High-throughput molecular dynamics simulations: Long and short range effects of mutations on substrate specificity

Fabian Bös[1], Natalia Currle-Linde[2], Peggy Lindner[2], Rolf D. Schmid[1], Jürgen Pleiss[1]

[1] Institute of Technical Biochemistry, University of Stuttgart, Allmanding 31, 70569 Stuttgart, Germany
[2] High Performance Computing Center Stuttgart (HLRS), Allmandring 30, 70550 Stuttgart, Germany

Juergen.Pleiss@itb.uni-stuttgart.de

Substrate specificity of serine β-lactamases

How do enzymes achieve substrate specificity? More specifically, how do structural elements and individual residues contribute to substrate recognition? Serine β-lactamases are an interesting model to study the effect of exchange of individual residues on specificity because new enzyme variants with extended substrate spectra are a major cause of increasing antibiotic resistance of pathogenic bacteria.

Lactam compounds are the most widely used antimicrobial agents applied for the treatment of infectious diseases. β-lactamases have raised major concern due to their capability to hydrolyze a broad spectrum of β-lactam antibiotics and thus confer resistance to bacteria. Therefore, many β-lactams, especially second, third, and fourth generation cephalosporins, have been designed that cannot be hydrolyzed by certain β-lactamases and thus are active against many pathogenic strains. However, the rapid evolution of bacterial β-lactamases, mainly due to the short generation time of bacteria and enhanced selective pressure imposed by application of antibiotics, keeps on generating new enzyme variants. A number of variants emerged that are able to hydrolyze even third and fourth generation cephalosporins, hence their name extended spectrum β-lactamases (ESBL) [Br01]. As the structures of many serine β-lactamases have become available, it became clear that remarkable differences in the biochemical properties are caused by a small number of mutations [MLF98]. Some of the mutations are in the substrate binding site and directly interact with the substrate, but many changes occur far from the substrate. However, the molecular basis of long and short range effects of various mutations is not clear, though there are assumptions on how they may alter the shape of the binding site [MMP02]. A general model that describes the effect of mutations on structure, dynamics, and flexibility of the lactamase and the impact to the interaction between the enzyme and the substrate is still lacking.

We have applied high-throughput molecular dynamics (MD) simulations of a TEM β-lactamase to establish a general, generic molecular model of specificity, which describes short- and long-range effects of mutations on structure and dynamics. The TEM β-

lactamases are the most commonly encountered β-lactamases in gram-negative bacteria. Up to 90% of ampicillin resistance in *E. coli* is due to the production of TEM-1 [Br01]. Up to date, over 120 different sequences of TEM β-lactamases have been isolated and identified from clinical isolates [BJ04]. Surprisingly, these new TEM variants arise from the original TEM-1 enzyme by altering only one to five amino acids at selected loci in the amino acid sequence of TEM-1. Thus the TEM β-lactamases are a challenging system for creating a general model.

Quality of molecular dynamics simulations

The model is based on MD simulations of enzyme variants in complex with covalently bound substrate intermediates, solvated in a water box. MD simulation is one of the principal tools for the study of the time dependent behavior of a molecular system. This method is now routinely used to investigate the structure, dynamics, and thermodynamics of biological molecules and their complexes.

A highly resolved structure of serine lactamase TEM-1 (PDB entry 1M40) [MWS02] was the starting point for estimating the quality of an unconstrained long time scale molecular dynamics simulation of this protein in water. The enzyme-substrate complexes were generated by placing the substrate in the binding site, covalently bound to the catalytic serine. After geometry optimization of the serine-substrate-complex partial charges were fitted using the *RESP* program of *AMBER 7.0* [Ca02] to reproduce the *ab initio* HF/6-31G* electrostatic potential calculated by *Gaussian 98* [Fr98]. The parm99 version of the all-atom *AMBER* force field [Co95] is used to represent the protein system. Using the *xLEaP* program all structures were solvated and the system was neutralized by adding ions. Energy minimization and MD simulations were performed using the *Sander* program of the *AMBER 7.0* software package. Multiple simulations of each enzyme-substrate combination were performed with ten different initial velocity distributions. After heating the enzyme-substrate-solvent systems and equilibration, a trajectory of at least 2 ns was generated. Their stability was verified by following the potential energy and the deviation of the backbone atoms from the experimental X-ray structure. For all simulations, the average structure over the last 500 ps of the trajectory deviated only by 0.9 to 1.2 Å from the X-ray structure. A further quality improvement was reached by averaging over all ten simulations down to a deviation of only 0.8 Å. In addition, not only the experimental structure was well reproduced by the model but also the flexibility, as simulated and experimentally determined B-factors show close similarity. From these studies, we conclude that the proposed molecular dynamics simulations are feasible using the available hard- and software, and the applied simulations parameters are appropriate to reliably simulate the system of a TEM β-lactamase.

