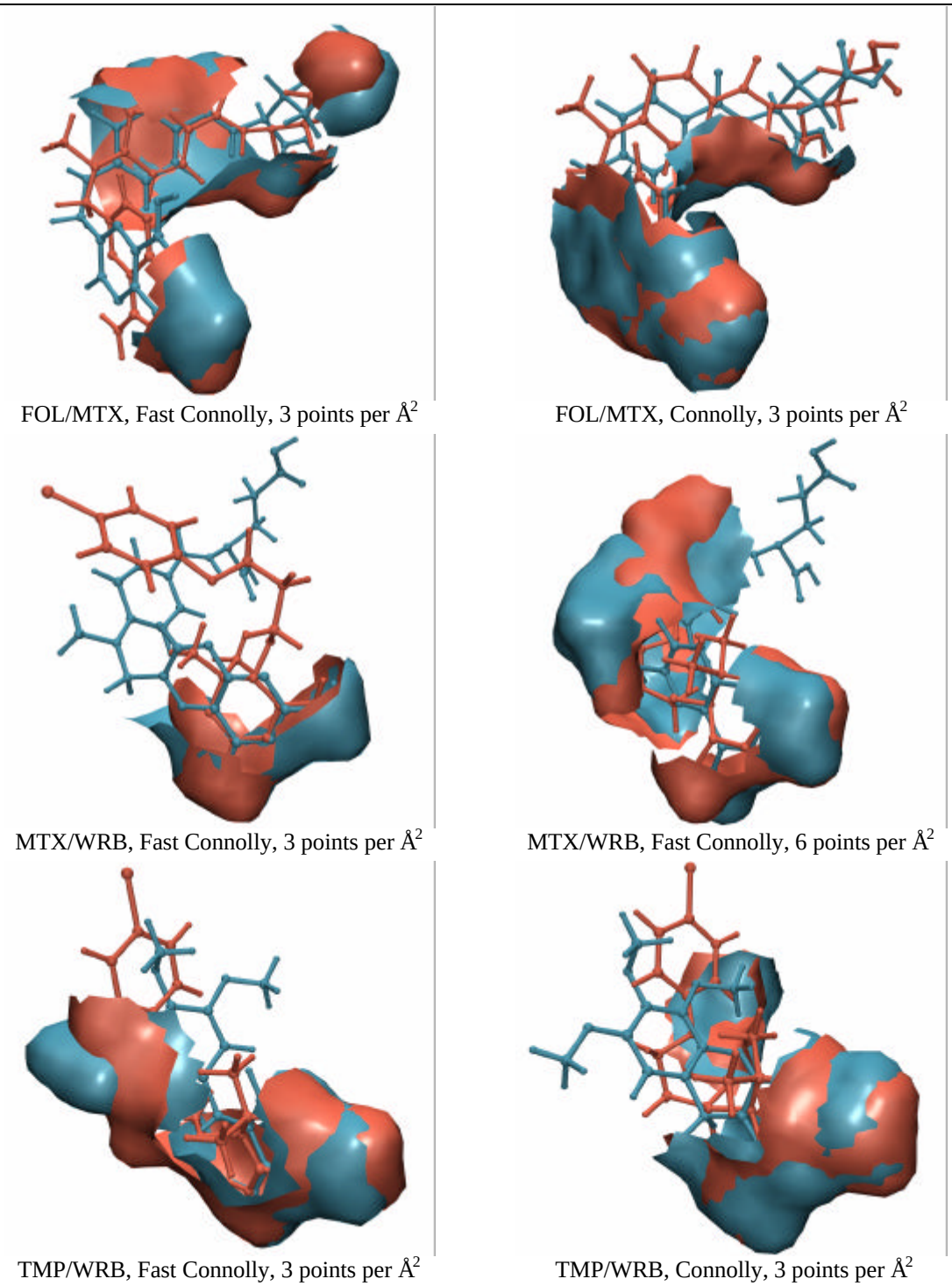


Figure 4-12: Line ups between comparisons performed by different surface types.
The standard surface set (Fast Connolly with 3 points per Å^2) is given on the left and the improved surface sets on the right.
top: The alignment on the right side is based on a much better surface similarity that contains almost the complete area around the heterocyclic ring systems.
middle: only the alignment based on the improved surface set (right) is correct. Watch the orientation of the red structure on the left image.
bottom: presents a similar situation as in the top row; the surface alignment of the improved set is much better due to more precise surface similarities.

The last group of calculations was performed with high resolution Connolly surfaces that had a point density of 6 points per Å^2. The size of the surface similarities were slightly larger or equal to the low resolution Connolly surfaces and the computational effort was comparable to the high resolution Fast Connolly calculations. In this case the RMSD between the detected surface alignments and the fit to the X-ray data did not differ significantly from the other two improved calculations. Only MTX and TMP showed a much better surface alignment while at the same time the fit to the X-Ray data was worse than in the low resolution Fast-Connolly experiments (see Table 4-8d), because only the regions around the heterocycles were considered to be similar. The opposite was the case in the comparison of MTX and WRB. Here the higher resolution surface allowed a better identification of the similarities in the surface regions over the phenyl ring systems in both structures. Three examples that compare the results obtained by the initial setup with those of the improved surface sets are given in Figure 4-12.

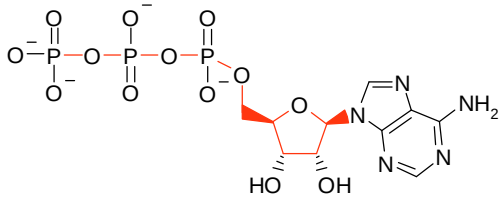
Comparing the results of group 2 and 3 (Table 4-8b and Table 4-8c) leads to the conclusion that in case of featureless surfaces an increase of the surface resolution has almost the same effects as a better placement of surface points and triangles. Increasing the point density is done easily and every surface generation algorithm provides a parameter to adjust that property. But a better point placement or a more sophisticated triangulation algorithm can usually be achieved only by a change of the generation algorithm which may not be possible in certain situations.

4.1.5. Testing different conformations

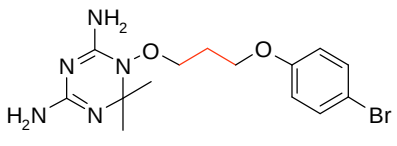
Real molecules are flexible and their actual shape can vary between many configurations that correspond to minima on the potential energy surface. The conformational flexibility of a molecule depends on several parameters including the number of rotatable bonds, the presence of rings and large groups and the environment (whether the compound is docked into an active site of a protein or is in solution). A fixed 3D structure is therefore not always a sufficient representation of a molecule but it provides all the information that is necessary to take flexibility into account. If all the atoms and bonds of a compound are known it is possible to search for new low-energy configurations on the potential energy surface using molecular or quantum mechanics.

Molecular surfaces do not provide information that is necessary to deal with flexibility. They can be seen as a view on a specific conformation that hides any information about the internal structure of the molecule. Therefore it can be difficult, if not impossible, to reproduce a surface similarity between two molecules if different conformations are used for the generation of their surfaces. To what extend the structures can vary to show still the same similarities depends on the surface comparison methodology.

The performance of the SURFCOMP program was tested on different conformations

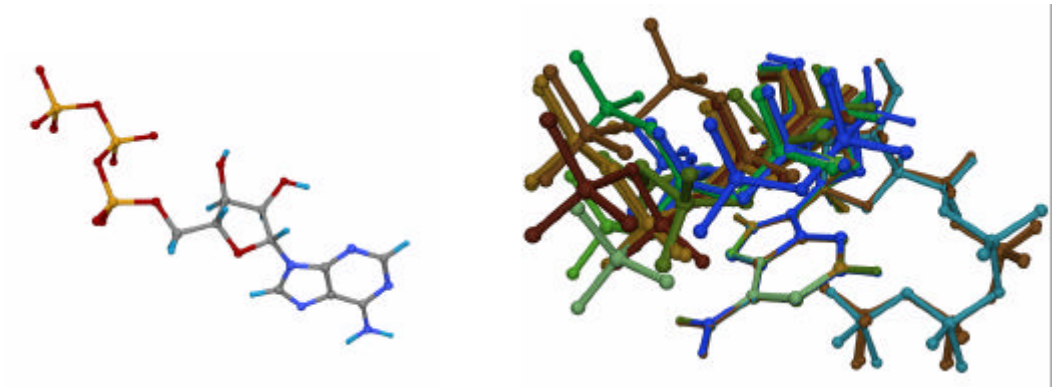


ATP^{4-}



WRB

Chart 4-3: 2D structures of adenosine triphosphate and Br-WR99210 (WBR).
The rotatable bonds considered for the conformational search are printed in red.



(a) (b)

Figure 4-13: Alignment of the generated conformations for ATP^{4-}.
(a) Stretched (natural) structure of ATP^{4-} when bound as a ligand to a protein and (b) alignment of the 14 different conformations as found by the random search. The conformations have been superimposed on the coordinates of the adenosine atoms. The conformations are colored according to their relative energy: blue represents the lowest energies and brown corresponds to high energy structures.

of ATP^{4-} and the DHFR antagonist Br-WR99210 (see Chart 4-3). For both molecules a set of conformations was calculated and the molecular surface of each conformation was compared with a template conformation. The detected similarities were evaluated by the result of a self-match of the template conformation, which in both cases represented an identical one-to-one association between all surface points.

ATP^{4-}. Adenosine triphosphate has usually four negative charges when bound to a protein. Therefore the three dimensional structure of ATP received four negative formal charges at the terminal oxygens of the three phosphate groups. This structure was used to generate a set of different conformations. Because of the large number of rotatable bonds a systematic search in the space of possible torsions was not possible and the random search facility of Sybyl 6.9 [2] was applied with a subset of the free bonds that includes the bond between the ribose and the adenosine, all the C-C and C-O bonds of the ribose, the connection between the ribose and the triphosphate and all the P-O bonds of the triphosphate (see Chart 4-3). The search returned 14 different conformations which were all more compact than the original conformation taken from a protein complex (see Figure 4-13). The energies of these structures varied from 28.83 kcal/mol to 32.56 kcal/mol.

With a random search the completeness of the set of conformations can be assessed by the number of times each conformation was detected by the algorithm. According to Saunders [116] the probability that a set is complete increases with the number *n* of hits for each conformation with $(1-(0.5)^n)$. Thus, if each conformation has been found five times there is a 96.9% chance that all possible conformations have been found. In the search for the ATP^{4-} molecule some clusters where only detected once in 1000 steps of the algorithm. The set is therefore not a representative sample of the available conformational space. Fortunately, for the purposes of the investigation no exhaustive list of low-energy conformers was needed, only a selection of sufficiently different shapes of the molecule.

The lowest energy conformation was taken as a template and compared with all other conformations. The comparisons were performed with Connolly surfaces at a resolution of 3 points/$Å^2$ and the corresponding electrostatic potentials mapped to the points (for the

Conformation	count[a]	E [$^{kcal}/_{mol}$][b]	CPs[c]	points	RMSD surf. [Å][d]	RMSD conf. [Å][e]
1	2	32.35	4	317	1.69	2.82
2	1	32.56	11	397	0.84	2.06
3	1	32.27	6	361	0.83	1.90
4	1	32.27	6	344	1.28	2.76
6	6	30.15	16	646	1.32	1.65
7	7	28.88	21	707	0.71	0.44
8	4	30.47	15	545	0.68	1.18
9	4	31.77	9	424	1.28	1.61
10	2	31.52	12	451	0.71	0.95
11	3	31.48	15	519	0.74	0.77
12	5	29.57	6	433	1.51	2.51
13	3	31.1	12	532	0.79	1.32
14	3	31.73	10	433	1.04	1.45

Table 4-9: Results of the surface comparison of ATP^{4-} conformations
The tests were performed with the lowest energy conformation No. 5 and all other conformations of ATP^{4-}.
[a] number of times this conformation was detected in the random search
[b] total energy of the conformation calculated as calculated during the random search
[c] number of critical points that form the similar regions
[d] RMSD between the similar surface regions and [e] between the conformations

experimental details see Table 4-10). For each pairwise similarity search the top ranking cluster was selected by means of the consensus scoring method and the structural RMSD between the template and the test structure was evaluated as a measure of the conformational difference.

The results, summarized in Table 4-9, reveal that the size of the detected surface similarities decreases with increasing RMSD between the compared conformations. This trend is rather qualitative but it agrees with the expectations. The same trend cannot be observed between the conformational RMSD of the structures and the RMSD of the similar surface areas. For the four most similar conformations (compared to the template)

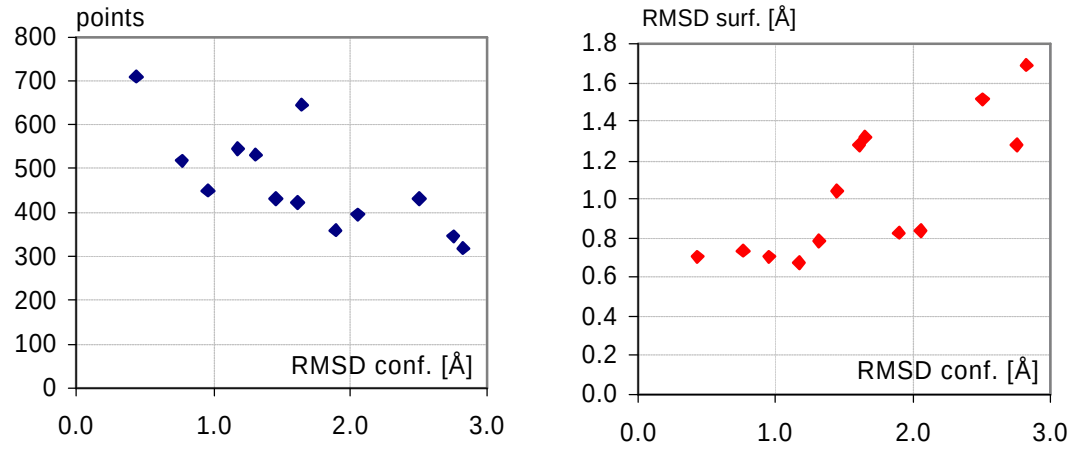


Figure 4-14: Surface similarity vs. conformational difference.
Relations between the conformational difference of the structures and (a) the size of the similar patches or (b) the goodness of the similar surface fit.

the quality of the surface fit is almost equal, and even the similar surface regions of most of the other conformations can be aligned quite well. This is because different 3D structures can nevertheless have common patches on their molecular surfaces. These regions will most probably get smaller and smaller but those parts that match can still fit very well.

Figure 4-15 shows the results of surface comparisons between two similar and two different conformations (4 and 13 in Table 4-9). It is remarkable how well parts of the molecular surfaces match each other even if the RMSD value between the corresponding structures is as large as 2.76 Å. Only the size of the patches for the less similar conformations is significantly smaller compared to the better matching structures. Furthermore, the search between the template and conformation 4 (Figure 4-15 a and b) is a good example for the case that two different structural elements can have a common molecular surface. On the other hand, the second example, comparing conformation 13 with the template, shows that although most of the structure and surface is almost identical, a single difference between the two structures, the position of the third phosphate group, prohibits the recognition of a large patch at the top of both surfaces.

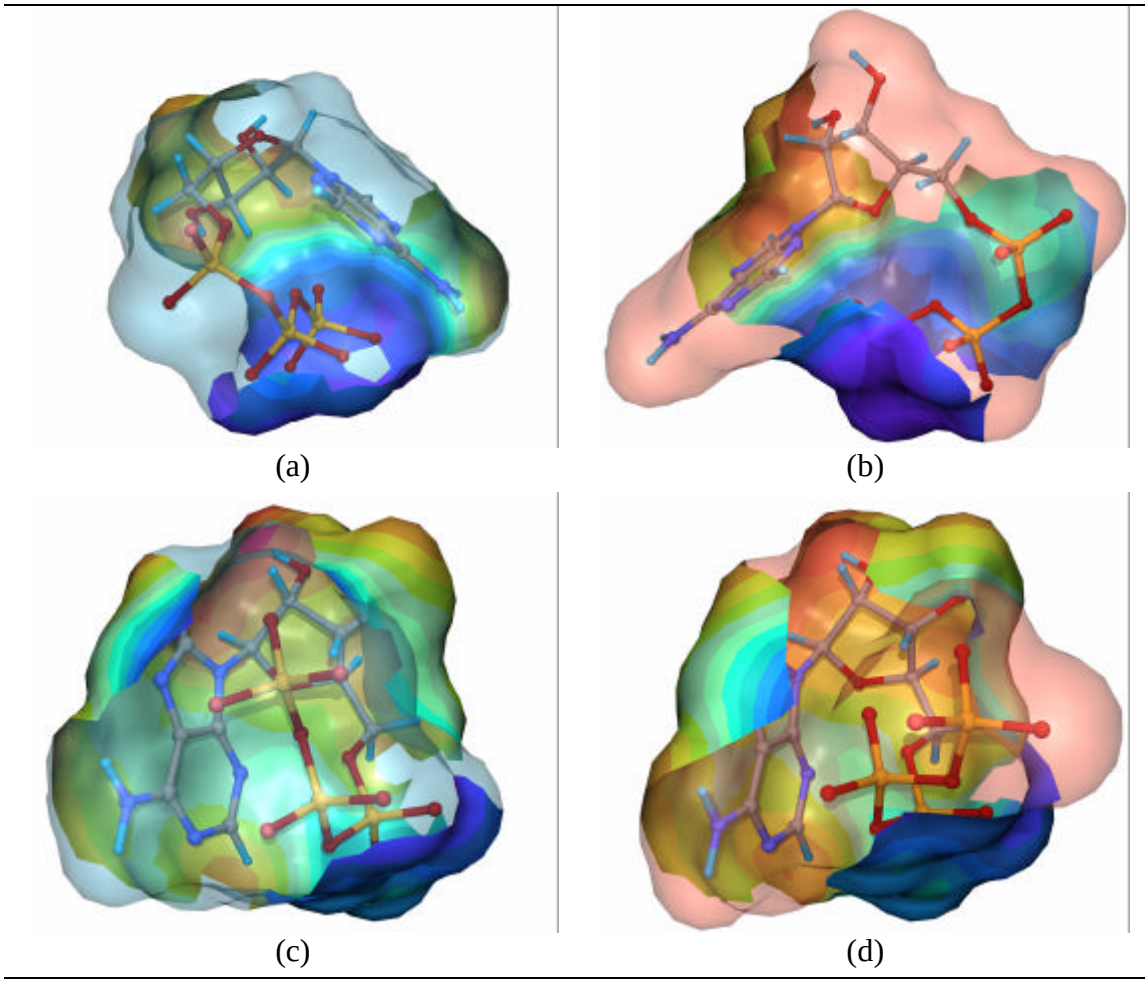
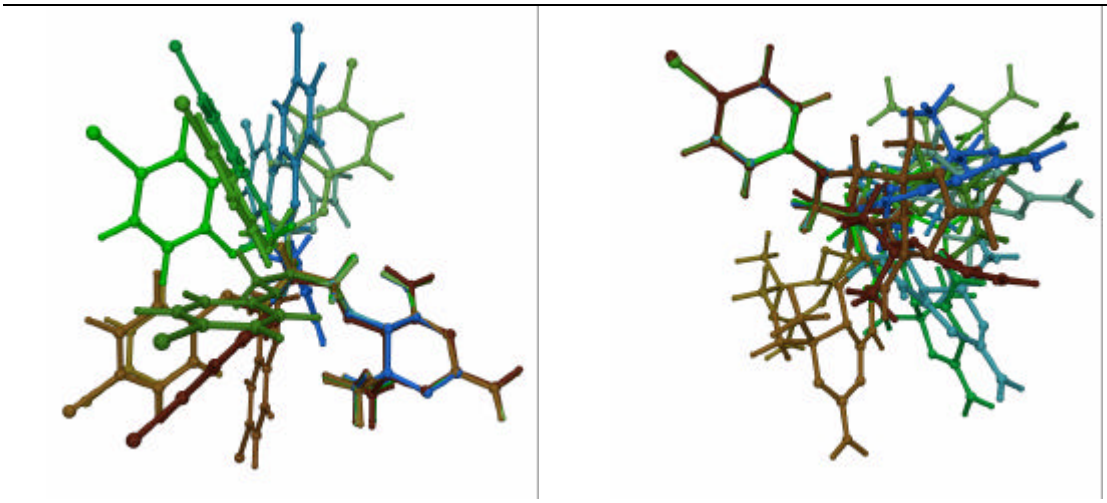


Figure 4-15: Surface alignments of the template and two calc. conformers of ATP^{4-}.
(a) and (b) contain the template (blue) together with a bad matching conformation (red, see 4 in Table 4-9), while (c) and (d) display the alignment of the template with a well matching conformer (see 13 in Table 4-9). The surfaces are color coded by the electrostatic potential, where blue corresponds to a negative and red to a positive charge. One can see that the surfaces that match in the first example do not cover corresponding parts of the molecular structure. E.g. the positive patch on the upper left corners belongs to the ribose in (a) and to the adenosine residue in (b).



(a) (b)

Figure 4-16: Alignment of the generated conformations for WRB.
The conformations are colored according to their relative energy: blue represents the lowest energies and brown corresponds to high energy structures. (a) Shows the relative orientation of all conformations when the structures are superimposed by the triazine rings and (b) gives the same situation for an alignment via the bromo-phenyl residues.

WRB. To investigate the actual influence of distinct conformational changes the DHFR ligand Br-WR99210 (WBR) was taken from the protein structure 1DG7. The molecule consists of two rather inflexible parts, a substituted triazine ring, which is responsible for the protein binding and a bromo-phenyl residue on the opposite end. These two parts have a very characteristic molecular surface which should be recognized easily between different conformations. The flexibility of the compound is mainly due to the ether bridge that connects the two rigid parts. Hence large changes in the 3D structure and thus in the surface can only happen in this region of the molecule. If the conformational search is focused on this area one should obtain a set of structures that will have a surface match over the rigid parts but no similarity in between. The question is to what extent these two similarities can be detected by a single surface alignment.

In this experiment a systematic search was used to generate a set of conformations. For that the two central bonds of the ether bridge were selected to rotate freely. The torsions around these bonds were changed in steps of 60 degrees which after minimization resulted in 36 different conformations having energies between 5.63 and 16.87 kcal/mol and RMSD to the original structures of 0.84 to 3.06 Å. A subset of 12 conformations is shown in Figure 4-16. From this picture one can see that the main conformational differences are the relative orientations of the triazine and bromo-phenyl residues.