Modelling of enzyme-substrate complexes

The serine ß-lactamase mechanism consists of two steps which both require a catalytic base: first, the acylation, where the ß-lactam substrate is covalently attached to the active site serine Ser70, followed by the deacylation, where the substrate is detached from the

Ser70. In recent publications the conserved residue Glu166 is identified as the catalytic base in acylation as well as in deacylation [MWS02, He03a]. This residue is located at the start of the so called ω-loop formed by the residues 161 to 179. The ω-loop is fixed by a salt bridge and hydrogen bonds involving Arg164 and Asp179 (Figure 1). Replacing the arginine residue at position 164 with a serine residue, as in the variant Arg164Ser (TEM-12), leads to the ESBL phenotype. Mutation studies on the TEM β-lactamase resulted in the finding that the Arg164Ser should increase the flexibility of the ω-loop and therefore allow a better positioning of the catalytic base Glu166 [Va99].

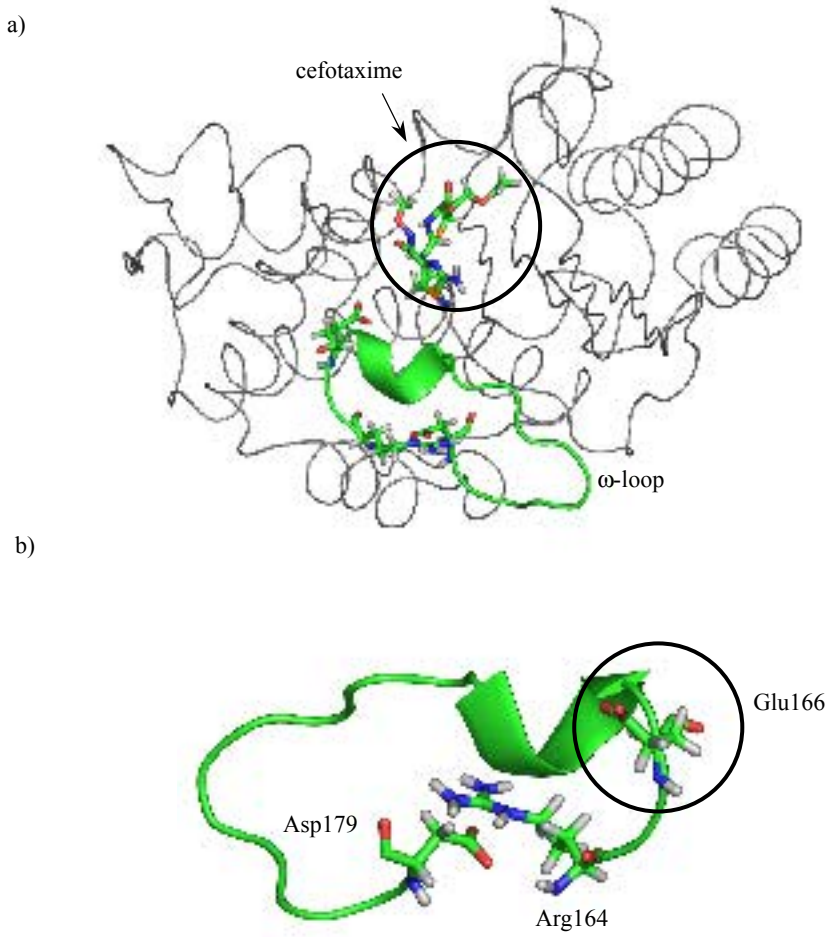

Figure 1: a) Model of TEM-1 β-lactamase with covalently bound cefotaxime. The ω-loop is highlighted in green. b) The ω-loop in detail with fixating residues Arg164 and Asp179 and catalytic base Glu166.