With each molecule in that subset a surface comparison against the original 3D structure from the PDB structure 1DG7 was performed. The surface type and the calculation of the physicochemical properties were equal to the ATP^{4-} tests and the experimental details are given in Table 4-10. To detect how much of the surface is preserved by each conformation the results of the SURFCOMP program were searched for the clusters that included the patches around the triazine and the bromo-phenyl ring. They could be found more easily when the alignments based on the surface similarities were compared against the two different structural superpositions shown in Figure 4-16. The consensus scoring was then used to rank the clusters according to one of these RMSD differences, the size of the similar patches and the chemical correlation. This

			triazine		bromo-phenyl	
Conf.	E [$^{kcal}/_{mol}$][a]	RMSD conf. [Å][b]	points	RMSD [Å][c]	points	RMSD [Å][c]
2	8.79	1.67	452	1.19	505	0.76
9	11.36	1.15	462	0.66	378	0.56
10	6.24	1.85	485	0.63	337	0.83
13	16.87	0.84	494	0.74	395	0.69
21	10.97	2.65	436	0.68	402	0.69
23	10.75	2.21	457	0.66	392	0.89
24	6.66	1.84	483	1.17	432	1.11
25	16.27	2.79	448	0.92	377	0.97
26	10.25	2.94	330	0.53	428	0.75
27	15.65	3.06	382	0.61	373	0.62
30	10.55	2.65	392	0.79	394	0.82
32	5.63	2.65	362	0.85	437	0.79

Table 4-10: Results of the surface comparison between the conformations of WRB.
[a] total energy of the conformation calculated as calculated during the random search
[b] difference between the calculated conformation and the original structure
[c] goodness of fit of the similar surface regions.

variation to the usual scoring procedure identified in all cases the largest possible surface similarities that were centered on one of the rigid areas in the molecules.

The results show that the correlation between the size of the similar patches and the RMSD of the conformations is still similar to the relationship detected by the ATP^{4-} example and that the matches between the single rigid parts are found in every comparison. However, in almost any case – even with very small differences in the 3D structure – the two rigid areas could not be detected by a single cluster. Only if features in the ether bridge were similar they were included into the clusters that represented the conserved areas.

These results, from the ATB and WRB tests, emphasize the fact that conformational changes are a critical perturbation when two different molecules are compared. However, individual features that do not change their conformation easily are most likely detected as similar even if the total 3D structures are very dissimilar.

filter parameter	**symbol**	**section**[a]	**value**	**property**[b]
Curvature cut-off range	c_{CR}	2.2.3	1.0 Å	
neighbourhood radius	r_{CP}	3.2	2.0 Å	
fuzzy threshold	F	3.5	0.4	ESP
shape threshold	R	3.6	0.5	STI
distance tolerance	T	3.7	1.0 Å	
Minimum distance	d_{min}	3.7	0.5 Å	
angular tolerance	f_{tol}	3.8	15.0 °	

Table 4-11: Experimental conditions used in the conformation tests.
[a] the section in the text where the filter is described
[b] the molecular surface property applied to the specific filter (ESP, electrostatic potential).

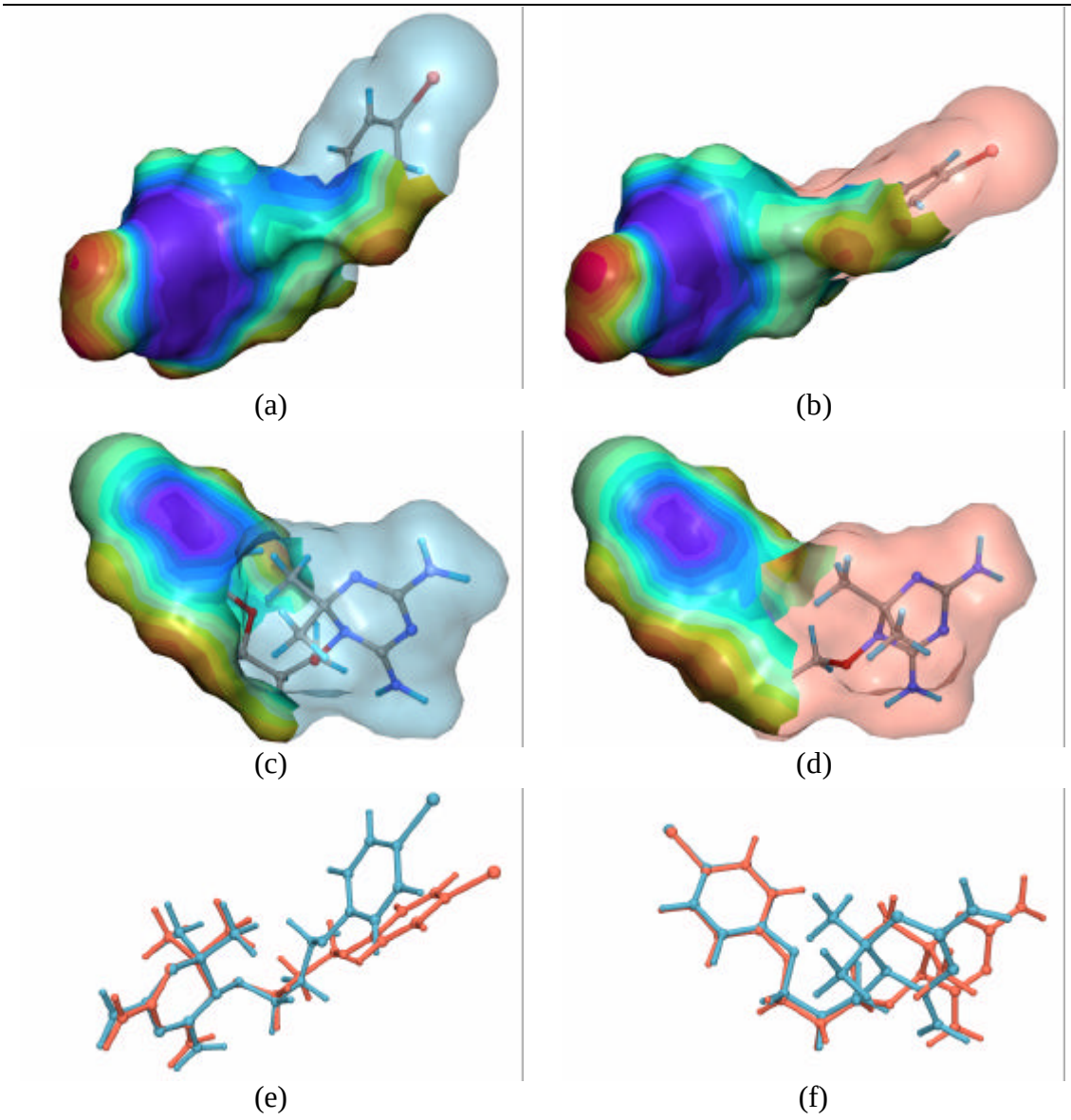
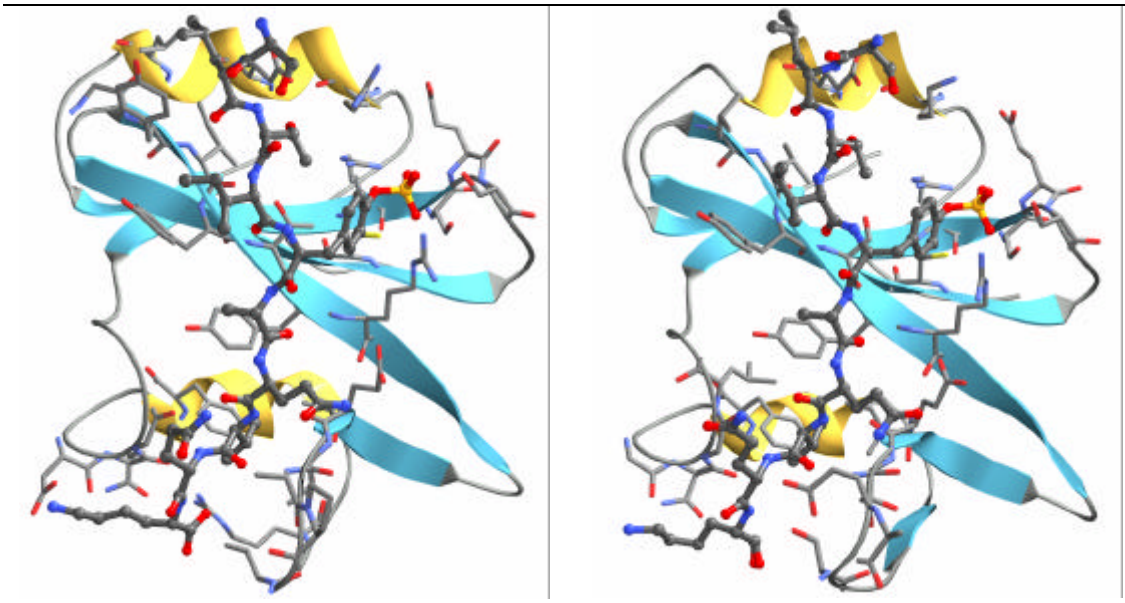


Figure 4-17: Separated surface similarities between WRB conformers.
In the top row the similar regions that matches the triazine areas are displayed (a, b), in the middle the similarity between the surfaces around the bromo-phenyl part are shown (c, d) and the bottom lines up the corresponding alignments between the two molecules based on the triazine (e) and bromo-phenyl (f) similarity.

4.2. Comparing Proteins: Surface Differences between SAP and EAT-2

SAP and EAT-2 are both representatives of SRC homology 2 (SH2) domains, which are key elements in tyrosine kinase regulation of cellular processes. The mechanism is usually triggered by the binding to peptide sequences that contain phosphorylated tyrosine residues (pTyr). SH2 domains consist of approximately 100 amino acids and can be found in a large number of proteins. Normally they can be found in higher eukaryotic cells but some evidence exists that they are also present in yeast [88]. The common fold of SH2 domains consists of a central β sheet core and a separate, small antiparallel β sheet which are flanked by two α helices, one on each side (see Figure 4-18). The phosphorylated tyrosine residue of a cognate ligand binds orthogonal to the β sheet core and residues from one side of the core and of the N-terminal α helix are forming



SAP (1D4W) resolution = 1.80 Å EAT-2 (1I3Z) resolution = 2.15 Å

Figure 4-18: Structure of the SAP-pSLAM and EAT-2 pSLAM complexes.
In both pictures the ligand peptide is displayed in bold, dark balls and sticks together with the amino acids of the protein in capped sticks that are located within 4.0 Å of a ligand atom. The PDB codes are given in the individual image captions together with the X-ray resolutions.

coordinative bonds to the ligand. The loops that connect the different structural elements can vary between the different members of the SH2 family and the affinity of a SH2 domain to a ligand peptide depends strongly on the first three amino acids that follow downstream of the pTyr [117].

When coordinated to pTyr-containing signal peptides, SH2 domains can form various protein/protein interactions with catalytic domains like tyrosine kinases or adaptor proteins like CRK or GRB2 [118]. Thereby they serve as an additional regulation mechanism in the orchestration of signal transduction that supplements the phosphorylation/dephosphorylation mediation via kinases and phosphatases. Hence, their function makes SH2 domains very interesting targets from the drug discovery point of view. Blocking SH2 domain dependent protein-protein interactions is a promising strategy for a variety of different diseases from cancer and osteoporosis to allergy and inflammatory diseases [21]. For the same reasons, selectivity between different SH2 domains is a very important factor. To avoid side effects it is absolutely necessary to target only one member of the SH2 family by an inhibitor. Therefore, the studies on the surfaces of SAP and EAT-2 concentrated on the differences of their cognate ligand binding sites.

SAP is a free SH2 domain that inhibits signal transduction events induced by a series of receptors on the surface of T lymphocytes and natural killer cells (NK). A mutation in the gene encoding SAP (*SH2DIA*) is involved in the X-linked lymphoproliferative disease (XLP), a rare immune disorder that renders the immune system unable to respond effectively to the Epstein-Barr virus [100]. SAP interacts with the consensus motif in the cytoplasmic tail of SLAM (CD150) in the phosphorylated and also in the dephosphorylated form, thereby blocking the recruitment of the SHP-2 phosphatase to that position in the receptor. Recently two groups independently discovered that the interaction of SAP with the SH3 domain of the SRC-family kinase FynT couples this kinase to SLAM [28;78].

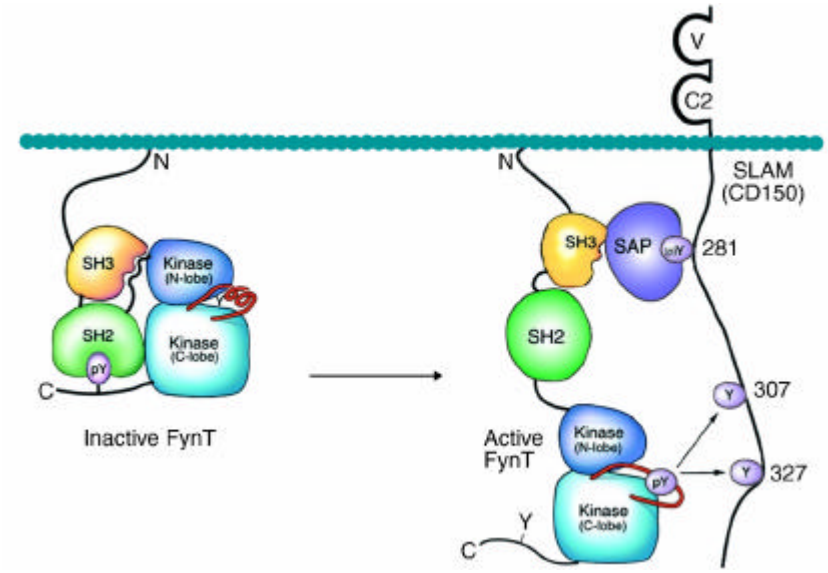


Figure 4-19: A mechanism for SLAM-induced recruitment and activation of Fyn.
The inactivated form is shown on the left side and the SAP-activated form is given on the right side. (figure taken from Chan et. al. 2003 [28]).

In an experimental study, Li et. al. discovered [85] that SAP has interesting relative binding affinities to variations of the native SLAM peptide. They tested the relative dissociation constants of parts of the signaling peptide of pSLAM against the full and dephosphorylated sequence (SLAM). It was found that the N-terminal part of pSLAM is more important for the binding than the C-terminal part which is unique among the members of the SH2 family (see also Figure 4-20).

EAT-2 is a very similar SH2 domain that is expressed in macrophages and b-lymphocytes [99]. EAT-2 too can be associated to SLAM and acts as a SHP-2 blocker but no interactions with the SH3 domain of FynT are reported. Analogously to SAP, it binds to the phosphorylated cytoplasmic tail but unlike SAP it does not bind to the dephosphorylated receptor. Therefore, in contrast to SAP the binding of EAT-2 to SLAM is significantly more dependent on the tyrosine phosphorylation. This selectivity towards pTyr and the different locations of SAP and EAT-2 make the system an interesting target for a selective blocking of the SH2 signal peptide interactions.

Several protein structures for SAP and EAT-2 in the native form and in complex with the phosphorylated and dephosphorylated SLAM-tail peptide (SLTI-(p)T-AQVQK) are available. An overview is given in Table 4-12. In this study X-ray structures of both proteins in complex with the phosphorylated SLAM were used to determine the

Structure	PDB	Technique	resolution	ref.
unliganded SAP	1D1Z	X-ray	1.40 Å	[107]
SAP in complex with p-SLAM	1D4W	X-ray	1.80 Å	[107]
SAP in complex with SLAM	1D4T	X-ray	1.10 Å	[107]
SAP bound to the N-Y-C peptide	1KA7	NMR		[68]
SAP bound to the N-pY peptide	1KA6	NMR		[68]
SAP/FynSH3/SLAM ternary complex	1M27	NMR		[28]
EAT-2 in complex with p-SLAM	1I3Z	X-ray	2.15 Å	[107]

Table 4-12: Available protein structures for SAP and EAT-2.

fragment	sequence
	RKSLTIpYA
	RKSLTI YAQVQK
	TIpYAQVQK
	RKSLTIpYAQVQK

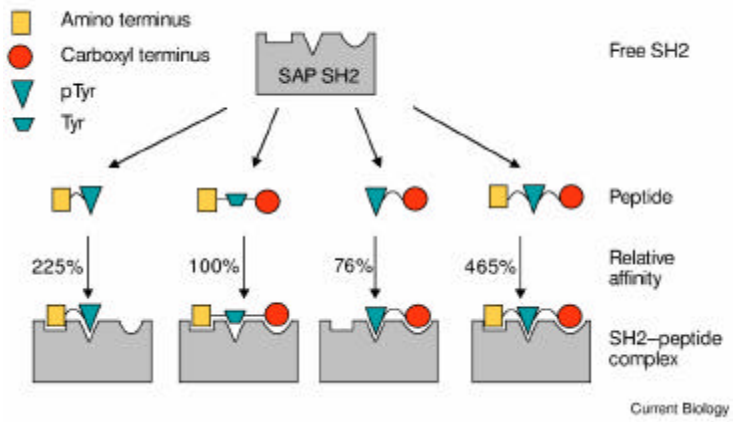


Figure 4-20: Relative binding affinities between SAP and different SLAM peptides.
The figure is taken from Li et. al. [85] and compares the relative binding affinities of different SLAM peptides with the binding sites of the SH2 domain SAP.

differences on the surface regions that are involved in the ligand binding (1D4W and 1I3Z). Sketches of both protein/ligand complexes are given in Figure 4-18 and the result of a sequence alignment is displayed in Figure 4-21.

Earlier studies revealed that the consensus sequence motive T/S-x-pY/Y-x-x-V/I is responsible for the SLAM recognition in SAP [83;85], where x represents any amino acid, and pY/Y (phospho-tyrosine or tyrosine) can be replaced by other amino acids. The three fixed residues of this motif are bound to three well formed cavities on the surface of SAP and corresponding binding pockets can be found in EAT-2. It was now of particular interest to investigate the cavities and to detect any differences in the molecular surfaces around those regions. If such differences are based on structural variations, they may highlight positions where a selective binding to SAP but not to EAT-2 could be successful. Differences due to different conformations will probably disappear if a new ligand induces a conformational change.

4.2.1. Surface Comparison

The investigation was focused on the molecular surface that was in close contact with the pSLAM peptide. Close contact was defined by selecting only those critical points on the surface which were located within 8.0 Å of the following atoms on the pSLAM peptide:

1. the carbon atom of the closer methyl group in the side chain of leucine 278 (CD1),
2. the oxygen of the hydroxyl group of threonine 279 (OG1),

```
SAP      1 MDAVAVYHGKISRETGEKLLLATGLDGSYLLRDSESVPGVYCLCVLYHGYIYTYRVSQTE
EAT2     1 MD-LPYYHGCLTKRECEALLLKGGVDGNFLIRDSESVPGALCLCVSFKKLVYSYRIFREK

SAP     61 TGSWSAETAPGVHKRYFRKIKNLISAFQKPDQGIVIPLQYPVEK
EAT2    60 HGYYRIETDAHTPRTIFPNLQELVSKYGKPGQGLVVHLSNPIMR

pSLAM  276 SLTI-pT-AQVQK
SLAM   276 SLTI-T-AQVQK
```

Figure 4-21: Sequence alignment between SAP and EAT-2.
The residues that are in close contact (6.0 Å) to the ligand peptide are displayed in blue (SAP) and red (SAP). A **|** means residue identity and **:,·** strong and weak chemical similarity.

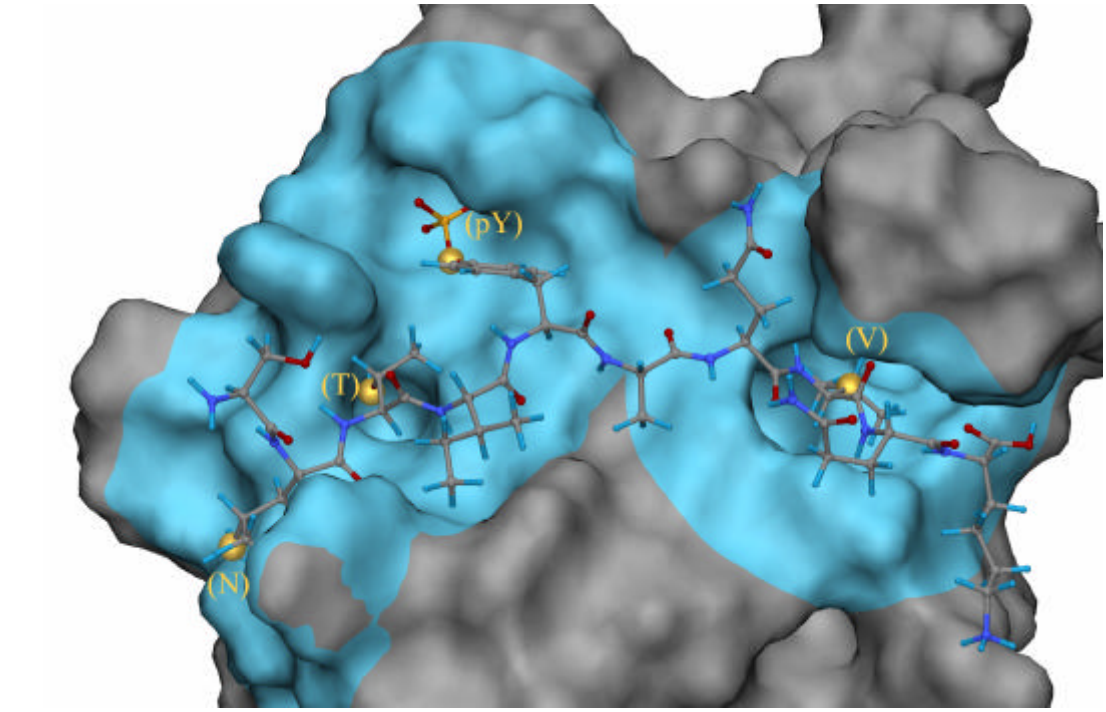


Figure 4-22: Surface regions considered in the comparison of the SAP and EAT-2.
To detect differences in the molecular surface beneath the ligand peptide, the areas on both molecules around the N-terminal residue (N), the threonine 279 (T), the phospho-tyrosine 281 (pY) and the valine 284 (V) of pSLAM were compared with each other. The yellow spheres indicate the atoms that served as central points of these regions and the blue patch defines the selected surface area.

3. the oxygen connecting the phosphate group with the sidechain of p-tyrosine 281 (OH) and
4. the β carbon in the sidechain of valine 284 (CB).