To asses the influence of the Arg164Ser variant we simulated this variant as the substrate free enzyme and in complex with the substrates benzyl-penicillin and cefotaxime. In all simulations with substrates, both benzyl-penicillin and cefotaxime showed the same orientation in the binding side (Figure 2a). In addition the carbonyl oxygen of the β-lactam ring was always orientated towards the oxyanion hole formed by nitrogen backbone atoms of Ser70 and Ala237 (Figure 2b). The bond lengths fulfil the criteria for a hydrogen bond and help to stabilize the covalently bound substrate in the binding site. To estimate catalytic activity, we measured the distances between the Glu166 carboxyl oxygen atoms and serine side chain oxygen atom (Figure 3). The distances were measured in the last 500 ps of the production phase and are averages over three simulations each with a different distribution of starting velocities. While for the benzyl-penicillin complex the distances were similar for wild-type (5.09±0.09 and 4.14±0.08 Å) and the Arg164Ser (5.13±0.20 and 4.22±0.24 Å) variant, the distance in the cefotaxime complex decreased for the Arg164Ser variant (3.07±0.19 and 3.29±0.18 Å) as compared to the wild type enzyme (4.62±0.09 and 3.90±0.12 Å). In the substrate-free enzymes, the distances increased for the variant (5.03±0.19 and 3.78±0.18 Å) compared to the wild type (3.73±0.12 and 2.71±0.06 Å). The local geometry measured by distances seems to be a sensitive probe of the ω-loop's flexibility and a quantitative parameter for the repositioning of the catalytic base Glu166, which is specific for the protein sequence and the structure of the substrate.

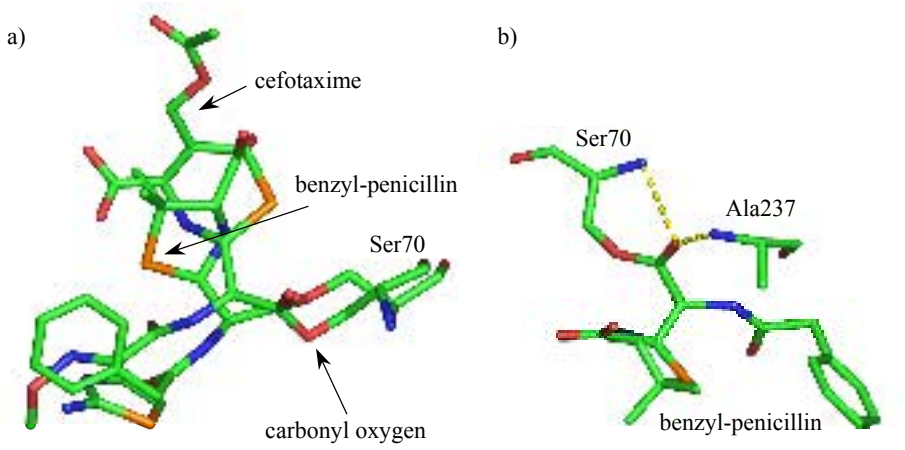

Figure 2: a) Orientation of covalently bound benzyl-penicillin and cefotaxime in the binding site. b) Fixation of the substrate in the binding site.

Currently, further geometrical and dynamical properties are calculated from the simulation trajectories and compared to experimental data on catalytic activity of the respective variant-substrate pair. The analysis includes the average structure of the β-lactamase and the substrate, the enzyme flexibility, correlated motions, hydrogen

bonding networks, distances between catalytic residues and the substrate, orientation of substrates inside the binding site, area and volume of the substrate binding site, and an estimation of binding energy. We expect that changes in one or more of the above mentioned parameters correlate with the experimentally determined catalytic activity. We investigate 20 clinically relevant TEM β-lactamase variants in complex with five different substrates, ranging from the early benzyl-penicillin to the third generation cephalosporin cefotaxime. For many combinations experimental data on kcat/Km are available from literature. As we have previously shown for lipases [SPS00] and esterases [He03b], geometrical parameters are highly predictive probes of biochemical activity.

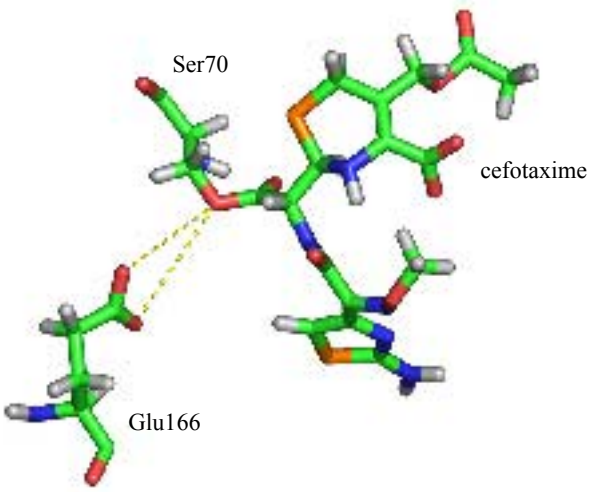

Figure 3: Distances between the carboxyl oxygen atoms of the catalytic base Glu166 and serine side chain oxygen atom in the serine-cefotaxime complex.

A workflow system for production and analysis of MD simulations

The two major bottlenecks are production and analysis of molecular dynamics simulations. Production of each of the 1000 MD simulations will take about 50 processor days each. Each trajectory file has a size of several GB and will be analyzed by up to ten different programs. Thus, production will be carried out in a grid environment including a Linux PC cluster (256 CPUs) at the Institute of Technical Biochemistry and on PC clusters at the High Performance Computing Center. While most projects on MD simulation are still managed by hand, this large-scale parameter studies involve up to thousand MD simulations and therefore will be controlled by a workflow system. Tools are provided to design complex parameter studies combined with control of job execution in a distributed computer network. Furthermore, the system will help users to run experiments which will find their right direction according to a given criteria automatically.

The system architecture of the parameter study tool consists of three components: the User Workstation (Client), the Parameter Study Server and the Object-oriented Database (Figure 4). The DataManager, JobManager and InfoModule of the workflow system communicate not directly with computing resources, but via Grid middleware adaptors. In the first version, an adaptor will be implemented to connect to UNICORE resources. UNICORE [Er03] provides a Grid solution, which allows users to execute their jobs on high performance computing resources and makes these resource available in a seamless and secure manner.

For the design of the individual parameter study, the scientist can use a graphical editor and a database. The database stores the relevant modules (structures of proteins and substrates, topology, run parameters, programs). A complete parameter study consists in preparing, performing, and analyzing multiple MD simulations, which differ in one or several parameters (initial velocities, substrate structure, lactamase variant). Using the graphical project editor (Figure 5), these building blocks are combined, either linearly or in parallel and loop definitions are added.

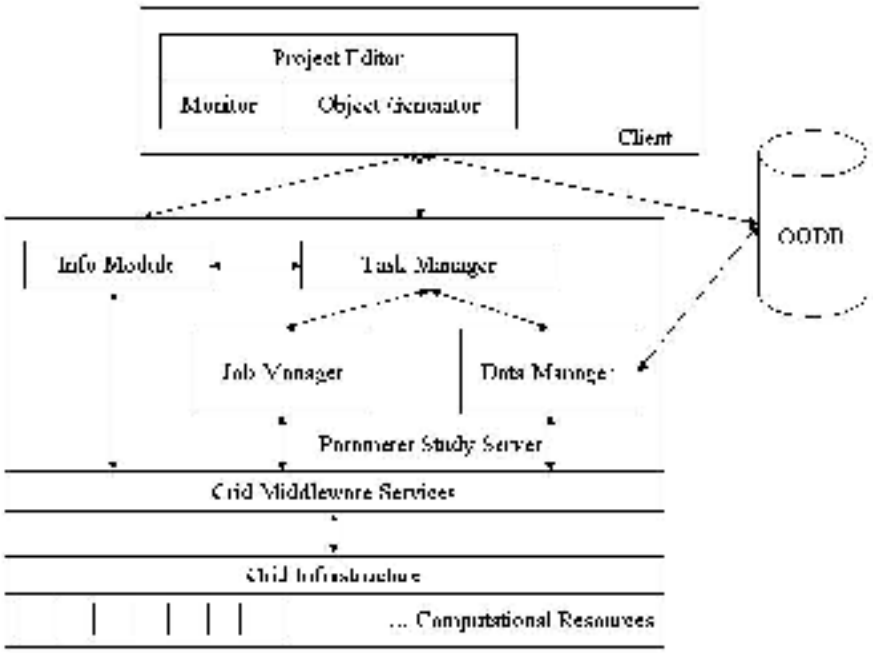

Figure 4: System architecture of the workflow system.

Figure 5 shows the example usage of the graphical editor to generate a complete MD simulation task. Starting from a user provided molecular system including the enzyme, the substrate and the surrounding solvent, three parallel branches will be generated. Every branch consists of the same molecular system, but uses a different initial velocity