The first center represents the N-terminal part of the ligand peptide and the last three atoms are placed within the three binding cavities of the proteins that bind the fixed residues of the consensus sequence motif. Figure 4-22 shows the molecular surface of SAP with the considered regions highlighted.

A surface similarity search with SURFCOMP was performed for each of the four corresponding centers on SAP and EAT-2. To work out all the possible differences the parameters were tuned in a way to retrieve only the most significant surface similarities (Table 4-13). For the physicochemical property used in the fuzzy filtering the electrostatic potential of the protein was selected, which was calculated as described in section 3.11 (p. 37). Initially the results of each comparison highlighted only the differences in one region. To get the overall view of the complete binding area the best clusters of all four computations were combined into one picture that gives a good overview of the surface differences of SAP and EAT-2 binding to pSLAM.

Figure 4-23 and Figure 4-24 show that differences between the binding surfaces are located at the N-terminal part, at the threonine binding pocket, around the pTyr-281 location and inside of the valine-284 cavity. The central pTyr-284 binding pocket seems to be different on the upper rim and on the left side where it flanks the threonine cavity. The latter difference is mainly due to a single surface feature that corresponds to the side chain of Lys-12 in EAT-2 and the guanidine group of Arg-13 in SAP which do not show any strong interaction with the residues of the ligand. The other difference in that part, covering the binding pocket from the upper left corner is caused by different conformations of the sidechains of the glutamic acids 34 (EAT-2) and 35 (SAP). It is

unlikely that these differences can serve as a starting point for a selective SAP/SLAM blocking.

Of more interest are the differences in the threonine binding pocket and the valine cavity because they cover two of the three structural motifs that seem to be responsible for the recognition of the ligand. The threonine cavity in EAT-2 is wider than but not as deep as the corresponding feature on the SAP surface. Furthermore the entrance to the cavity from the right (in Figure 4-23 and Figure 4-24) is steeper in SAP than in EAT-2. The situation around the valine pockets is even more interesting, because the differences there are larger and more complex. The finger that encloses the cavity from above the surface is much more negatively charged in EAT-2 than in SAP and the shape of that region is also quite divergent. The most important differences are found at the bottom of the pocket. There SAP has two little extra cavities that are separated by a small ridge. On EAT-2 the bottom of the valine pocket is rather flat and has no pronounced hole or ridge.

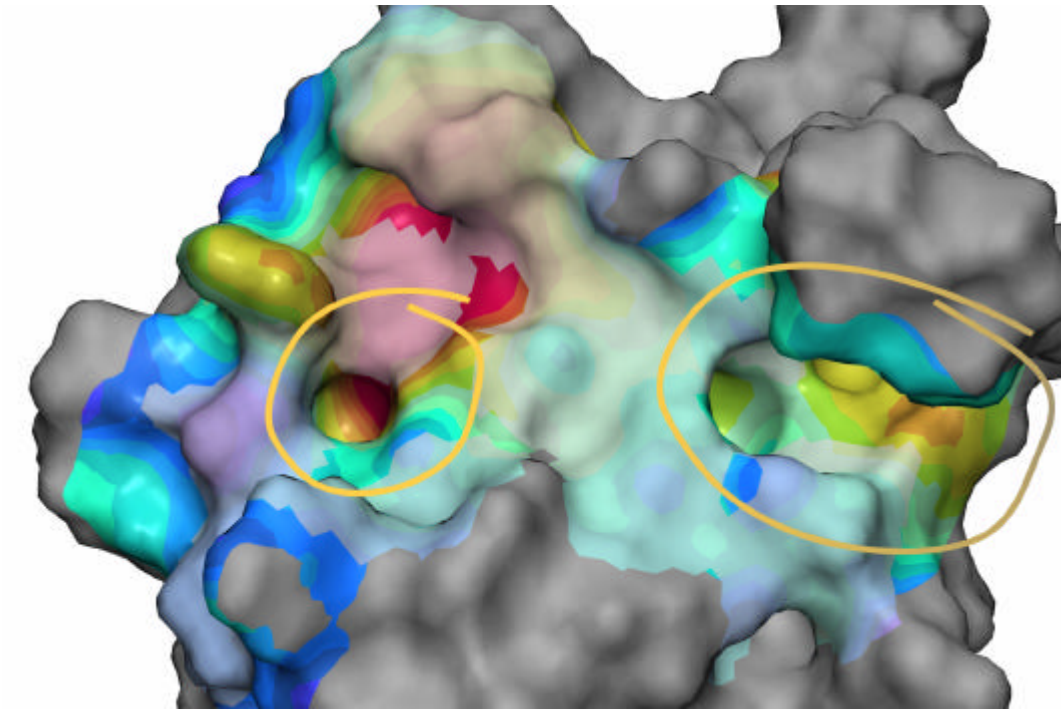
It is noteworthy that the corresponding surface patches of the proteins, which are in contact with variable parts of the consensus sequence motif T/S-x-pY/Y-x-x-V/I, are highly conserved. Neither the region beneath the Ile-280 nor the patch close to Ala-282 and Glu-282 show any significant differences, although they do not have a lot of features. These findings support the consensus motif from the perspective of the surfaces, because a flat and featureless region does not provide many anchor points which are necessary for discrimination.

4.2.2. Structural Investigations

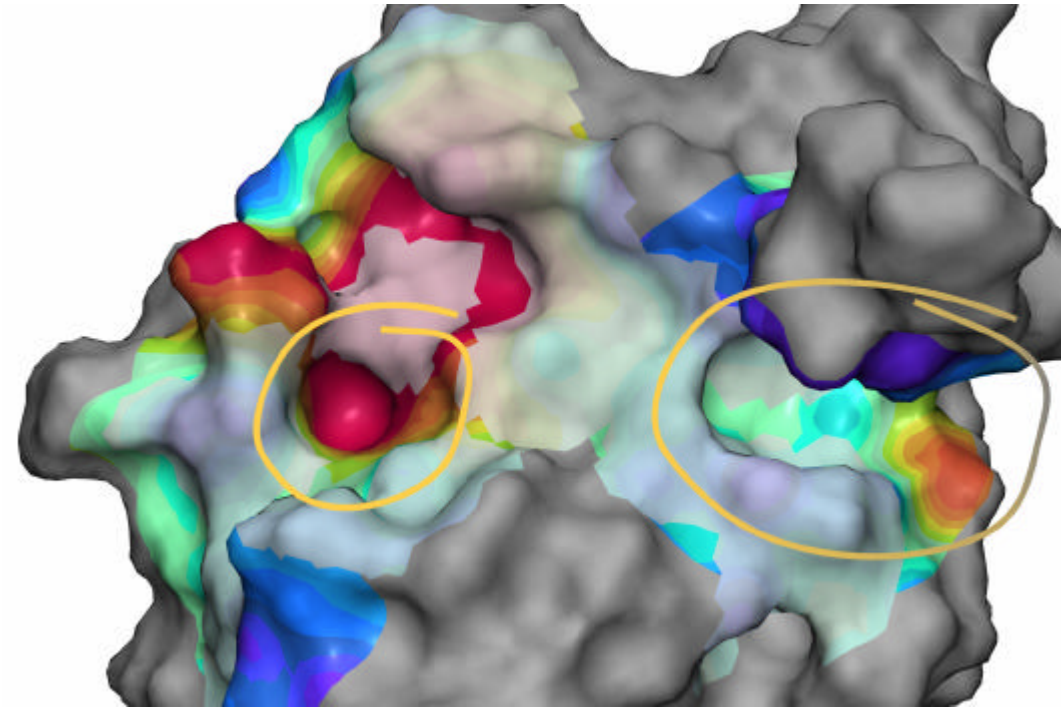
To evaluate the potential of the differences to serve as starting points for SAP/EAT-2 selectivity the structural configurations that lead to the dissimilarities in the binding surfaces have to be examined. Therefore the clusters that represented the best picture of the surface differences at each of the four sites were exported into the molecular modeling package SYBYL 6.9 [2] together with the corresponding structural and surface data. In the molecular viewer the residues that are responsible for the differences in that area could be identified easily and the surface was regenerated only for those amino acids to focus the eye of the observer on the relevant parts.

filter parameter	symbol	section[a]	value	property[b]
curvature cut-off range	c_{CR}	2.2.3	2.0 Å	
neighbourhood radius	r_{CP}	3.2	2.0 Å	
fuzzy threshold	F	3.5	0.3	ESP
shape threshold	R	3.6	0.6	STI
distance tolerance	T	3.7	1.0 Å	
minimum distance	d_{min}	3.7	0.5 Å	
angular tolerance	f_{tol}	3.8	15.0 °	

Table 4-13: Experimental conditions used in the SAP/EAT-2 comparisons.
[a)] the section in the text where the filter is described
[b)] the molecular surface property applied to the specific filter (ESP, electrostatic potential).

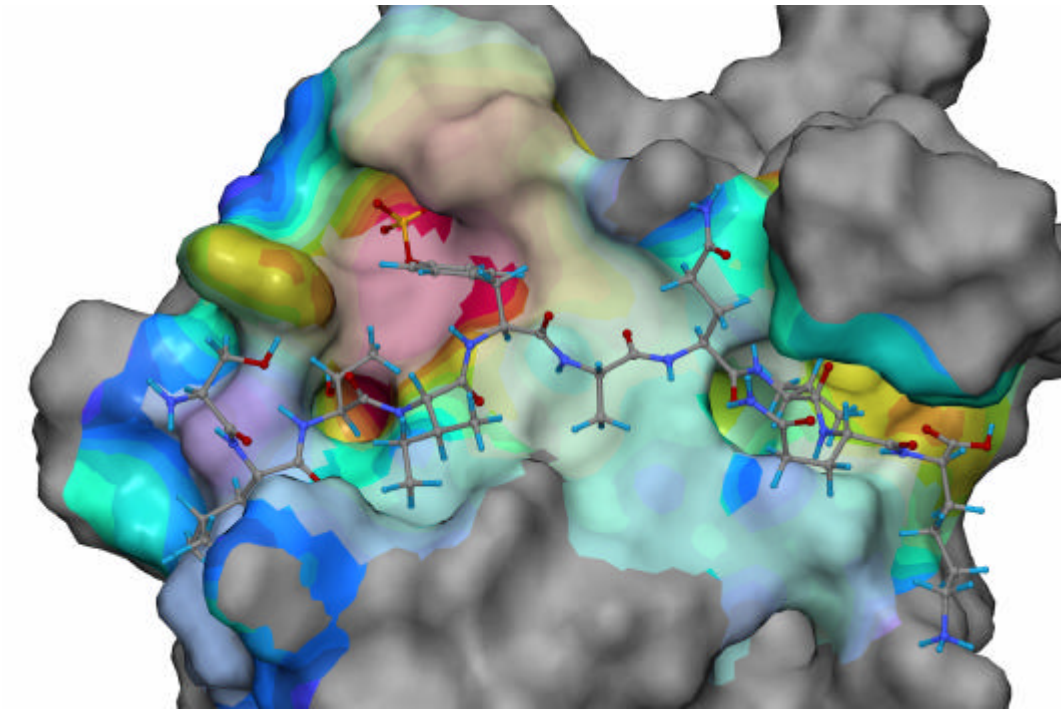


SAP

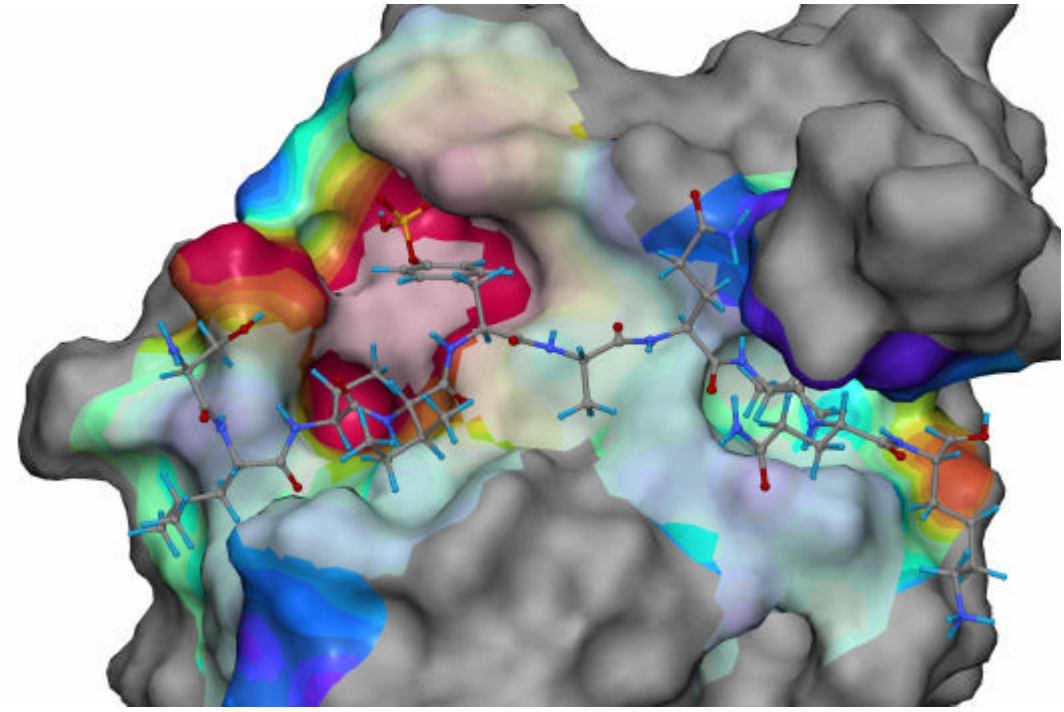


EAT-2

Figure 4-23: Surface differences between SAP (above) and EAT-2 (below).
The figure shows the differences in the surface areas that are involved in the pSLAM binding in intensive colors. The similar surface is highlighted with less intensive colors while the surface areas that were not compared are displayed in gray. The colors are encoding the electrostatic potential (ESP) of the surfaces, where blue indicates negative and red positive areas. The yellow circles indicate the dissimilarities that have been investigated in more detail on a structural level (see text).



SAP



EAT-2

Figure 4-24: Surface differences between SAP and EAT-2 with structures.
The figure shows the surfaces in the same way as Figure 4-23 but in combination with the structure of the pSLAM peptide.

The residues that form the threonine cavity in SAP and EAT-2 are very similar and the relative conformations of the residues in each pocket are also highly conserved (Figure 4-25). But the surfaces are nevertheless divergent at several points which are related to the differences in the amino acid sequence. As mentioned above a significant dissimilarity is caused by the patches that are placed around Arg-13 in SAP and Lys-12 in EAT-2. The most important difference, however, is located right at the center of the cavities where a glycine residue in SAP (Gly-16) is exchanged against a cysteine residue in EAT-2 (Cys-15). The missing side chain causes the pocket of SAP to extend deeper into the protein than in EAT-2, where the side chain of the cysteine is blocking the way. In the crystal structure of SAP the larger cavity is occupied by two water molecules which seem to be tightly bound to the protein as judged by their low B-factors of 15.25 $Å^2$ for the inner and 17.89 $Å^2$ for the outer water respectively. In EAT-2 the corresponding pocket holds only one molecule of water which is much more mobile (B-factor of 39.07 $Å^2$).

Similarly to the situation of the threonine cavity, the pocket that binds the Val-284 residue of the pSLAM ligand consists of some conserved and some divergent residues both in SAP and EAT-2. In contrast to the threonine cavity, the shapes of these valine cavities differ not only at the center but also at the peripheral sections. However the most

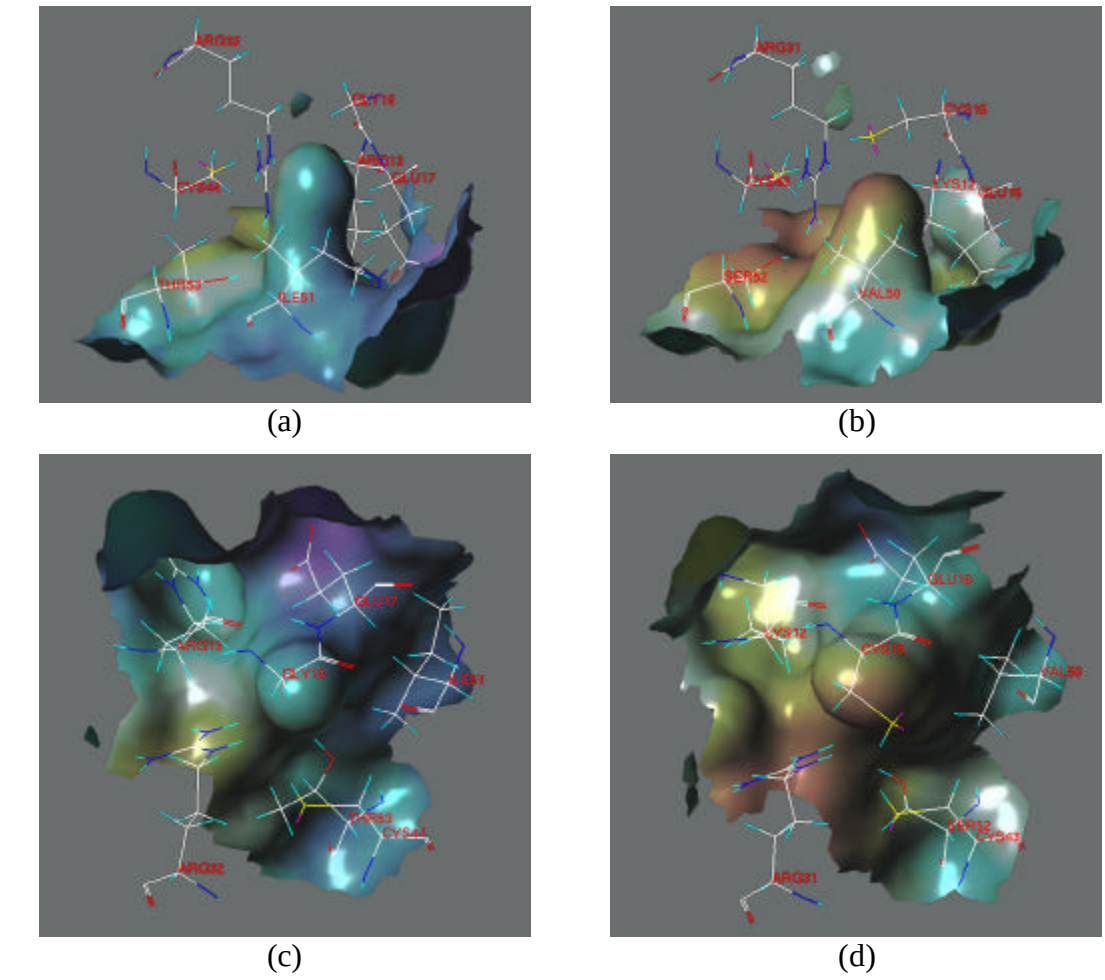


Figure 4-25: Structural conformation of the threonine cavity in SAP and EAT-2.
All four images are presenting the inside of the cavities' surfaces of SAP (left) and EAT-2 (right). In the top pictures (a + b) the different depth of the pockets is illustrated and in the bottom row (c + d) the effect of the cysteine sidechain is shown. From these pictures one can figure out easily how the mercapto-methyl group is limiting the extension of the cavity.

interesting part is again the central pocket. In the middle of the valine cave EAT-2 has only a single shallow hole that is enclosed by a leucine (Leu-93) and isoleucine (Ile-65) residue. SAP has two deeper but smaller cavities at the same position which share a common entrance similar to the entrance of the single EAT-2 hole. These two cavities are encircled by two phenylalanine residues (Phe-77 and Phe-87), one alanine (Ala-66) and one leucine (Leu-43). In contrast to the threonine binding site the valine pockets in SAP and EAT-2 do not contain bound water molecules in the crystal which is due to the hydrophobic character of the residues involved. To illustrate how different the depth of the two pockets actually is, consider that the bottom of the cavity in SAP is formed by Leu-43. This residue corresponds to Leu-42 in EAT-2 which is buried deep inside the protein and does not have any contact to solvent molecules.

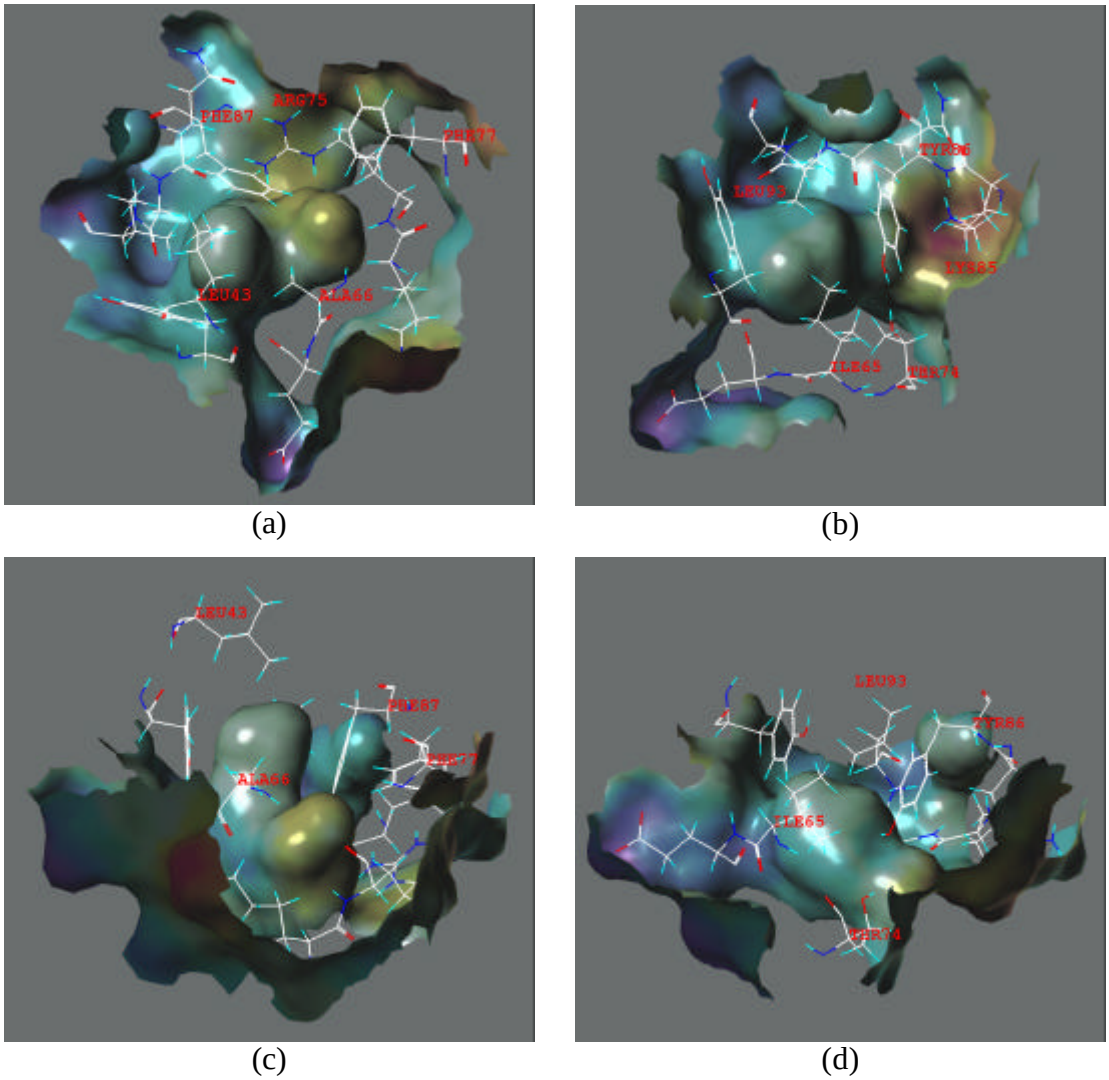


Figure 4-26: Structural conformation of the valine cavity in SAP and EAT-2.
All four images are presenting the inside of the cavities' surfaces of SAP (left) and EAT-2 (right). The top row shows which residues in both molecules are defining the borders of the cavities. The bottom row shows how Ile-65 and Leu-93 prevent the further extension of the pocket into the inner parts of EAT-2 (d) while the same hole reaches to Leu-43 in SAP (c).