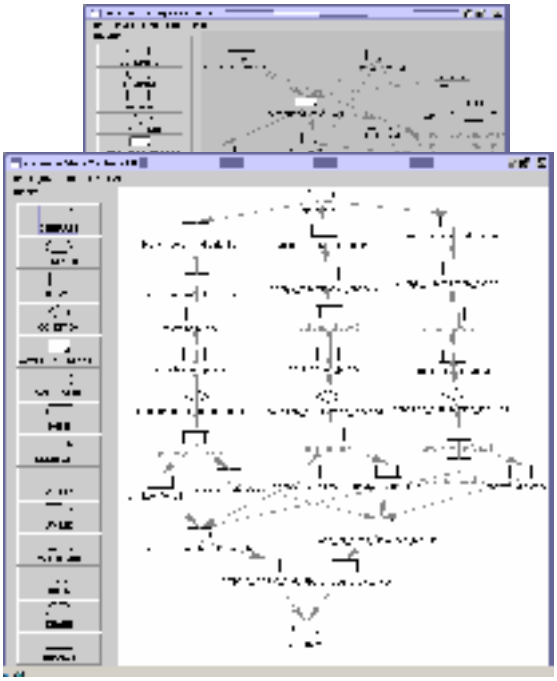

Figure 5: The graphical user interface for creation of the workflow process.

distribution for the simulation in the parameter file. The workflow of every branch follows the aforementioned scheme for MD simulations: First the system will be relaxed to a local energy minimum (energy minimization phase). Subsequently the system is adapted to the simulation temperature in the equilibration phase. An automatic control of the system's relaxation into equilibrium would be of great interest to save calculation time. This can be achieved by monitoring system properties like structure and potential energy, which can be extracted from a plain text simulation output file. The actual system conformation must be compared to all previous conformations of the system by calculating the root-mean-square deviation (RMSD). The equilibrium state is reached when temperature, potential energy, and RMSD values have become stable during a predefined time. Once these conditions are met, the equilibration phase will be terminated and the production phase is started. The resulting trajectory from the production phase will be analyzed. In this example the flexibility of the enzyme and distances between catalytic amino acids will be regarded. At the end of the task the results from all three branches will be compared to each other and the task is terminated.

It remains to be emphasized that the available tool-set for branching, synchronisation and amalgamation of user processes provides the possibility to generate a high performance, hierarchical and multiprocessor-capable parameter study.

References:

[BJ04] Bush K., Jacoby G.: TEM Extended-Spectrum and Inhibitor Resistant ß-Lactamases. URL: http://www.lahey.org/studies/webt.asp, 2004

[Br01] Bradford, P.A.: Extended-spectrum beta-lactamases in the 21st century: characterization, epidemiology, and detection of this important resistance threat. Clin Microbiol Rev 14, 2001; pp 933-951

[Ca02] Case D.A. et. al.: AMBER 7, University of California, San Francisco, 2002

[Co95] Cornell W.D. et. al.: A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. J Am Chem Soc 117, 1995; pp. 5179-5197

[Er03] Erwin D. et. al.: Joint Project Report for the BMBF Project UNICORE Plus, Grant Number: 01 IR 001 A-D, Duration: January 2000 to December 2002', ISBN 3-00-011592-7, 2003;

[Fr98] Frisch M.J. et. al.: GAUSSIAN 98 (Revision A.7), Gaussian, Inc., Pittsburgh PA, 1998

[He03a] Hermann J.C., Ridder L., Mulholland A.J., Höltje H.D.: Identification of Glu166 as the General Base in the Acylation Reaction of Class A-Lactamases through QM/MM Modeling. J Am Chem Soc, 125 (32), 2003; pp. 9590-9591

[He03b] Henke E., Bornscheuer U.T., Schmid R.D., Pleiss J.: A molecular mechanism of enantiorecognition of tertiary alcohols by carboxylesterases. Chembiochem 4, 2003; pp. 485-493

[MLF98] Matagne A., Lamotte-Brasseur J., Frere J.M.: Catalytic properties of class A beta-lactamases: efficiency and diversity. Biochem J 330, 1998; pp. 581-598

[MMP02] Majiduddin F.K., Materon I.C., Palzkill T.G.: Molecular analysis of beta-lactamase structure and function. Int J Med Microbiol 292, 2002; pp. 127-137

[MWS02] Minasov G., Wang X., Shoichet B.K.: An Ultrahigh Resolution Structure of TEM-1 -Lactamase Suggests a Role for Glu166 as the General Base in Acylation. J Am Chem Soc, 124 (19), 2002; pp. 5333 -5340

[SPS00] Schulz T., Pleiss J., Schmid R.D.: Stereoselectivity of *Pseudomonas cepacia* lipase toward secondary alcohols: a quantitative model. Protein Sci 9, 2000; pp. 1053-1062

[Va99] Vakulenko S.B. et. al.: Effects on substrate profile by mutational substitutions at positions 164 and 179 of the class A TEMpUC19 beta-lactamase from Escherichia coli. J Biol Chem 274 (33), 1999; pp. 23052-23060