4.3. Elucidating the Phosphatase Activity of SAP

As mentioned in the preceding section the SH2 domain SAP has some unique properties compared to other members of that family. Its binding affinities to signaling peptides of the phosphorylated/dephosphorylated SLAM type are more dependent on the residues upstream of the pTyr than on the C-terminal amino acids [85]. Furthermore SAP is known to block the activity of the SHP-2 phosphatase. Recently, a series of biological experiments by Schweighoffer et. al. [120] discovered that SAP shows a phosphatase activity which cannot be found for other representatives of the SH2 family (such as EAT-2, SHIP, SRC or FYN). This functional similarity to protein-tyrosine phosphatases cannot be verified by a corresponding match of the protein sequences (see Figure 4-27). The problem is thus an interesting test case for the theory that similar protein functions are reflected by similar molecular surfaces [131].

As a reference protein for the similarity searches PTP1, a member of the protein-tyrosine phosphatases (PTPases) family, was selected. These enzymes, in concert with protein-tyrosine kinases, regulate a large number of cellular events, including proliferation and differentiation, metabolism, cytoskeletal organization, neuronal development, and the immune response [67]. PTP1B consists of a single domain which has its active site located at the bottom of a shallow cleft. This site is formed by a sequence of eleven amino acids that represents a common motif of the PTP family and includes the catalytic cysteine and arginine residues. This cysteine residue acts as a nucleophilic agent in the catalytic dephosphorylation reaction.

To elucidate the molecular features that cause the phosphatase activity of SAP the 3D structures of SAP in contact with the peptide fragment SLAM (PDB identifier 1D4W) was compared to an inactive mutant of tyrosine phosphatase PTP1B complexed with bis(para-phosphophenyl)methane, Bppm (PDB identifier 1AAX, see also Figure 4-28) [109]. As can be expected from the low sequence similarity of the two proteins, a direct match between the two structures could not be established by means of the alpha carbon atoms or the protein backbone. However, although the structural features of the two proteins are rather different, the corresponding molecular surfaces around the active sites seem to have similar motifs. Hence a series of surface comparisons between the crystal structures of PTP1B, SAP and EAT-2 (1AAX, 1D4W and 1I3Z respectively) were performed. In these experiments the protein surfaces were restricted to the active sites by selecting only those surface points that were located within 8 Å of the ligands' phosphate groups. For the similarity search these points were then augmented by the ESP of the proteins and the experimental details of that setup are given in Table 4-14. For the sake of simplicity let us assume that the residue 215 in the crystal structure 1AAX is still the

```
SAP      1 -MDAVAVYHGKISRETGEK------LLLATGLDGSYLLRD---------------SESV
             |· ·· |·:|· ·  ::      |:     : ||:| :                ·:·
PTP1B   46 YRDVSPFDHSRIKLHQEDNDYINASLIKMEEAQRSYILTQGPLPNTCGHFWEMVWEQKS

SAP     38 PGVYCL----------CVLYHG------YIYT----------------YRVSQTETGSW
             ||  |          |· |          |:                 | | | |  ·
PTP1B  105 RGVVMLNRVMEKGSLKCAQYWPQKEEKEMIFEDTNLKLTLISEDIKSYYTVRQLELENL

SAP     65 SAETAPGVHKRYFRKIKNLIS----------AFQKPDQGIVIPLQYPVEK---------
           ::: :  : : :: ·  :               |:  :·| : | : ||
PTP1B  164 TTQETREILHFHYTTWPDFGVPESPASFLNFLFKVRESGSLSPEHGPVVVHCSAGIGRS
```

Figure 4-27: Sequence alignment between SAP and PTP1B.
The residues that are in close contact (6.0 Å) to the ligand peptide are highlighted in blue (SAP) and red (PTP1B). A **|** means residue identity and **:,·** strong and weak chemical similarity.

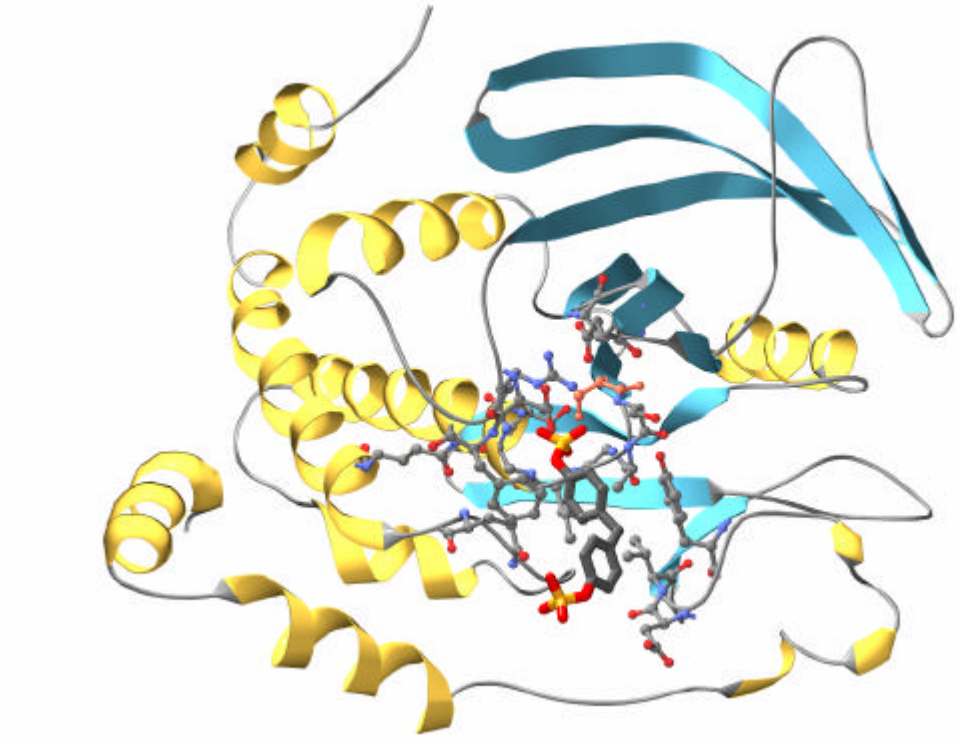


Figure 4-28: Crystal structure of the C215S mutant of PTP1B (1AAX).
The residues of the protein that are within 6.0 Å from the ligand are shown as small ball & sticks. The ligand that is reaching into the active site is rendered with bold ball & sticks. The mutated serine residue is highlighted in red.

natural cysteine.

The investigations discovered a significant surface similarity between the active sites of SAP and PTP1B (see Figure 4-29); it consists of one ridge on one side of the ligands' phenyl rings and two concave patches in the cavity around the phosphate groups of the ligands. In both molecules the phenyl ring of the ligands are surrounded by these similar features and the rest of the cleft that holds them is very well aligned in the superposition of the similar patches. It is interesting to note, that the result of the surface similarity search comes close to an alignment obtained by the plain superposition of the ligands' phenyl rings. No corresponding surface similarity could be detected between the active sites of PTP1B and EAT-2 which correlates well with the biological data.

With the established alignment, one can now look for similar constellations of amino

filter parameter	**symbol**	**section**[a]	**value**	**property**[b]
curvature cut-off range	c_{CR}	2.2.3	2.0 Å	
neighbourhood radius	r_{CP}	3.2	2.0 Å	
fuzzy threshold	F	3.5	0.6	ESP
shape threshold	R	3.6	0.5	STI
distance tolerance	T	3.7	2.0 Å	
minimum distance	d_{min}	3.7	0.5 Å	
angular tolerance	f_{tol}	3.8	15.0 °	

Table 4-14: Experimental conditions used in the SAP/PTP1B comparisons.
[a)] the section in the text where the filter is described
[b)] the molecular surface property applied to the specific filter (ESP, electrostatic potential).

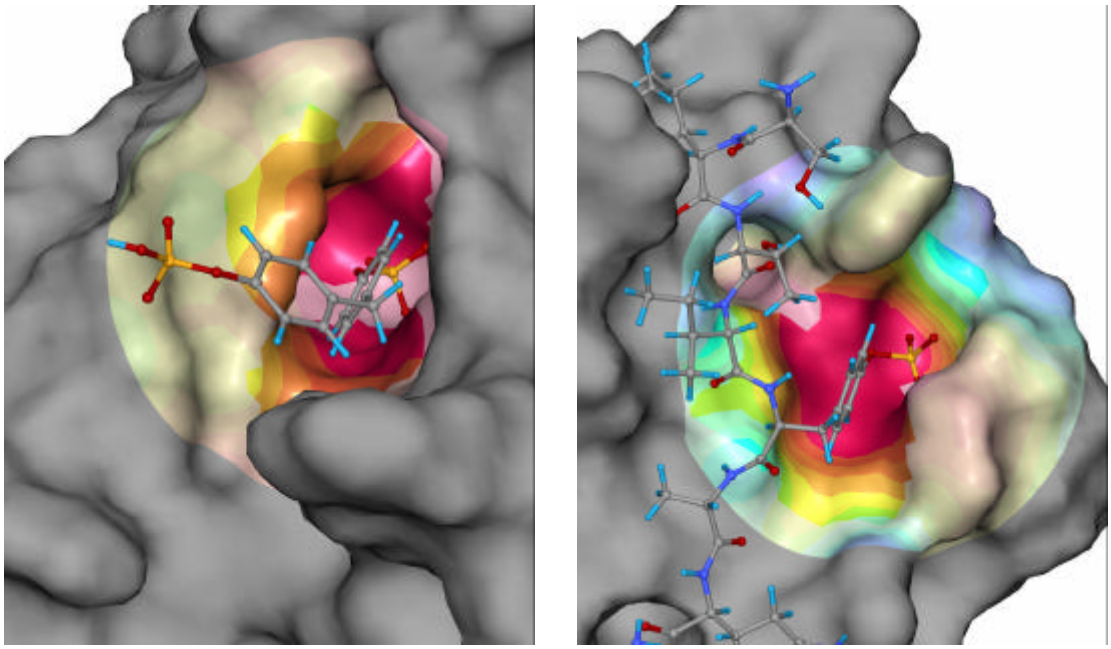
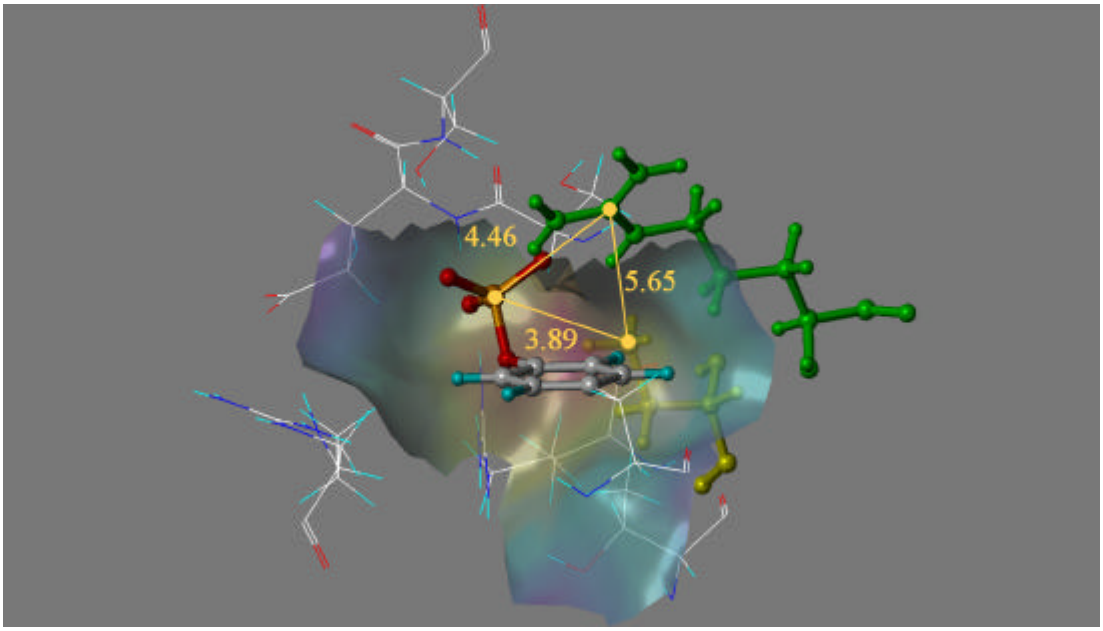


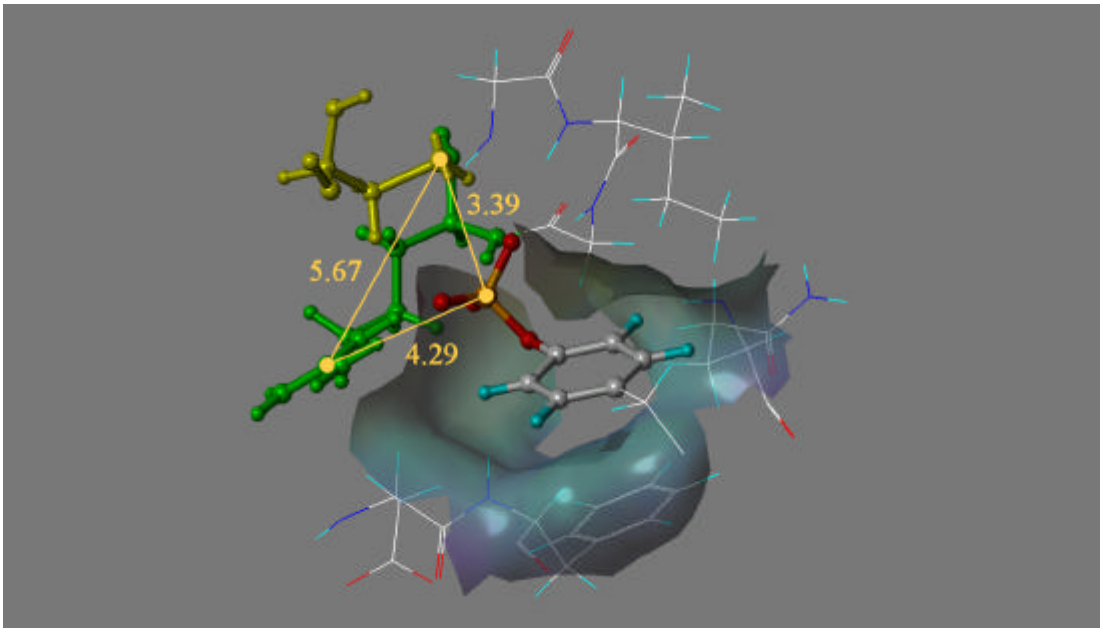
Figure 4-29: Similar surface areas in the active site of PTP1B and SAP.
The similar surfaces in PTP1B (left) and SAP (right) are highlighted in strong colors while the different parts are indicated by less intensive colors. The gray areas are not considered in the surface comparison. The colors are coding the electrostatic potential on the surface; blue represents negative and red positive regions. In both pictures the ligands (Bppm left and pSLAM right) are displayed in balls and sticks with CPK color codes for the elements.

acid residues within the active sites. In both cavities a cysteine and at least one arginine residue are present. These two side chains are involved in the catalytic cleavage of the phosphate group in PTP1B and it is suggested that they are also responsible for the phosphatase activity of SAP. The triangles formed between the cysteine sulfur atom, the central carbon atom of the arginine's guanidine group and the phosphor atom of the ligand are very similar (see Figure 4-30 and Figure 4-31 on page 80). Distances between two atoms in these triangles do not differ by more than 0.5 Å, but the triangles do not coincide in the alignment. However, aligning the triangles would bring the ligands out of a position so that they would not fit into the other active site.

It is suggested that the similar surface regions in both active sites are necessary for the molecular recognition of the ligand structures. In both cases the phenyl ring fits well into the shape of the similar ridge and cleft motif. These structural features may be necessary to bring the substrate in close contact with the catalytic residues. Surface comparison alone cannot answer the question whether the reaction is indeed controlled by the cysteine/arginine residue pairs that are located at different parts of the cleft, because it is a static method that does not consider any dynamic processes. Further structural studies are needed to elucidate the mechanism of the catalytic reaction.



SAP



PTP1B

Figure 4-30: Orientation of the catalytic residues in the active sites.
The residues are displayed to show their orientation with respect to the phosphate groups of the ligands (CPK ball and sticks). The yellow triangle indicates the distances between the ligands' phosphor and the cysteine sulfur and the central carbon of the arginine's guanidine group. The picture clearly shows that the residues in PTP1B (below) are rotated by approximately 180° compared to their counterparts in SAP (above).

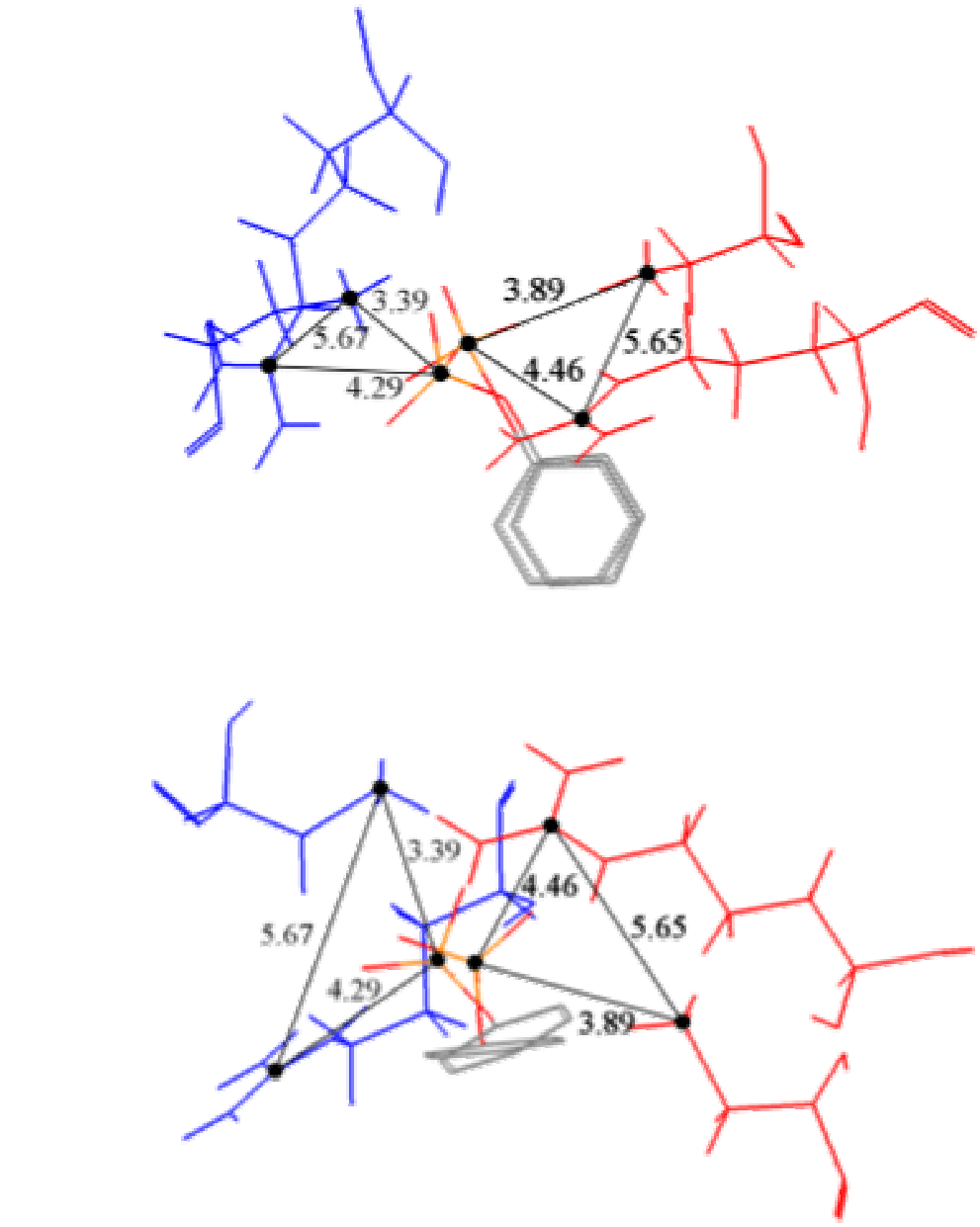


Figure 4-31: Orientation of the catalytic residues and the ligand's phosphate groups
The alignment is based on the surface similarity found between the active sites of SAP and PTP1B. The triangles describe the distances between the important residues and the phosphor atom of the ligands: The distances between the phosphor atoms and the cysteine sulfurs are 3.39 Å (SAP) and 3.89 Å (PTP1B). The carbon atoms of the guanidine group in the arginine residues are placed at distances of 4.46 Å and 4.29 Å and the distance between the two residues is 5.67 Å and 5.65 Å, respectively. The residues of SAP are represented by red and that of PTP1B by blue lines, the phosphate groups and phenyl rings of the ligands are displayed in CPK colors. Top view (above) and side view (below).